



Estimating Consumer Tipping Behavior: Review and Recommendations

Prepared for: Internal Revenue Service

Prepared by: Fors Marsh Group, LLC

FINAL

Version 2.0

February 2014

The views, opinions, and/or findings contained in this report are those of Fors Marsh Group, LLC and should not be construed as official government position, policy, or decision unless so designated by other documentation. This document was prepared for authorized distribution only. It has not been approved for public release.

Table of Contents

Introduction.....	1
Literature Review Summary	2
Review of Methodologies.....	6
Recommended Approach	17
Appendix A – Reviewed Articles.....	24
Appendix B – Annotated Citations.....	30
Appendix C – Search Engines and Search Terms	69

Introduction

This report is intended to provide guidance to the IRS as it attempts to develop estimates of tipping and stiffing rates, tipping income, and ultimately, the gap between actual and reported tip income at the aggregate level and by sector. This guidance is based on the results of past research on tipping behavior as well as lessons learned from authors' own work in this area. The first section of this report summarizes the results from a comprehensive annotated bibliography of academic and government literature on tipping. This bibliography, which can be found in Appendix B, includes summaries of research examining average tipping rates as well as individual and establishment characteristics associated with tipping. In anticipation of subsequent sections, the bibliography also summarizes articles that do not directly address tipping, but are relevant to the development of research designs that could be used to collect and analyze data on tipping.

The second section reviews different methods for the collection and analysis of tipping data, and their potential benefits and drawbacks. Topics addressed include sample sources, specifically samples drawn from address-, telephone-, and Internet-based samples; the mode used to collect the data from the sample, including in-person interviews, paper surveys, and Internet surveys; and the design of the survey, including long-recall cross-sectional, short-recall repeated cross-sectional, and longitudinal designs. Finally, this section describes potential methods for analyzing the data, including the use of disaggregated means as well as model-based approaches.

Finally, the third section presents recommended approaches for collecting and analyzing tipping data based on the reviews in the first two sections. This includes both immediate steps pertaining to instrument development as well as pilot testing prior to full scale implementation.

Literature Review Summary

A preliminary set of articles was identified using a bibliography of tipping-related research compiled by Dr. Michael Lynn.¹ Additional articles were identified through backward and forward citation searches starting from the articles identified in the Lynn bibliography. Google Scholar was used to identify more recent research that cited the articles from Lynn's bibliography. Gated articles were accessed through a local University Library System. However, to mitigate the potential for selection bias, queries for articles relevant to tipping and survey methodologies were made using several search engines and archives. This set of search engines and databases included general interest academic archives and search engines such as Google Scholar, JSTOR, and the Social Science Research Network (SSRN) as well as specialized business and accounting-related archives such as Business Source Complete and ProQuest's Accounting & Tax database. Themes and keywords for this search were identified based on an initial review of articles obtained from the Lynn bibliography and the backward and forward searches. From these articles, additional backward and forward searches were conducted to identify additional articles. From the resulting compilation of articles, authors influential to the tipping literature were identified based on total numbers of articles written/published and/or number of citations. These researchers were consulted in order to obtain any previously unidentified tipping-related papers/research, whether published or unpublished.

Many articles touch on multiple topics that are relevant to determining a methodology for data collection and analysis of tip-stiffing and tip rates. Consequently, articles cannot be sorted into mutually exclusive categories based on themes. To facilitate review of evidence from the compiled literature on specific topics, each citation includes a list of the article's themes. The reader can use his or her word processor/PDF reader's search or find functions to quickly discover articles that address a given theme. A list of all themes with descriptive text is included in Table 1. A list of the reviewed articles is provided in Appendix A, with the associated annotations presented in Appendix B. Descriptions of search engines, search terms, and related themes derived from the search are provided in Tables 2 and 3 in Appendix C.

¹http://tippingresearch.com/uploads/Tip_Bibliography.pdf

Table 1. Themes

Theme	Description
METHODOLOGY	Methodology used by article along with relevant benefits and drawbacks.
NATIONAL AVERAGE TIPPING RATES	Article's findings, if any, with respect to U.S.-wide stiffing/tipping rates.
INDUSTRY/SERVICE	Article's findings, if any, with respect to differences in average stiffing/tipping rates across industries/establishment types.
CASH VERSUS CREDIT	Article's findings, if any, with respect to differences in stiffing/tipping rates between establishments/customers who accept/use cash versus credit.
SERVICE CHARGE	Article's findings, if any, with respect to differences in stiffing/tipping rates between establishments that do or do not include automatic tip/service charge.
BILL SIZE	Article's findings, if any, with respect to differences in stiffing/tipping rates between establishments/customers based on bill size.
GEOGRAPHY	Article's findings, if any, with respect to differences in average stiffing/tipping rates across geographic regions and jurisdictions.
INCOME	Article's findings, if any, with respect to differences in average stiffing/tipping rates of customers with different levels of income.
EDUCATION	Article's findings, if any, with respect to differences in average stiffing/tipping rates of customers with different levels of educational attainment.
AGE	Article's findings, if any, with respect to differences in average stiffing/tipping rates between customers based on AGE.
GENDER	Article's findings, if any, with respect to differences in average stiffing/tipping rates between men and women.
RACE/ETHNICITY	Article's findings, if any, with respect to differences in average stiffing/tipping rates of customers with racial/ethnic characteristics.
TIPPING KNOWLEDGE	Article's findings, if any, with respect to customers' understanding of tipping norms (i.e., percent of bill).

Methodology: With respect to the methodologies related to collecting data on consumer/producer expenditure and reporting, the current literature covers many of the trade-offs between maximizing data quality, making causal inferences, and ensuring that the sample and their recorded behavior is representative of the population of interest. Panel-based survey designs, such as the original NPD Group diary panel (McCrohan & Pearl, 1991; Pearl & McCrohan, 1984) and the Bureau of Labor Statistics' Consumer Expenditure Survey, can potentially allow analysts to make inferences about the effects of interventions on individual behavior because of the ability to control for individual-level factors that do not vary over time (Parker, Souleles, & Carroll, 2012). However, panel-based survey designs can also potentially increase respondent burden, leading to increased attrition and selection bias. In addition, consumer diary panels may induce changes in respondent spending behavior, leading to less valid predictions for individuals outside the sample (Crossley & Winter, 2012). A similar trade-off comes with experiments, whether in labs (Alm & Jacobsen, 2007) or fields (List,

2011)², which allow for controlled environments, and thus the estimation of treatment effects at the expense of external validity. Nonpanel surveys, while providing limited ability to make causal inferences about the effects of different interventions on expenditure and reporting, can potentially produce more representative samples because of a relatively lower burden being placed on respondents and consequently higher response rates. However, long recall periods may lead to lower quality of responses because of the inability of respondents to accurately recall the timing of spending occasions (Crossley & Winter, 2012).

With respect to the effects of survey modes and instruments, web-based surveys can lead to more accurate responses than paper-based or in-person/telephone surveys because of the ability of respondents to more easily skip past irrelevant questions, and, in the case of in-person/telephone surveys, the increased time respondents have to look up information necessary to accurately answer questions. In addition, self-administered surveys, which are now primarily web-based, may be more accurate than in-person or phone interviews because of the anonymity that self-administered web-based surveys afford (Crossley & Winter, 2012). However, a sample of individuals with web access may not be perfectly representative of the population of interest because of an individual's probability of web access being related to individual characteristics as well as his or her geography.

Industry/Service: The majority of tipping research has focused on the restaurant industry, but a few studies have focused on other industries where tipping is prevalent. For instance, previous studies have investigated tipping rates for luggage handlers, taxi drivers, bartenders, parking attendants, hotel bellmen, and barber/hair stylists. Koku (2005) concluded that there is a difference between tipping rates in the restaurant industry and outside of it by interviewing customers of each sector. Similarly, Paul and Gardyn (2001) identified higher tip percentages for restaurant servers than for barbers, taxi drivers, food delivery workers, hotel bellmen, and several other professions. However, more research is necessary to provide a more direct comparison between customers' tipping behavior in restaurants and other service industries.

A relatively significant amount of research has been conducted to investigate alcohol's effect on tipping rates. Most of that research demonstrated that customers who consume alcohol provide higher tip percentages than those who do not. Even after controlling for the relationship between bill size and tip percentage and a host of other variables, Lynn (1988) identified a significant effect of alcohol consumption on tip percentage. Similarly, Bodvarsson and Gibson (1997) found that tip rates varied across establishments, but establishments that were licensed to serve alcohol received higher tips.

Cash Versus Credit: Several studies have investigated the difference in tipping rates between restaurant patrons who pay their bill with a credit card and those who pay with cash. Although some articles failed to find a significant difference between payment methods, the majority of research on this subject seems to indicate that customers who paid with credit cards tipped at higher rates than those who paid with cash.

Bill Size: Some research has focused on what is known as the magnitude effect of tipping. The magnitude effect refers to the tendency for customers to leave bigger proportional tips for smaller

² List, J. A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. *The Journal of Economic Perspectives*, 25(3), 3-15.

bills compared with bigger bills. Chapman and Winqvist (1998) concluded that customers provide higher tips on smaller checks, and that the tip percentage decreases as the total bill increases by demonstrating the effect in restaurants and barbershops/hair salons.

Regional Differences: Geographic difference such as between urban and metropolitan areas and between different census regions or divisions of the country, have been investigated previously in studies that contained sufficient sample sizes. Typically, customers in urban areas have been shown to tip at higher percentages than ones in rural areas, and there are some studies showing significant differences in tipping behavior and knowledge between customers from the Northeast region of the country and those in other parts of the country. For instance, McCrohan and Pearl (1983) found that Northeast customers tipped at higher rates than those from middle parts of the country, and Lynn (2006) reported that Northeast customers had a higher knowledge of tipping norms than those in the South. However, the urban/rural difference and regional could potentially be explained by differences in racial composition as well as education and income levels. Studies have demonstrated that these variables significantly influence tipping behavior and tipping knowledge, with higher income levels and educational levels leading to higher tipping and knowledge of tipping norms.

Age and Gender: have been investigated in a number of studies on tipping behavior, and the influence these variables exert on tipping behavior has been somewhat mixed. Gender differences have been inconsistent across studies, whereas age has been linked to significantly higher tipping rates (Pearl & Vidmar, 1988) and greater knowledge of tipping norms (Lynn, 2006). However, differences in tipping behavior and tipping knowledge due to age could be confounded by other factors, such as higher income levels, differences in payment methods, and educational differences. For instance, Lynn (2006) showed in his analysis that higher knowledge of tipping norms with increasing respondent age was nonsignificant when other factors such as education, income, and metro status were included in the analysis.

Race Ethnicity: One of the most researched topics in tipping compliance is how tipping behavior differ between various racial/ethnic groups. Numerous studies have researched this topic, investigating not only actual tipping behavior, but knowledge of tipping norms as well. The findings of these studies have some robust conclusions; primarily, that Black customers tip at lower rates than White customers in the restaurant industry. Though less researched, studies investigating racial/ethnic differences in other tipping industries have reached similar conclusions: Black customers usually tip at lower rates, stiff more often, and tend to leave “flat” tipping rates at higher percentages than White customers, as noted in Lynn’s 2004 study on Black-White tipping differences among service industries (though this effect differs across certain service industries).

Although a significant amount of research has investigated the differences between White and Black tipping behavior, some work has looked into differences between Asian and Hispanic customers. Studies by Lynn in 2006 and 2013 on tipping rates and knowledge of tipping norms indicated that Hispanics tip at lower rates than Whites and have lower levels of tipping knowledge. Asian customers in these studies were not shown to tip at significantly different levels or be less knowledgeable about correct tipping norms. However, these racial/ethnic groups have not been as thoroughly researched as Black and White customers.

Review of Methodologies

For the purposes of developing estimates of consumer tipping by industry, multiple approaches can be taken with respect to the method used to collect the data as well the method used to analyze the resulting data. A data collection approach is defined along several dimensions; specifically, the choice of sampling source, survey mode, and survey design. Methodologies used to analyze the data can be roughly categorized into simple, nonparametric approaches and parametric, model-based approaches.

Sampling Sources

The primary factors to consider when choosing a sampling source are the representativeness of the resulting sample with respect to tipping behavior and the costs associated with recruiting and retaining the sample. Sampling-related bias can result from an unrepresentative frame, non-response, incompletes, and, in the case of a longitudinal panel, attrition. In addition, to mitigate the potential for additional nonresponse that results from transitioning sampled respondents to the survey, survey sources are often coupled with survey modes. To minimize total survey error, biases that result from the survey mode have to be considered when choosing a sampling source. Table 2 discusses the benefits and drawbacks of different sampling sources.

RDD Sample: Random digit dialing (RDD) uses randomly generated phone numbers to select a sample for participation in a survey. RDD sampling is helpful in the sense that it may allow for coverage of the population that has unlisted numbers, but there are problems associated with cell phone users. First, in some RDD methodologies, cell phone users are not reachable, which excludes individuals without household lines from the sample, affecting the generalizability of results. Second, even if cell phone users can be reached, it may not be possible to determine the participant's location from his or her area code because many cell phone users retain their old numbers when moving to different regions of the country, forcing researchers to rely on self-reporting. Self-reporting of basic demographics has a negative influence on response bias analyses because little to nothing is known about the nonrespondents. This is compounded by the fact that RDD response rates continue to decrease with the widespread adoption of caller ID and call screening.

Address-Based Sample: An address-based sampling (ABS) source relies on home address and demographic information from the frame file, which is provided by third-party vendors and the U.S. Postal Service (USPS). This source allows for measureable sample coverage across a population and a fairly well-known probability of the sample selection. Additional household data can be purchased and appended to ABS files to assist in more targeted sampling and further response bias analyses. Although the costs for a mail paper sample are not low, cost per complete has been found to be lower than that for RDD studies³ with certain populations. However, response rates for mail-based paper surveys, which are the most commonly used data collection mode with a mail paper sample,

³ Medway, R.L., Viera, L., Turner, S.R. & Marsh, S.M. (2009). *List-assisted mail as an alternative to random digit dial in a survey of the young adult population*. Paper presented at the 64th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

have continued to decrease over time as people increasingly use the Internet as their medium for correspondence (Dillman et al., 2009b).

Traditional Internet Sample: Traditional Internet samples are collected via an opt-in procedure where individuals choose to join a survey administrator's panel. This panel acts as a potential pool of respondents who are then queried to participate in individual surveys or diaries. An opt-in sample might not be representative of the population of interest with tipping or expenditures on tipping services. Unlike RDD or ABS methods, randomly sampling from an email-based Internet frame is made difficult by the lack of an algorithm that can randomly select email addresses, due to inconsistencies in email address conventions (Dillman, 2009). Yet, similar to the RDD and ABS methods, traditional Internet samples can fit with various data collection modes. Although the online survey is the most straightforward mode for Internet samples, it should be noted that a longitudinal diary approach could also be used, such that all those potential respondents are contacted and recruited to report their tipping behavior for a predetermined amount of time (see Survey Design). Some examples of different Internet samples include:

- **GfK Knowledge-Panel®:** The KnowledgePanel is an Internet-based panel that uses a probability-based sampling strategy where the survey frame is derived from the USPS Delivery Sequence File. Individuals are invited to participate in the panel by mail, followed by telephone calls for those who do not respond to the initial invitation. Households are sampled without replacement, avoiding potential bias that may result from respondents participating in the panel twice. For those individuals selected for participation without computers or an Internet connection, a netbook is provided. This process attempts to mitigate the selection bias associated with web surveys while preserving the benefits associated with a computer interface. The primary benefit of the KnowledgePanel relative to the opt-in panels described below is that knowing the probability of selection allows researchers to estimate error. However, these estimates will always be deficient capturing all aspects of non-response unaddressed by demographic post-stratification. Further, the procedures used to setup and maintain panel membership and participation serve as an additional component of error difficult to fully model and correct for.

Blended Online Sample (Ipsos Ampario): Ipsos' blended sample approach combines the use of its Ampario online sampling method in addition to its iSAY online panel—an online panel of 800,000 members and their households. Ampario is a new, nonprobability sampling procedure that Ipsos has developed that invites respondents by invitations, banner ads, and other means on 100 to 400 websites that have partnered with Ipsos. These two methods are combined into a single sample using Ipsos' proprietary Cortex routing system, which allocates and reallocates a sample given respondent eligibility. Simply put, when respondents are not eligible for one survey, they are immediately redirected to other surveys in progress. In traditional one-off opt-in surveys, noneligible respondents are lost sample, a considerable cost. Finally, Bayesian methodology, which requires previous information regarding the overall sample of interest in order to mix with current information for the final distribution of results, is used to form final distribution. As is the case with a traditional online sample, Ipsos' blended sampling could work with several different data collection modes, but it is best served with an online-based

questionnaire, which could include a cross-sectional administration or a longitudinal diary approach. However, because of the opt-in nature of the Blended Sample, it is not possible to model the probability of response, and thus account for that source of potential bias in survey estimates.

- NPD Group Online Sample: NPD Group utilizes sophisticated techniques both at the sample design stage and post-survey weighting stage to reduce bias and increase representativeness of the sample, but it is not a probability sampling technique.. Although there are certain demographic groups that have less representation online and are not represented in correct proportions as they would be in the U.S. Census, they are large enough that they can be sampled appropriately to represent the U.S. population.

Recruitment of panelists is done using a wide variety of opt-in sources (email, affiliate marketing, co-registration, banners, etc.). The wide variety of sourcing ensures a large representation from various strata of the U.S. population. All sourcing is balanced and ensures no single source provides a disproportionate percentage of recruits. A number of other steps are put in place to prevent fraudulent prospects from joining the panel. This, combined with other behavioral data collected, is used to monitor recruitment source quality and guide media planning for recruitment. NPD limits the number of surveys a panelist can start in a day, week, and month to avoid survey fatigue. Response rates are tracked at an individual panelist level—if panelists fail to participate consistently over time, NPD removes them from the active panel.

The sample for a particular study is drawn from this panel to demographically represent the U.S. population. Sophisticated algorithms take varying response rates by demographic groups into account to provide stratified quota for each of the targeted cells. Once the sample is collected, the cells that fall short in demographic representation during sampling are weighted during processing the data. Again, because of the opt-in nature, it is not possible to model the probability of response, and thus account for that source of potential bias in survey estimates.

Table 2. Summary of Trade-Offs in Sampling Sources

			INTERNET SAMPLES			
Qualities of Sampling Plans for Current Research	RDD Telephone Sample	Address-Based Sample	Traditional Internet Probability Sample	GfK Knowledge Panel® Probability Sample	Ipsos Nonprobability Sample	NPD Nonprobability Sample
General population coverage	Medium	High	Low	Medium/High	Medium	Medium
Known probability of selection	Medium	High	Low	High	Low	Low
Response rate and cooperation rate	Low	Medium	UNK	Low	UNK	UNK
Cost per complete	High	Medium	Low	Medium	Low/Medium	Medium

Survey Mode

As is the case for sampling source, the choice of survey mode can impact the representativeness of the sample by influencing the demographics of those who choose to actually take or complete the survey. In addition, the burden that a particular choice of survey mode places on respondents can influence the accuracy of the data obtained from the survey for a given respondent. Issues of selection bias and measurement error thus have to be considered when choosing the survey mode.

Web-based questionnaire: Some of the many benefits to online surveys include more rapid and reliable transmission of completed questionnaires as well as more flexibility in skip patterns. This can also reduce respondent fatigue by withholding non-applicable items (Crossley & Winter, 2012; Dillman et al., 2009b). Related to this is the lower cost of administering a web survey versus other modes due to the lack of need to send or code a physical questionnaire or have an interviewer make contact with the respondent. The accuracy of responses may also be improved relative to in-person or telephone surveys because of the ability of respondents to retrieve relevant information—a benefit that results from the ability of respondents to answer web-based questions when convenient. Another benefit of web-based (and mail-based) surveys is that social desirability effects that result from the presence of an interviewer can be mitigated. On the other hand, interviewers can diminish the effects of respondent confusion by helping clarify ambiguous questions or following up on inconsistent responses—advantages that a web-based or mail based survey may lack. There is also evidence of a “primacy” effect in responses to visual based surveys (i.e., web- and mail-based ones), where respondents are more likely to pick the first option given in a list of discrete responses (Dillman et al., 2009).

Mail-based surveys: Decreased coverage of telephones and difficulty in estimating coverage in web panels has led to increased use of address-based sampling (ABS) and, subsequently, mail-based surveys. In addition to this positive association with ABS, mail surveys have actually maintained relatively higher response rates than telephone- and web-based surveys. However, mail-based surveys also have a number of weaknesses. They are generally less flexible when it comes to skip logic than web-based surveys. In addition, a mail-based survey may be significantly more costly than

a web-based one in terms of time and money because of the significant variable costs needed to publish survey and mail pieces, transport these pieces between survey administrators and respondents, and code the responses. This disparity is likely to be greatest for larger survey efforts requiring big sample sizes and/or significant follow-up.

In-person surveys: Although there may be many variations of this approach, for the current project in-person surveys may involve interviewers waiting outside of restaurants and other service-industry establishments with the purpose of asking patrons a battery of items associated with their tipping behavior. This approach could allow for immediate recall of a behavior as well as confidence ratings of the data if the interviewer was trained (as done by the Bureau of Labor Statistics [BLS] in “the New Orleans Test”), thus possibly ensuring more reliable data (Crossley & Winter, 2012). However, the cost can be quite prohibitive, particularly with respect to the number of survey administrators and/or transportation required to ensure that different demographic groups, establishment types, and geographic areas are properly represented in the sample. BLS, in particular, has conducted a number of longitudinal in-person studies over the years under its National Longitudinal Surveys program, and has used a number of techniques to keep respondent attrition low. These techniques include giving their researchers access to local resources to track down any respondents who might have moved or passed away since the previous survey and corresponding with the participant to encourage survey compliance (thank-you letters and pamphlets highlighting the data and knowledge gleaned from the survey effort are two examples). To mitigate the social desirability issues previously discussed with this method, BLS has incorporated computer-based response options so respondents can listen to sensitive questions with headphones and type in their responses without their interviewer’s knowledge.

Phone survey: Phone surveys can either be administered by working off a purchased consumer directory or through RDD. With the advent of cell phones, many households no longer use landline phones, and that makes a portion of the population difficult to reach in a cost-effective manner (Pew, 2012).⁴ RDD also lends itself to difficulty in measuring non-response bias given the lack of knowledge of the sample frame and, specifically, the nonrespondents. Overcoming low response rates (and potential selection bias) can require frequent calls, increasing the cost of this mode. Another potential issue is that these types of real-time surveys do not give respondents enough time to refer to their schedules or other sources of information concerning past expenditures compared with web and mail surveys (Crossley & Winter, 2012). This will tend to undermine data quality.

Diary study: Following Pearl and McCrohan (1984) and McCrohan and Pearl (1991), a diary panel can be used to provide data over a certain time span for each observation (i.e., tipping behavior of interest) and has been used for both servers and customers in the past. The fact that respondents are expected to record their expenditures near the time when the expenditure was made can mitigate the effects of recall on response accuracy that would plague a recall-based survey. However, this lack of recall bias can come at the cost of not properly capturing seasonal fluctuations in expenditure and tipping behavior if the diary period is short and/or infrequent. In addition, research burden on the participant is quite high and compliance (in the form of attrition and recorded expenditures) significantly drops off over time. It is also possible that the act of recording

⁴ Pew Research (2012) “Assessing the Representativeness of Public Opinion Surveys.” <http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/>

expenditures may induce a downward trend in expenditures over time (Crossley & Winter, 2012). This learning effect is a well-known research confound whereby subjects modify their future behavior in response to the knowledge and skill they gain by being part of the study.

Mixed-mode surveys: Using multiple survey modes has the potential benefit of increasing response rates because of differences in mode preferences across different respondents (Dillman et al., 2009; Crossley & Winter, 2012). However, given that mode has an effect on response quality, data gathered using different modes will not necessarily produce comparable responses (Dillman et al., 2009). Measurement error due to mode effects may be exacerbated if modes with low degrees of recording error are combined with a mode with a high degree of recording error versus the use of a single mode with a low degree of recording error. There is consequently a trade-off between nonresponse/selection and the potential for measurement error.

Table 3. Summary of Trade-Offs for Alternative Consumer Study Modes

Qualities of Modes for Current Research	Web-Based Questionnaire	Mail-Based Paper Survey	In-Person Interview	Phone Survey	Diary Study*
Interviewer effects	Low	Low	High	High	Low
Learning/Testing effects	Low	Low	Low	Low	Medium
Respondent controls when to participate (at a convenient time)	High	High	Low	Low	Medium
Dynamic question branching	High	Low	High	High	High
Quick data turnaround	High	Low	Low	High	Medium
Immediacy of recall	Low	Low	High	Low	High
Administration costs	Low	Medium	High	Medium	High

*Through electronic diary

Study Design

To obtain a picture of expenditure and tipping behavior that is representative of a given period of interest, several study designs can be employed. These designs would differ with respect to the number of times individual respondents are interviewed, the period over which the interviews take place, and the length of the period over which the respondent is required to recall their tipping behavior. A longitudinal, or diary, study would involve surveying individual respondents about their tipping behavior multiple times over the course of the period of interest. A cross-sectional study involves surveying each respondent once over a short period, while requiring that they recall their tipping behavior for the entire period of interest. Finally, a repeated cross-section would only require that respondents provide information about their tipping behavior for the period immediately preceding the interview, but the interviews would be conducted over the entire period of interest.

Longitudinal: A longitudinal study requires repeated observations of the same subject over a specific length of time. Because the same subjects are tracked over time in a longitudinal study, researchers

can more reliably attribute a change in behavior to an observed variable. In terms of the proposed methodologies, a longitudinal diary study could illuminate changes over the course of a week or across seasons in consumers' tipping behavior. Longitudinal studies could also be used to track the tipping rate over time with a multiyear effort. In addition, when examining the causes of tipping behavior, longitudinal data allows one to control for unobserved individual level factors that affect tipping, enhancing the researchers' ability to make causal inferences. However, these two latter benefits may not be relevant for the purposes of this project. Asking participants to record their tipping behavior for every service-related purchase immediately afterward over a specified period of time (e.g., one week) would allow for data collection among several different service industries without the need for recall. However, attrition among longitudinal studies is certainly higher and places a higher burden on the respondent. Furthermore, longitudinal studies tend to be more expensive than cross-sectional studies that merely ask for participation for a short duration of time.

Cross-sectional: Unlike longitudinal studies, cross-sectional studies do not utilize repeated observations of the same respondent. Instead, cross-sectional studies aim to survey people of different populations at one point in time, allowing for researchers to compare different populations simultaneously. At another time, the researcher surveys a different sample that is estimated to be congruent to the previously surveyed sample. This form of surveying avoids the high costs and high attrition rates associated with longitudinal studies. All of the proposed data collection methods could potentially use a cross-sectional approach. Mail-, online-, and phone-based surveys frequently use single contacts with participants in order to aggregate data for a given population. Similarly, diary studies can take a cross-sectional approach in the sense that participants are asked to provide feedback about tipping behavior over a 24-hour period. In the process, they would rely less on respondent recall, but avoid the burden of high costs and attrition associated with a longitudinal diary study. However, it is more difficult to be sure that changes in variables of interest within populations are due to outside factors, because respondents are being grouped as opposed to following the same respondents over time. In addition, estimates derived from a single cross-sectional survey with a short recall would not accurately reflect annual tipping rates if expenditure and tipping rates vary by season or day of the week.

Repeated Cross-Sectional: Repeated cross-sectional studies, also known as synthetic panels, offer an alternative to longitudinal and single cross-sectional studies (see Parker, Souleles, & Carroll (2012)). Data from multiple cross-sections of survey data would be pooled and respondents sorted into strata defined by multiple, unchanging characteristics (gender, ethnicity for individuals, establishment type and location for establishments/managers). Changes in mean outcome variables (bill size/tipping) for individual strata could then be tracked over time to discern seasonal trends in reported tipping. Unlike single cross-sectional studies, this design/methodology allows variation over time in respondents' tipping (in the case of a consumer) or tip reporting (in the case of server/establishment surveys). In addition, these types of studies are less susceptible to issues associated with longitudinal studies related to survey nonresponse and attrition. The original tipping studies conducted by IRS/NPD, while using data collected through a diary, treated their data as a repeated cross-section for the purpose of analysis.

Analytic Considerations

The goal of the IRS tipping project is to produce estimates of establishment and/or employee tip income that will inform the development of policies that encourage tip reporting. Given that tip income, tip reporting propensity, and optimal policies to encourage tip reporting are likely to vary by sector and geography, estimates of tip income at the industry-location level will likely be more useful to the IRS than more aggregated data. As individual establishments and employees may be less likely to provide accurate responses to surveys that ask about tip income (Simpson, 1997), consumers have been the focus of past research in this area. However, because compliance-based policies are inevitably going to focus on specific types of establishments and locations, consumer tipping data is only useful if consumer tipping can be linked to particular industry-locations. Given that most establishments likely draw the bulk of their customers and tipping revenue locally, this implies that to produce accurate estimates of tipping revenue for particular industry-locations, estimates will have to be produced for relatively small geographic units. This section considers two methods of estimating tipping rates for small geographic areas and their implications for the design of the data collection instrument: Disaggregated Means (DM), and Multilevel Regression and Post-Stratification (MRP).

Disaggregated Means: The simplest approach to estimating tipping rates for particular geographic units would be to simply take the mean tipping rate for all respondents located in a particular geography. Specifically, the estimate is calculated as:

$$T_{jk} = \frac{1}{n} \sum_i^n T_{ijk}$$

Where T_{ijk} is the tipping rate of individual i for sector j in location k and n is the number of individuals in location k . Besides its simplicity, the advantage of DM is that it makes few assumptions relative to a model-based approach such as MRP (see below). The disadvantage of DM is that the number of observed tipping incidents for a given establishment type/location strata may be very small given a nationally representative survey of typical size ($N = 5000$).⁵ Consequently, bill sizes/tip rates for given sectors/locations from the survey will likely be particularly noisy for a nationally representative sample. Indeed, for very small levels of geographic aggregation, such as counties, there may be no observations for a given establishment type to make the estimate. For this reason, a model-based strategy, like that undertaken by McCrohan and Pearl (1991), may, under certain assumptions, be used to extract precise predictions of tipping rates at a more disaggregated level. One such modeling-based approach, MRP, is discussed below.

Multilevel Regression and Post-Stratification: One means of linking customer-level tipping data to establishments while mitigating issues related to noise in small strata is MRP (Gelman & Little, 1997⁶; see Buttice and Highton, 2013, for a recent review and critique). MRP has attained popularity

⁵ Buttice, M. K., & Highton, B. (2013). How does multilevel regression and poststratification perform with conventional national surveys? Forthcoming, *Political Analysis*.

⁶ Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23(2): 127-135.

by social scientists who wish to obtain geographically disaggregated estimates of a quantity of interest.

Analyzing consumer tipping data using MRP would first involve estimating models of consumer expenditure and tipping that take the form:

$$\widehat{E}_{ijk} = \beta X_{ijk} + \alpha G_k + C_j$$

$$\widehat{T}_{ijk} = \beta X_{ijk} + \alpha G_k + C_j$$

Where E is the amount spent by respondent i for a service in sector j in location k ; T is a tip rate calculated by dividing a reported dollar amount in tips by E or by directly asking the respondent for a tip rate; X is a set of observable respondent-level demographic variables such as race, socioeconomic status, etc., that are likely to influence both tipping and expenditure; and G is a set of location-specific factors such as whether the location is part of a rural or urban region that capture variability in expenditure and tipping by sector that is not explained by differences between locations in X . Locations are defined as the market area of the establishment. Although it is likely that the size of a given market area will vary by establishment, it might be more practical to assume that an establishment draws most of its customers from the county or metropolitan area in which it is located. Finally, C is a constant. After estimating model parameters β , α , and C , predictions are generated for strata defined by all N combinations of values of X and G covariates. Poststratification is then used to generate an average tipping rate for a given establishment type/location:

$$T_{jk} = \sum_s^N \frac{E_{sj} P_{sk}}{\sum_s^N E_{sj} P_{sk}} T_{sj}$$

Where P is the population of a given stratum in a given location, taken from census data. Estimates for the average tipping rate for a given sector/location is thus the average tipping rate across all strata, weighted by the strata's expenditure at a given establishment type and the proportion of a location's population in the strata. The benefit of using a quasi linear, additive model to produce predictions for individual strata rather than using nonparametric estimates from the survey is that, if the linear model provides reasonably accurate estimates of expenditure and tipping rates, the resulting strata-level predictions are likely to suffer less from sampling variability in small to moderate sample sizes than nonparametric estimates. The resulting estimated sector-location tipping rates can be multiplied by an establishment's reported bill size to arrive at a prediction for its tip income. This estimate can then be compared with reported tip income to arrive at estimates of tip reporting.

Note that the model described above is more flexible than that presented in McCrohan and Pearl (1991) insofar as it (1) disaggregates tipping occasions by industry for the purpose of the regression and (2) incorporates consumer-level demographic data into predictions. Although the model in McCrohan and Pearl (1991) only allowed predicted tipping rates to vary by establishment type and by limited degree geography (size of metropolitan area and census division), MRP may produce predictions of tipping rates by establishment type for a location that varies not just by metropolitan status and census division but, because of the poststratification step, also by the demographics of a particular locality.

Integrating Data Collection and Data Analysis

Obtaining usable information from consumer tipping data will require that the design of the data collection instrument anticipate the requirements of the methodology used to analyze the data. With the assumption that this methodology will incorporate features of both a DM and MRP, this section reviews some items to consider when designing a survey instrument.

Observable Variables: The poststratification stage of MRP requires counts of demographic strata defined by the individual-level variables in the regression stage for the geographic units of interest (i.e., market areas of establishments). Given this requirement, a review of available 2010 Census or 5-year American Community Survey (ACS) data would allow for a determination of what strata counts are available. This will, in turn, inform the construction of the survey instrument to ensure that relevant demographic data is obtained from respondents. If, for instance, we could obtain data on number of individuals of a given age-race-income strata by county, we would want to make sure we could obtain data—either from the respondents or the survey frame on age, race, and income—similar to the original IRS tipping study (Pearl & McCrohan, 1984; McCrohan & Pearl, 1991), so as to post-stratify by income group, age, and region using strata counts taken from Census data.

Geographic Variation: MRP accounts for regional variation in outcomes of interest (in our case, tipping), by including region-level variables that are thought to predict that outcome. To model the effect of region-level variables on tipping, we will require that our survey/diary sample be drawn from variable localities. With respect to geography, the academic literature on tipping has generally focused on differences in tipping between individuals located in metropolitan and nonmetropolitan areas. This suggests that our geographic variable should be some indicator of urban status or population density. However, this might pose a problem for estimating a multilevel regression in a nationally representative sample given that the overwhelming majority of the country's population lives in urban areas. Consequently, it would probably be advisable to oversample rural areas. To do this, however, it will be necessary to define our urban-rural typology before fielding the survey/diary. Specifically, we will want to decide on the urbanization categorization. One simple categorization scheme is the Rural-Urban Continuum Codes (RUCC) produced by the U.S. Department of Agriculture⁷. RUCC codes incorporate information on a county's population density as well as its proximity (adjacency) to metropolitan areas. The advantage of the use of adjacency is that it may better reflect the proximity of an individual residing in a county to large numbers of other people than would be the case if only the county's population density were considered. One downside to the RUCC relative to a simple measure such as population density is that it is tied to counties. If we decide to use a geographic unit other than counties, using the RUCC scheme would require some means of assigning a status to the alternative unit, which would be simple if the unit were nested within counties, but less so if counties were nested within the alternative unit or if the borders did not align with counties, such as in the case of Designated Marketing Areas (DMA).

Temporal Variation: If tipping is seasonal as past research has suggested, computing an annual average estimate of tipping would be complicated by the potential unrepresentativeness of the sample with respect to tipping. This would be the case within a short recall cross-sectional survey to differences in propensity to respond across the year to the day of the week, or in a diary panel

⁷ http://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation.aspx#_UrMWBfRDu6M

because of attrition. Although this may be mitigated by modeling tipping and expenditure behavior using time effects in order to create a synthetic panel (in the case of repeat cross sections), if the lack of variation is extreme enough, then parameter estimates on the time effects will be imprecise. We might thus want to consider stratifying the sample over days of the week and the year (in the case of a repeat cross section) or have some means of mitigating panel attrition, perhaps by oversampling individuals from demographic groups that have a high probability of attriting⁸ or else by having some procedure in place to bring on additional panelists.⁹

Establishment Types/Sectors: One of the goals of the project is to examine variation in tipping rates by industry and establishment type. This implies the use of an establishment typology. The degree to which survey design will be affected by the need for an establishment typology will depend on the type of information we can obtain from respondents. If we can obtain the name of the establishment where a transaction took place, we may possibly be able to classify the establishment after the survey has been completed depending upon our needs. If that is not feasible, however, we will likely need to obtain information on establishment type from the consumers. In that case, we will have to design the survey such that the options for establishment classification are intuitive and, perhaps most important, limited enough so as not to increase respondent burden to such a level as to increase nonresponse, attrition, or otherwise undermine response quality. The original IRS/NPD diary panel (Pearl & McCrohan, 1984) arguably did a good job of dealing with this trade-off. Individuals were asked to classify establishments into one of six broad categories and then, in a second question, asked to name the type of food served. Consequently, respondents were not confronted with a large typology of establishment types in one list. Defining establishment types and eating occasions by multiple dimensions and then having a separate question for each dimension allows for a detailed typology while minimizing respondent burden. The chosen typology will also have to be meaningful such that the parameters relating the individual and geographic variables to expenditure and tips will be precisely estimated (i.e., not heterogeneous) when estimated for a given type. Also, this taxonomy must be extended to include establishments other than restaurants. It is thus important that we consider how we are likely to obtain information on establishment type, as that will likely inform the degree of trade-off between collecting accurate information and the precision of the categorization. Another consideration trade-off with having a large number of establishment types is the potential lack of variation in terms of expenditure and tipping behavior one will see if the number of individuals who actually used the service is too small. Larger sample sizes may be necessary to obtain at least some variation in spending and tipping for establishment types for which individual patronage is infrequent.

⁸ Frankey and Hillygus (2013) found that non-White respondents were more likely to attrit from the American National Election Study.

⁹ McCrohan and Pearl (1991), for example, used a panel that was replenished quarterly to match strata population targets.

Recommended Approach

Based on the benefits and drawbacks of the methodologies reviewed in this report, the following section provides recommendations for the IRS in developing estimates of tipping and stiffing rates, tipping income and, ultimately, the gap between actual and reported tip income both at the aggregate level and by sector. Given many of the unanswered methodological questions in the literature, this report recommends a two-stage process whereby a small set of methods tests will be conducted prior to full-scale administration. Specifically, we recommend examining the performance of a web-based, repeated cross-section survey administered to both a probability and non-probability internet-based panel. The choice of a probability or nonprobability web panel could be adjudicated in a validation phase (see below).

Sample Source

As discussed in the earlier section, all the sample sources covered (RDD, ABS, or the traditional Internet based samples) have a variety of strengths and weaknesses pertaining to sample-related bias. Although phone and address-based frames may arguably be more representative of the U.S. population as a whole than Internet-based panels, response rates are generally low and have been declining over time (Pew, 2012; Keeter et al., 2006¹⁰; Curtin, 2005¹¹). These low response rates would likely become even more problematic if, as is recommended below, a web-based mode is used to conduct the survey, given the author's experience with low conversion rates of individuals recruited using these methods to a web-based survey. Further, these more traditional methods may become less mandatory as traditional Internet-based sampling sources continue to evolve, minimizing deficiencies of idiosyncratic recruiting methods prevalent with single source opt-in panels. In fact, recent research on "blended" approaches that use multiple online respondent sources have been found to yield results more similar to dual frame RDD.¹² In addition, the GfK Knowledge Panel® continues to use a probability based sampling strategy where the survey frame is derived from the USPS Delivery Sequence File.

While none of these methods has a clear advantage with respect to sample-related bias, the same cannot be said for issues related to cost. As already discussed, recruiting individuals using RDD or ABS are likely to be very resource-intensive. In the case of the former, it might take many attempts to contact a given individual before receiving a response, resulting in high labor costs. In the case of ABS, the requirement that the request be printed and transported to the potential respondent carries obvious costs, and response times may be slow. By contrast, recruiting a sufficient number of individuals from Internet-based panels will likely be less costly because of the panelists' stated willingness to participate and the ease of scaling given relatively low variable costs. Even in the case of the GfK Knowledge Panel®, which recruits its panelists using more costly ABS methods, the

¹⁰ Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly*, 70, 759-779.

¹¹ Curtin, R. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, 69, 87-98.

¹² Vidmar, J., Bricker, D., Young, C., Clark, J., Roshwalb, A., & El Dash, N. (2013). *Using non-probability online surveys for exit polling: The case of the 2012 U.S. Presidential Elections*. Paper presented at the 68th annual meeting of the American Association for Public Opinion Research (AAPOR), October 7, 2013.

recruitment costs would be lower than those for phone- or address-based frames because of the low costs associated with contacting individuals through email.

Consequently, we recommend the use of an Internet-based sample. Further, we recommend pilot testing both probability and non-probability samples in an attempt to validate the quality of the data resulting from samples recruited from each source.

Survey Mode

With respect to the survey mode, this report recommends the use of a web-based survey. The primary reasons being minimization of measurement error and relative cost. Because the survey will require individuals to record their expenditures and tips and categorize the types of establishment for at least a day, the amount of information they may *potentially* have to recall and enter is substantial. In fact, the sheer amount of possible survey branches and associated instruction would make a paper-/mail-based survey extremely burdensome, increasing the probability of nonresponse, attrition, or otherwise incomplete, inaccurate documentation of tipping occasions, undermining the quality of the data. With respect to in-person and phone-based surveys, data quality issues may arise because of interviewer effects as well as the inability of the respondent to invest time in recalling accurate information about his or her tipping behavior. By contrast, a computer-based interface can make finding the type of establishment and entering tipping expenditures relatively easy, through dynamic branching, instruction, and look-ups.

Another clear advantage of web-based modes is related to cost. In-person, phone-based, and mail-based surveys all have high variable costs which are likely to be substantial due to the large number of people that will be required to estimate tipping rates on low frequency behaviors like casino gambling. By contrast, web-based modes can be scaled at relatively low cost.

Survey Design

The primary considerations for survey design are the ability of a specific design to obtain information on tipping that is representative across both individual and time as well as the degree to which different designs increase respondent burden, and thus risk nonresponse/attrition and/or poor data quality. Given these considerations, this report recommends the use of a repeated cross-sectional design. Given that each individual is only surveyed once, in contrast to a consumer diary (longitudinal design), where an individual is expected to record the details of tipping occasions multiple times, respondent burden, and thus the unrepresentativeness of the final sample can be considerably improved. The one-shot nature of the cross-sectional design may also mitigate the risk that the survey itself will influence behavior. One of the primary benefits of a longitudinal design, the potential to make inferences about the causes of individual expenditure behavior, is arguably of limited relevance in this context as the IRS is primarily concerned with estimating tipping and stiffing behavior rather than explaining individual differences related to consumer tipping. Finally, the costs associated with gaining longer term commitments and incentivizing participation can be considerably higher for longitudinal designs.

With the repeated cross-sectional design, we further recommend a short-recall period to increase the accuracy of recall, reduce respondent burden, and consequently minimize the role of measurement error. Shorter recall periods mean that the tipping occasion reported by a given respondent is not

representative of their yearly tipping. However, because of seasonal differences in tipping behavior and the frequency of tipping occasions for specific industries, the repeated nature of the survey increases the potential for variation in both the days of the week and season for tipping occasions in the sample. This variation then allows for the further development of period-specific estimates of tipping using poststratification weighting techniques.

To obtain a large enough sample of respondent-day observations to ensure that there is sufficient variability in low frequency tipping occasions for analysis, the number of respondents used in a repeated cross-sectional study may have to be very large or the recall length extended with the implied increase in measurement error. It should be noted that the IRS' initial tipping study conducted 30 years ago roughly averaged 60,000 respondent-day observations each year (approximately 4,200 respondents over a 14 day period each year). Although this sample size was largely driven by the existing NPD diary data collection this IRS study was attached to, this is roughly the magnitude that we would expect would be necessary to adequately capture the "opportunity for tipping" on low frequency behaviors like casino gambling. For example, as seen in Table 4 below, we estimate needing approximately 76,000 respondent-days to capture 350 casino gambling occasions. This would entail 76,000 respondents if the recall was 24 hours and fewer if the recall length was extended. Although we strongly recommend a short recall period, the day or days this represents should be determined in the piloting stage of the study as prior research does not provide explicit guidance on this key detail. Table 4 provides estimates for the sample size required to obtain different frequencies of tipping occasions by sector. A one-day recall is assumed to remain conservative with the projected estimates.

An alternative to relying on a large nationally representative sample to capture sufficient variation in infrequent activities is to oversample from regions where the activity is expected to be more frequent. This strategy would be most suitable for activities like gambling, where establishments are geographically clustered. Potential complications that result from oversampling arise from the fact that individuals residing in gambling localities may not be representative of the total U.S. population with respect to tipping rates. These differences may reflect the fact that gamblers in high-gambling localities are less likely to be on vacation when they gamble. There may also be systematic differences with respect to demographic characteristics between high gambling and low gambling regions that influence gambling-related tipping. In a model-based approach such as MRP, this could be accounted for by including an indicator for residence in a tipping locality as well as an indicator if the individual were on vacation when the gambling took place. If the assumptions of the model were accurate, relevant differences between gamblers in high gambling regions, gamblers in low gambling regions, and those who gamble on vacation could be accounted for in the final estimate through post-stratification. An alternative approach that avoids the model based assumptions would be to calculate a weighted mean tipping rate, where respondents from oversampled localities would be given a smaller weight such that the weighted sample is representative of the national population with respect to geography. However, this would result in a smaller effective sample of gamblers and gambling occasions, which would increase variance in the final estimate, potentially limiting the benefits of oversampling.

Table 4. Estimated Annual Occurrence

	Occasions per year	Likelihood per day	Required sample for 350	Estimated Number of Occasions for a given Sample Size				
				10,000	30,000	60,000	120,000	240,000
Eating out/take-out fast food*	183.0	0.501	698	5,014	15,041	30,082	60,164	120,329
Eating out/sit down*	68.0	0.186	1,879	1,863	5,589	11,178	22,356	44,712
Salon**	6.3	0.017	20,373	172	515	1,031	2,062	4,123
Hotels/motels**	0.6	0.002	223,826	16	47	94	188	375
Taxi/Limo**	0.6	0.002	210,415	17	50	100	200	399
Casino***	1.7	0.005	76,314	46	138	275	550	1101

Notes: * Estimates of occasions per day taken from Pearl and McCrohan (1981). ** Estimates of occasions per day generated from the detailed monthly expenditure file of the Consumer Expenditure Survey¹³.

***Estimate is an average based on data taken from Shinogle, Norris, Park, Volberg, Haynes, & Stokan (2011) and Volberg, Nysse-Carris, and Gerstein (2006)¹⁴.

Next Steps

This report lays out a general recommended approach; it also leaves open a number of key choices—such as the use of a probability or nonprobability sample, the period of recording/recall, and the type of model (MRP versus DM). These choices are critical as they may lead to invalid predictions due to the data (e.g., selection bias and measurement error) and issues with the model (e.g., included variables and functional form assumptions). Both issues can be relatively difficult to remedy after data has been collected. If the data is measured with error or if there is substantial response bias, it will be unclear what precisely is being modeled and additional rounds of data collection might be prohibitively expensive.

If the dataset does not contain a large range of potentially observable respondent characteristics, then testing alternative model specifications might be impossible. For this reason, before settling upon a final method, we believe it will be important to conduct a set of method studies to examine the validity and feasibility of our recommended approaches.

¹³ For the purposes of calculating the number of occasions per year, a non-zero monthly expenditure on a given activity is assumed to equate with one occasion in that month for the individual respondent. The number of occasions per year is then the fraction of person-months with non-zero expenditure multiplied by 12. Note the assumption that an individual engages in a maximum one expenditure a month likely depresses the number of occasions. Consequently, these estimates should be viewed as conservative.

¹⁴ Shinogle, J., Norris, D. F., Park, D., Volberg, R., Haynes, D., & Stokan, E. (2011). Gambling prevalence in Maryland: A baseline analysis. Volberg R.A., Nysse-Carris K.L., and Gerstein D.R. (2006). 2006 California Problem Gambling Prevalence Survey. Estimates based on Table 4.15 on pg. 26 and Table 3 on pg. 31, respectively. Respondents who list “Past Year Participation,” assumed to gamble at a Casino once per year; “Monthly Participation,” 12 times per year; “Weekly Participation,” 52 times a year. Note that casino gambling is legal in both Maryland and California. In addition, California is in close proximity to Nevada. Consequently, the fraction of the population who reports gambling at a casino, and especially those who visit the casino frequently, may be larger than in the national population. As a result, Table 4 may inflate the number of casino gambling occasions that would be obtained in a nationally representative sample.

Instrument Development: The first step, which would focus on instrument development and choice of recall length, should occur even before a pilot study is initiated. Survey usability testing can be used to identify problems in the self-administration of the surveys and interpretation of survey items and instructions. In its most basic form, usability testing is a pretest in which participants are asked to think aloud while completing the survey instrument and describe their thought process for determining their answer to the survey item. Hearing participants vocalize this “inner speech” provides insight into the respondents’ understanding of the question wording, response categories, and survey organization. After completion of the survey, additional cognitive probing can be done to explore understanding of concepts that did not emerge during the “think aloud” process. If issues are identified, the survey can be refined and additional cognitive interviews will be conducted to verify the changes. In this respect, survey development and usability should be performed iteratively.

One of the primary focuses of this test would be to understand the process through which people recall their expenditures in order to make a consistent decision on one or multiple days of recall. If usability testing, for example, demonstrates that user’s performance is similar in both one- and two-day recall, we would suggest including this variable in subsequent pilot testing.

Pilot Testing: Once an instrument or instruments have been developed, we would suggest a pilot test to further examine the measurement characteristics of the instrument while also examining the use of probability and nonprobability internet panels. As discussed above, the trade-offs between cost and quality are not entirely clear between these two sample sources and would benefit from an empirical test prior to full scale implementation. In addition, to the degree that the usability testing yields ambiguous results with respect to the effects of recall length on accuracy, recall length may also be used to define the set of instruments subject to testing in the pilot phase. We would recommend conducting a test of approximately 20,000 respondent-days (10,000 each method), within one month, spread over approximately 30 Designated Market Areas (DMA). Initial analyses would include an examination of relative differences in estimates, indicators, as well as response characteristics. Although this will provide some evidence as to the consistency of these methods, it will provide little by way of validation evidence. For this, it will be critical to identify a benchmark data source.

One potential source of validation data is point of sale (POS) electronic billing records. Organizations like Restaurant Sciences collect electronic billing records/guest checks to compile useful data for the restaurant industry. This data, including bill and tip totals, can also be purchased by third parties. However, because not all tips are paid using a credit or debit card, such estimates will likely provide an underestimate of total tip income, and therefore cannot be taken as accurate. One way of generating comparable predictions would be to only model expenditures and tipping rates that are paid using a debit or credit card. The dependent variables of the tipping and expenditure models would then be zero if the payment or tip were made using cash, and equal to the amount expended or the tip rate otherwise.

One issue with this type of validation is that the validation metrics would only apply to electronic tipping in restaurants, and would not necessarily say much about the ability of the model to predict nonelectronic, nonrestaurant tip revenue. This type of selection bias would be expected if restaurants that report electronic payments were systematically different from those that do not, with respect to their tip rates. This would be the case, for example, if restaurants with the means to

report their electronic tips were generally better organized. Better organization may be reflected in better service quality and thus higher tips. Another issue with this type of validation is that electronic payment data will likely only be available for restaurants, and thus this data has less to say about the validity for model predictions for nonrestaurant sectors. With these caveats in mind, this out-of-sample data source could provide an extremely valuable source of validation independent of respondent survey data.

Model Validation: Implementing a MRP approach places additional requirements on the data collection instrument. Specifically, for the model to be estimated, the sample will likely have to be stratified geographically in order to obtain variation in the geographic variables. This takes the unweighted sample away from being representative and thus potentially leads to less precise national-level estimates. In addition, depending on the proposed model specification, obtaining information for the individual or geographic variables may increase respondent burden and thus the risk of non-response or attrition. Consequently, model based approaches, and specifically MRP, should be validated in the Pilot stage with respect to its ability to predict regional-level tipping rates.

In the spirit of Buttice and Highton (2013), a potential means of validating the model would be to use the disaggregate mean estimates of tipping from relatively large Restaurant Sciences samples for a set of approximately 30 geographic regions. The number of observations in the given region will be larger than in the primary survey, allowing more precise, non-parametric estimates of tipping behavior in that region. Regions should be chosen for the validation exercise based on dimensions relevant to tipping rates. Specifically, based on prior literature on tipping, we may believe population density or proximity to an urban center is associated with tipping rates. In that case, the sample of validation regions should vary with respect to their level of urbanization. Note that, because the limited number of observations in the pilot sample, urbanization categories may have to be more aggregated than for the final sample in order to obtain sufficient variation in the geographic covariates (i.e. to obtain observations from less dense, rural regions). If the additive assumptions underlying MRP hold, the MR estimates would be expected to look similar to estimates from these region-specific surveys. Of course, this latter validation step does not account for potential systematic measurement error that can affect the accuracy of responses to any survey.

The deviation between the prediction and the ‘observed’ establishment level revenue can be modeled using establishment-level and locality level covariates to provide further guidance with respect to sources of bias. Specifically, we can estimate:

$$|\overline{T}_o - T_{jk}| = \beta O_{ojk} + \alpha G_k$$

In this equation T_o , is the observed tip rate of restaurant o. The left hand side is therefore the difference between the predicted tip rate of establishments in its sector and locality. We model this as a function of both establishment-specific characteristics, O, and locality characteristics (G). Note that the locality definitions and characteristics do not have to match those in the models of consumer tipping behavior. This is important because it allows us to incorporate additional geographic information that explains model error. We might find, for instance, that zip-code tabulation area income explains some of the error in the predicted tip rates. In that case, that would suggest in the full survey, we would want to ensure that we are able to identify the zip code of the respondents for the purpose of modeling. We might also find that, within establishment types,

organizational features such as the size of the establishment affects error. To account for this, for the final data collection instrument, we might want to ensure that we are able to collect relevant information about the establishment in order to incorporate those characteristics into our sector typology for the purposes of either DM or MRP, even if it comes at the price of increased respondent burden and risk of selection bias.

Appendix A – Reviewed Articles

- Alm, J., & Embaye, A. (2013). Using dynamic panel methods to estimate shadow economies around the world, 1984–2006. *Public Finance Review*, 41(5), 510–543. Themes: METHODOLOGY
- Alm, J., & Erard, B. (2013). *Using public information to estimate informal supplier income*. Working paper. Themes: METHODOLOGY, INDUSTRY/SERVICE
- Alm, J., & Jacobson, S. (2007). Using laboratory experiments in public economics. *National Tax Journal*, 60(1), 129–152. Themes: METHODOLOGY
- Anderson, J. E., & Bodvarsson, O. B. (2005). Do higher tipped minimum wages boost server pay? *Applied Economics Letters*, 12, 391–393. Themes: GEOGRAPHY, NATIONAL AVERAGE TIPPING RATES
- Anderson, J. E., & Bodvarsson, O. B. (2005). Tax evasion on gratuities. *Public Finance Review*, 33, 466–487. Themes: GEOGRAPHY
- Ayres, I., Vars, F. E., & Zakariya, N. (2005). To insure prejudice: Racial disparities in taxicab tipping. *The Yale Law Journal*, 114, 1613–1674. Themes: RACE/ETHNICITY
- Azar, O. H. (2007). The social norm of tipping: A review. *Journal of Applied Social Psychology*, 37(2), 380–402. Themes: BILL SIZE
- Bodvarsson, O. B., & Gibson, W. A. (1997). Economics and restaurant gratuities: Determining tip rates. *American Journal of Economics and Sociology*, 56(2), 187–203. Themes: INDUSTRY/SERVICE, BILL SIZE, GEOGRAPHY
- Borzekowski, R., & Kiser, E. K. (2008). The choice at the checkout: Quantifying demand across payment instruments. *International Journal of Industrial Organization*, 26(4), 889–902. Themes: GEOGRAPHY
- Boyes, W. J., Mounts, W. S., Jr., & Sowell, C. (2004). Restaurant tipping: Free-riding, social acceptance, and gender differences. *Journal of Applied Social Psychology*, 34(12), 2616–2625. Themes: INDUSTRY/SERVICE, GENDER, INCOME
- Brewster, Z. W. (2012). Racialized customer service in restaurants: A quantitative assessment of the statistical discrimination explanatory framework. *Sociological Inquiry*, 82(1), 3–28. Themes: RACE/ETHNICITY
- Brewster, Z. W., & Mallinson, C. (2009). Racial differences in restaurant tipping: A labour process perspective. *The Service Industries Journal*, 29(8), 1053–1075. Themes: RACE/ETHNICITY
- Chapman, G. B., & Winqvist, J. R. (1998). The magnitude effect: Temporal discount rates and restaurant tips. *Psychonomic Bulletin & Review*, 5(1), 119–123. Themes: BILL SIZE, INDUSTRY/SERVICE

-
- Crossley, T. F., & Winter, J. K. (2012). Asking households about expenditures: What have we learned? In *Improving the measurement of consumer expenditures*, National Bureau of Economic Research. Themes: METHODOLOGY
- Curtin, R. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, 69, 87-98.
- Davis, S. F., Schrader, B., Richardson, T. R., Kring, J. P., & Kieffer, J. C. (1998). Restaurant servers influence tipping behavior. *Psychological Reports*, 83, 223-226. THEMES: GENDER, GEOGRAPHY
- Even, W. E., & Macpherson, D. A. (in press). The effect of the tipped minimum wage on employees in the U.S. restaurant industry. *Southern Economic Journal*. Themes: GEOGRAPHY, NATIONAL AVERAGE TIPPING RATES, METHODOLOGY
- Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26, 132-139. Themes: METHODOLOGY
- Feinberg, R. A. (1986). Credit cards as spending facilitating stimuli: A conditioning interpretation. *Journal of Consumer Research*, 13(3), 348-356. Themes: CASH VERSUS CREDIT
- Fernandez, G. A. (2004). The tipping point—gratuities, culture, and politics. *Cornell Hospitality Quarterly*, 45(1), 48-51. Themes: TIPPING KNOWLEDGE, RACE/ETHNICITY
- Filion, K., & Allegretto, S. A. (2011). *Waiting for change: The \$2.13 Federal subminimum wage* (Briefing Paper No. 297). Economic Policy Institute and Center on Wage and Employment Dynamics. Themes: GEOGRAPHY, NATIONAL AVERAGE TIPPING RATES, GENDER
- Frankel, L. L., & Hillygus, D. S. (2013). Looking beyond demographics: Panel attrition in the ANES and GSS. *Political Analysis*. Advance online publication. doi:10.1093/pan/mpt020 Themes: METHODOLOGY
- Frash, R. E., Jr. (2012). Eat, drink, and tip: Exploring economic opportunities for full-service restaurants. *Journal of Foodservice Business Research*, 15, 176-194. Themes: INDUSTRY/SERVICE
- Garrity, K., & Degelman, D. (1990). Effect of server introduction on restaurant tipping. *Journal of Applied Social Psychology*, 20(2), 168-172. Themes: CASH VERSUS CREDIT
- Green, L., Myerson, J., & Schneider, R. (2003). Is there a magnitude effect in tipping? *Psychonomic Bulletin & Review*, 10(2), 381-386. Themes: INDUSTRY/SERVICE, BILL SIZE
- Greenberg, A. E. (2014). On the complementarity of prosocial norms: The case of restaurant tipping during the holidays. *Journal of Economic Behavior & Organization*, 97, 103-112. Themes: GEOGRAPHY

-
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009–1055. Themes: METHODOLOGY
- Hill, D. J., & King, M. F. (1993). An exploratory investigation into consumer knowledge of tipping etiquette: Accuracy, antecedents and consequences. In W. Darden & R. Lusch (Eds.), *Proceedings of the symposium on patronage behavior and retail strategy: Cutting edge III* (pp. 121–135). Themes: TIPPING KNOWLEDGE
- Jargon, J. (2013, September 4). IRS rule leads restaurants to rethink automatic tips. *The Wall Street Journal*. Retrieved from <http://online.wsj.com/news/articles/SB10001424127887323893004579055224175110910>. Themes: SERVICE CHARGE
- Keeter, S., Kennedy, C., Dimock, M., Best, J. & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly*, 70, 759-779.
- Kerr, P. M., Domazlicky, B. R., Kerr, A. P., & Knittel, J. R. (2006). An objective measure of service and its effect on tipping. *The Journal of Economics*, 32(2), 61–69. Themes: INCOME, GENDER, CASH VERSUS CREDIT
- Klee, E. (2004). *How people pay: Evidence from grocery store data*. Federal Reserve Board. Retrieved from http://www.newyorkfed.org/research/conference/2006/Econ_Payments/Klee_b.pdf
Themes: AGE, INCOME
- Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692. Themes: METHODOLOGY
- Koku, P. S. (2005). Is there a difference in tipping in restaurant versus non-restaurant service encounters, and do ethnicity and gender matter? *Journal of Services Marketing*, 19(7), 445–452. Themes: INDUSTRY/SERVICE, GENDER, RACE/ETHNICITY
- Koku, P. S. (2007). Some significant factors that influence tipping in service encounters outside the restaurant industry in the United States. *Services Marketing Quarterly*, 29(1), 23–45. Themes: INDUSTRY/SERVICE
- Lynn, M. (1988). The effects of alcohol consumption on restaurant tipping. *Personality and Social Psychology Bulletin*, 14(1), 87–91. Themes: BILL SIZE, INDUSTRY/SERVICE
- Lynn, M. (2004). Black-White differences in tipping of various service providers. *Journal of Applied Social Psychology*, 34(11), 2261–2271. Themes: INDUSTRY/SERVICE, RACE/ETHNICITY
- Lynn, M. (2004). Ethnic differences in tipping: A matter of familiarity with tipping norms.

-
- Cornell Hospitality Quarterly*, 45(1), 12–22. Themes: TIPPING KNOWLEDGE, RACE/ETHNICITY
- Lynn, M. (2006). Geodemographic differences in knowledge about the restaurant tipping norm. *Journal of Applied Social Psychology*, 36(3), 740–750. Themes: TIPPING KNOWLEDGE, RACE/ETHNICITY, AGE, INCOME, EDUCATION, GENDER, GEOGRAPHY
- Lynn, M. (2006). Tipping in restaurants and around the globe: An interdisciplinary review. In M. Altman (Ed.), *Handbook of Contemporary Behavioral Economics: Foundations and Developments* (pp. 626–643). M. E. Sharpe Publishers. Themes: CASH VERSUS CREDIT, BILL SIZE, RACE/ETHNICITY, GENDER
- Lynn, M. (2011). Race differences in tipping: Testing the role of norm familiarity. *Cornell Hospitality Quarterly*, 52(1), 73–80. Themes: RACE/ETHNICITY
- Lynn, M. (2012). The contribution of norm familiarity to race differences in tipping: A replication and extension. *Journal of Hospitality & Tourism Research*. Advance online publication. doi:10.1177/1096348012451463. Themes: RACE/ETHNICITY
- Lynn, M. (2013). A comparison of Asians', Hispanics', and Whites' restaurant tipping. *Journal of Applied Social Psychology*, 43(4), 834–839. Themes: BILL SIZE, RACE/ETHNICITY
- Lynn, M., & Gregor, R. (2001). Tipping and service: The case of hotel bellmen. *International Journal of Hospitality Management*, 20, 299–303. Themes: INDUSTRY/SERVICE
- Lynn, M., & Latane, B. (1984). The psychology of restaurant tipping. *Journal of Applied Social Psychology*, 14(6), 549–561. Themes: CASH VERSUS CREDIT, GENDER, BILL SIZE
- Lynn, M., & McCall, M. (2000). Gratitude and gratuity: A meta-analysis of research on the service-tipping relationship. *The Journal of Socio-Economics*, 29(2), 203–214. Themes: METHODOLOGY
- Lynn, M., & McCall, M. (2000). *Beyond gratitude and gratuity: A meta-analytic review of the predictors of restaurant tipping*. Working paper, School of Hotel Administration, Cornell University. Themes: CASH VERSUS CREDIT, BILL SIZE, GENDER, RACE/ETHNICITY
- Lynn, M., & Thomas-Haysbert, C. D. (2003). Ethnic differences in tipping: Evidence, explanations, and implications. *Journal of Applied Social Psychology*, 33(8), 1747–1772. Themes: RACE/ETHNICITY, AGE, INCOME, EDUCATION, GENDER
- Lynn, M., & Williams, J. (2012). Black-White differences in beliefs about the U.S. restaurant tipping norm: Moderated by socio-economic status? *International Journal of Hospitality Management*, 31(3), 1033–1035. Themes: RACE/ETHNICITY
- Lynn, M., Zinkhan, G., & Harris, J. (1993). Consumer tipping: A cross-country study. *Journal of Consumer Research*, 20, 478–488. Themes: GEOGRAPHY

-
- McCall, M., & Belmont, H. J. (1996). Credit card insignia and restaurant tipping: Evidence for an associative link. *Journal of Applied Psychology*, 81(5), 609–613. Themes: CASH VERSUS CREDIT
- McCrohan, K. F., & Pearl, R. B. (1983, August). *Tipping practices of American households: Consumer based estimates for 1979*. 1983 Program and Abstracts Joint Statistical Meetings, Toronto, CA. Themes: NATIONAL AVERAGE TIPPING RATES, GEOGRAPHY, INCOME, CASH VERSUS CREDIT
- McCrohan, K. F., & Pearl, R. B. (1991). An application of commercial panel data for public policy research: Estimates of tip earnings. *Journal of Economic and Social Measurement*, 17, 217–231. Themes: NATIONAL AVERAGE TIPPING RATES, CASH VERSUS CREDIT, INDUSTRY/SERVICE, GEOGRAPHY
- Morran, C. (2013, September 5). Are these the final days of automatic 18% tips at restaurants? *Consumerist*. Retrieved from <http://consumerist.com/2013/09/05/are-these-the-final-days-of-automatic-18-tips-at-restaurants/>. Themes: SERVICE CHARGE
- Neuman, S. (2013, September 5). IRS to count automatic gratuities as wages, not tips. *NPR*. Retrieved from <http://www.npr.org/blogs/thetwo-way/2013/09/05/219290573/irs-to-count-automatic-gratuities-as-wages-not-tips>. Themes: SERVICE CHARGE
- Noll, E., & Arnold, S. (2004). Racial differences in tipping: Evidence from the field. *Cornell Hospitality Quarterly*, 45, 23–29. Themes: RACE/ETHNICITY
- Papp, T. G., & Burkhammer, A. L. (2001, March). *An investigation of server posture and gender on restaurant tipping*. Paper presented at the 22nd Annual Industrial Organizational Psychology and Organizational Behavior Graduate Student Conference, Pennsylvania State University. Themes: GENDER
- Parker, J. A., Souleles, N. S., & Carroll, C. D. (2012). The benefits of panel data in consumer expenditure surveys. *National Bureau of Economic Research*. Themes: METHODOLOGY, INDUSTRY/SERVICE
- Paul, P., & Gardyn, R. (2001). The tricky topic of tipping. *American Demographics*, 23(5), 10–11. Themes: NATIONAL AVERAGE TIPPING RATES, INDUSTRY/SERVICE, GEOGRAPHY
- Pearl, R. B. (1984). *A survey approach to estimating the tipping practices of consumers*. Special report on regression analysis to the Internal Revenue Service under contract TIR-81-21, Survey Research Laboratory, University of Illinois, Champaign, IL. Themes: GEOGRAPHY
- Pearl, R. B., & McCrohan, K. F. (1984). Estimates of tip income in eating places, 1982. *Statistics of Income Bulletin*, 3(4), 49–53. Themes: METHODOLOGY, NATIONAL AVERAGE TIPPING RATES, INDUSTRY/SERVICE

-
- Pearl, R. B., & Sudman, S. (1983). *A survey approach to estimating the tipping practices of consumers*. Final report to the Internal Revenue Service under Contract TIR-81-21, Survey Research Laboratory, University of Illinois. Themes: NATIONAL AVERAGE TIPPING RATES, CREDIT VS CASH, INCOME, GEOGRAPHY, INDUSTRY/SERVICE
- Pearl, R. B., & Vidmar, J. (1988). *Tipping practices of American households in restaurants and other eating places: 1985–86*. Supplementary report to the Internal Revenue Service under Contract TIR 86-279, Survey Research Laboratory, University of Illinois, Champaign, IL. Themes: CASH VERSUS CREDIT, GEOGRAPHY, INCOME, EDUCATION, AGE
- Pearl, R. B., & Vidmar, J. (1988). *Tipping practices of American households in restaurants and other eating places: 1985–86*. Supplementary report to the Internal Revenue Service under Contract TIR 86-279, Survey Research Laboratory, University of Illinois, Champaign, IL. Themes: GEOGRAPHY, INDUSTRY/SERVICE
- Pew Research Center. (2012). Assessing the representativeness of public opinion surveys. 1-25. <http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/>
- Rind, B. (1996). Effects of beliefs about weather conditions on tipping. *Journal of Applied Social Psychology, 26*(2), 137–147. Themes: INDUSTRY/SERVICE
- Sanchez, A. (2002). The effect of alcohol consumption and patronage frequency on restaurant tipping. *Journal of Foodservice Business Research, 5*(3), 19–36. Themes: RACE/ETHNICITY, AGE, CASH VERSUS CREDIT, INDUSTRY/SERVICE
- Schwer, R. K., & Daneshvary, R. (2000). Tipping participation and expenditures in beauty salons. *Applied Economics, 32*, 2023–2031. Themes: SERVICE/INDUSTRY, INCOME, AGE, GENDER
- Seiter, J. S., & Weger, H., Jr. (2013). Does a customer by any other name tip the same? The effect of forms of address and customers' age on gratuities given to food servers in the United States. *Journal of Applied Social Psychology, 43*, 1592–1598. Themes: METHODOLOGY, AGE
- Simpson, H. (1997). Tips and excluded workers: The New Orleans test. *Compensation and Working Conditions*, Bureau of Labor Statistics, 32–36. Themes: SERVICE/INDUSTRY, GEOGRAPHY
- Speer, T. (1997). The give and take of tipping. *American Demographics, 19*(2), 51–54. Themes: INDUSTRY/SERVICE, GEOGRAPHY, INCOME, GENDER.
- Star, N. (1988). *The international guide to tipping: When, where, and how much to tip in the U.S. and around the world*. New York, NY: Berkley Books. Themes: INDUSTRY/SERVICE, GEOGRAPHY
- Thomas-Haysbert, C. D. (2002). The effects of race, education, and income on tipping behavior. *Journal of Foodservice Business Research, 5*(2), 47–60. Themes: INDUSTRY/SERVICE, RACE/ETHNICITY, INCOME, EDUCATION

Appendix B – Annotated Citations

Alm, J., & Embaye, A. (2013). Using dynamic panel methods to estimate shadow economies around the world, 1984–2006. *Public Finance Review*, 41(5), 510–543.

METHODOLOGY: Article uses a model-based approach to estimate the size of the shadow economy for 111 countries across the world for the period 1984 to 2006. The shadow economy is defined as the production of goods and services that are not included in government accounts. To estimate the shadow economy, the authors model the demand for currency, defined as the amount of cash over M2. Cash-based transactions are assumed to be relatively easy to hide from the state. Consequently, economies dominated by shadow activities are expected to also be cash-based, all other things being equal. Cash demand is modeled as a function of proxies for levels of development such as urbanization and per capita income as well as country-level characteristics that are thought to influence the incentive to conceal income from the government (thus increasing the demand cash), including bureaucratic quality, the tax rate, and the level of inflation. The use of panel data provides more observations and thus degrees of freedom than prior country-specific, time-series based analysis of cash demand while also allowing the authors to correct for endogeneity in the predictors. The resulting model is used to predict cash demand as well as a counterfactual set of predictions where there is no incentive to hide income (when government quality, and thus enforcement is at its maximum, the tax rate is zero, and there is no inflation). The predictions for cash demand where there is no shadow economy is subtracted from the total predicted cash demand to arrive at an estimate of cash demand that is due to tax evasion. This estimate is then multiplied by money velocity and divided by GDP to arrive at an estimate of the shadow economy as a fraction of GDP. The results indicate a negative association between the size of the shadow economy and the level of development. However, the mean for OECD (Organization for Economic Cooperation and Development) countries across the entire period is still a substantial 16.9%, and 13.3% for the United States in 2006.

Alm, J., & Erard, B. (2013). *Using public information to estimate informal supplier income*. Working paper.

METHODOLOGY: Authors use responses from the 2001 Current Population Survey (CPS) to estimate informal supplier (self-employment) income and tax noncompliance. Specifically, they develop estimates of national informal supplier income using income information provided by self-employed respondents working in 11 industry categories in which informal suppliers will be prominent. To estimate the income of “Food Caterers and Roadside Stands,” the authors use responses from the Bureau of Labor Statistics’ Consumer Expenditure Survey (CES). They then compare these estimates

with income reported to IRS National Research Program (NRP) in these industries to arrive at industry category-level estimates of tax noncompliance. Supplementary CPS surveys were used to identify second jobs and second job income was imputed based on the assumption that secondary income comprised 26.5% of wages. This fraction was in turn estimated using a subsample of respondents who reported income for both jobs. In addition, self-employed, informal income misclassified as wages was assumed to comprise 4.08% of wages. The resulting estimates of total self-employed income (\$156.4 billion) for the 11 CPS industry categories exceeded reported income estimated from the NRP (\$50.9 billion), but was lower than an estimate of total income derived from NRP data (reported income + audit detected + estimated non-detected).

Alm, J., & Jacobson, S. (2007). Using laboratory experiments in public economics. *National Tax Journal*, 60(1), 129–152.

METHODOLOGY: Provides a review of literature using laboratory experiments in the field of public economics. The article lays out the requirements for the successful expectation of an experiment studying the effect of incentives on behavior, including control over the experimental environment such that monetary incentives be explicitly linked to behavior, that instructions are clear, that the experiment not be too long or complicated, and that instructions should not use terminology that hints at the research question that the experiment addresses, which the authors argue could potentially influence the subjects' behavior. Common criticisms of experiments include the argument that the mainly university student subject pool of most laboratory experiments is not representative of the wider population whose behavior and motivations the experiment is trying to analyze/explain (though the authors argue this concern is unfounded), that subjects modify their behavior as a result of the awareness that they are participating in an experiment, and that certain factors that affect behavior in the real world, such as the threat of prison time, cannot be plausibly simulated in a laboratory setting. Consequently, results of an experiment may not generalize outside of the laboratory setting. The article also discusses the use of laboratory experiments to address questions related to the determinants of tax compliance behavior. These experiments typically find that audits increase compliance (though there are diminishing marginal returns as the audit rate increases), that the fine rate increases compliance (though the effect is small), and that higher marginal tax rates lead to lower compliance. Higher income is found to lead to greater compliance. Targeted audits have been found to be more effective in increasing compliance than random audits. Democratic participation and an effective social norm supporting tax compliance increase individual compliance.

Anderson, J. E., & Bodvarsson, O. B. (2005). Do higher tipped minimum wages boost server pay? *Applied Economics Letters*, 12, 391–393.

DESIGN OVERVIEW: Authors investigate if there is any difference in server pay between states with varying levels of subminimum wages and tip credits for tipped staff. A probit analysis was used, and there were 100 total observations in the analysis: one observation for waiters and one observation for bartenders for each of the 50 states (Washington, D.C., was not mentioned in the article and was likely excluded). Data was pulled from the Bureau of Labor Statistics “Wages by Area and Occupation” file (additional data was pulled from the U.S. Census Bureau, the National Restaurant Association, and the Bureau of Economic Analysis). Analysis controlled for the percentage of firms exempted from state and federal minimum wage laws, and restaurants’ revenue as a proportion of the GDP, in addition to other control variables such as age and whether the state has a state income tax.

AVERAGE TIPPING RATES: OLS regression findings indicate that there was a very small difference between states with no minimum wage or tip credit versus states with no tip credits and wages that exceed federal standards, but that overall there was no noticeable difference between the minimum wage of waiters and reported wages.

Anderson, J. E., & Bodvarsson, O. B. (2005). Tax evasion on gratuities. *Public Finance Review*, 33, 466–487.

DESIGN OVERVIEW: The authors used state-level data from the Bureau of Labor Statistics (BLS) to determine if total reported pay is affected by factors that are hypothesized to affect underreporting of tips. The BLS’ Occupational Employment Statistics (OES) surveys are used to estimate the mean and median hourly pay for over 750 occupations, and the authors used restaurant-related occupations for testing their model. Two variables are included to proxy average customer tipping rate (i.e., premium full-service restaurants as a percentage of full restaurants in the state and the percentage of each state’s population living in urban areas). They also included several control variables to account for slight differences in job characteristics and locations.

GEOGRAPHY: Reported pay is higher in areas with a higher tipped minimum wage and in states with no income tax. IRS audit rates do not appear to have an effect on reported pay by restaurant employees. The most important result from their analyses was that higher tax rates raise the employee’s reported pay, such that one percentage point increase in a state’s minimum income tax rate results in servers reporting 13 cents more in pay.

Ayres, I., Vars, F. E., & Zakariya, N. (2005). To insure prejudice: Racial disparities in taxicab tipping. *The Yale Law Journal*, 114, 1613–1674.

DESIGN OVERVIEW: 12 taxicab drivers (6 Black, 4 White, and 2 “other minorities”) completed surveys immediately after dropping off customers for a total of 1,066 completed surveys. Tips were calculated by subtracting the fare from the total cost of the ride. Drivers recorded sex, race, age, passenger dress (proxy for wealth), and driver experience. They also recorded other interaction characteristics, including whether they paid with cash.

RACE/ETHNICITY: White drivers were tipped 61% more than Black drivers (20.3% versus 12.6%) and 64% more than “other minority” drivers (20.3% versus 12.4%). Black drivers were 80% more likely to be stiffed than White drivers (28.3% versus 15.7%) and “other minority” drivers were 131% more likely (36.4% versus 15.7%). The mean tipping percentage of Black customers was 42% of the mean tipping percentage of White customers (9.2% versus 21.6%). Hispanic customers’ mean tipping percentage was just over half of White customers’ mean tipping percentage (12.0% versus 21.6%). Asians tipped 75% of the White customers’ mean tipping percentage (16.2% versus 21.6%). White customers stiffed the driver (10.6%) less frequently than Blacks (39.2%), Hispanics (34.3%), and Asians (15.8%). Using a regression analysis and controlling for random driver effects, time, manner, and place effects, Black drivers are tipped 9.1% less than White drivers. In the most complete regression, Black passengers tipped 9% less than White passengers.

Azar, O. H. (2007). The social norm of tipping: A review. *Journal of Applied Social Psychology*, 37(2), 380–402.

DESIGN OVERVIEW: A literature review of various tipping-related areas, including both theoretical motivations behind tipping behavior and empirical studies on the subject. Areas of focus include the relationship between service quality and tipping behavior, patronage frequency, bill size, service quantity, and other variables.

Bodvarsson, O. B., & Gibson, W. A. (1997). Economics and restaurant gratuities: Determining tip rates. *American Journal of Economics and Sociology*, 56(2), 187–203.

DESIGN OVERVIEW: Authors test several hypothesis derived from economic theory on the determinants of tipping. Data is based on 697 respondents to a survey conducted in 7 Minnesota restaurants. Data collected included bill and tip size, number of food and drink items ordered, number of people at the table, whether the respondent visited the establishment at least once a month, and an assessment of service quality. To account for potential measurement error in tipping due to social desirability bias, the tip rates reported by customers were passed by the servers who

gave an assessment of their plausibility. Their answer was affirmative. Tip amounts and tip rates were analyzed using both descriptive statistics and multivariate regression analysis.

INDUSTRY/SERVICE: Tip rates varied across establishments; establishments that were licensed to serve alcohol received higher tips.

BILL SIZE: Tip amount was positively related to tip amount bill size, and only marginally related to service quality, consistent with the existence of a lower bound on the amount customer's tip.

GEOGRAPHY: Tips (amounts and rates) were higher in restaurants located in St. Paul than in St. Cloud, consistent with tips being higher in larger urban areas.

Borzekowski, R., & Kiser, E. K. (2008). The choice at the checkout: Quantifying demand across payment instruments. *International Journal of Industrial Organization*, 26(4), 889–902.

DESIGN OVERVIEW: Article examining roughly 1,500 households over the course of three months from March through May of 2004. The survey was conducted as part of the University of Michigan Survey of Consumers, a telephone-based survey that covers various aspects of consumer behaviors and attitudes. Various scenarios were presented to respondents, including one suggesting that a “flash” debit service has been introduced to see changes in behavior and another that attempts to “age” the cohort to see changes in behavior. Overall, it was reported that debit cards were overcoming the use of cash and checks for consumers. However, given that the scenarios presented ask about usage when purchasing items from a supermarket, payment methods will likely be very different for tipping situations, because checks are often not appropriate or accepted for tipping situations or establishments.

GEOGRAPHY: Of the four regions, the West region had the highest predicted market share of debit and credit usage (53% for the two) compared with 46.5% for the South, 41.6% for the Northeast, and 38.4% in the Midwest.

Boyes, W. J., Mounts, W. S., Jr., & Sowell, C. (2004). Restaurant tipping: Free-riding, social acceptance, and gender differences. *Journal of Applied Social Psychology*, 34(12), 2616–2625.

DESIGN OVERVIEW: Study investigating tipping behavior using in-person survey intercepts at 18 different restaurant locations, 10 surveys per restaurant. Analysis was used to determine if social acceptance and free-riding influence tipping behavior. Additional variables included customer gender. In-person intercepts were used at each restaurant, asking respondents various questions about their demographics, the size of their party, whether they are a local resident of the area (used as a proxy to determine if they were a repeat customer), how often they eat out and how often they have eaten at the restaurant in the past month, and ratings about the quality of their meal.

Respondents were also asked if they had any alcohol or not. Surveys were only asked during dinner hours to maintain consistency; roughly 90% of respondents agreed to respond to the survey, and a third of surveys were confirmed with the servers of the restaurant for accuracy.

Furthermore, restaurants were classified into four different restaurant types

INDUSTRY/SERVICE: Alcohol consumption had a significant impact on the tipping percentage such that respondents who indicated they had consumed alcohol left higher tips.

GENDER: Men tipped less than women, even when other factors were held constant. In addition, men's tips were found to be more significantly influenced by party size.

INCOME: Higher levels of income were related to higher tipping rates.

Brewster, Z. W. (2012). Racialized customer service in restaurants: A quantitative assessment of the statistical discrimination explanatory framework. *Sociological Inquiry*, 82(1), 3–28.

DESIGN OVERVIEW: A paper survey was given to servers from a sample of 18 chain-style restaurants. Overall, 200 completed surveys were gathered. The aim of the survey was to determine whether servers discriminate against customers of various races (based on questions asking if the quality of service will vary by race). The author acknowledged that explicit questions about racial tendencies in this way will lead to some lack of variability in reporting behaviors because people will wish to report in a way consistent with a social-desirability bias. Respondents were given a series of five scenarios (in which the customer race was held constant as Black customers in various configurations) and asked whether the customers were good or bad tippers (on a 5-point scale). Respondents were also asked what they considered to be good and bad attributes of diners and to provide ratings of the dining behaviors of the Black individuals in the scenarios. Servers were also asked if they preferred to serve various situations (such as groups with or without children, social classes of their clients, etc.).

RACE/ETHNICITY: Overall, nearly 1 in 5 servers reported an explicit preference for serving White clients. In addition, on the 4-point scale regarding service-quality discrimination (1 = never and 4 = always), the mean score was 1.49, indicating that a reasonable number of servers were willing to report some discriminatory behaviors against their customers. Findings seem to indicate that once discriminatory tendencies toward other groups are taken into consideration (such as children, etc.), that servers who report more positivity toward Blacks are less likely to discriminate against them in their service. However, given their use of a proxy variable for discriminatory behaviors, the findings have to be considered with caution.

Brewster, Z. W., & Mallinson, C. (2009). Racial differences in restaurant tipping: A labour process perspective. *The Service Industries Journal*, 29(8), 1053–1075.

DESIGN OVERVIEW: Literature review of two theoretical frameworks that try to explain the reasons for lower tipping behavior among Blacks. The two frameworks that are discussed are that (1) Blacks are unaware of tipping norms, hence leading to lower tipping behavior and (2) that Blacks tip at lower rates because service providers (i.e., waiters) treat Black customers poorly because they anticipate poor tips, creating a cyclical problem.

Chapman, G. B., & Winqvist, J. R. (1998). The magnitude effect: Temporal discount rates and restaurant tips. *Psychonomic Bulletin & Review*, 5(1), 119–123.

DESIGN OVERVIEW: Subjects included 50 undergraduate students participating for course credit. Subjects completed a questionnaire that included two sections: an intertemporal choice and three tipping scenarios. The tipping scenarios comprised a taxi ride, a restaurant dinner, and a haircut. Each scenario included a brief description and asked how much the participant would tip based on bill size. They were presented with four different magnitudes for each tipping setting. Participants were also asked how much they had paid and tipped the last time they had used each of the service scenarios.

INDUSTRY/SERVICE AND BILL SIZE: Tip percentages decreased with bill magnitude for each of the three tipping scenarios, but ANOVA revealed a significant effect of magnitude for the haircut and restaurant dinner scenarios. The magnitude effect (i.e., tip percentages decrease significantly as the bill size increases) was found to be present in both of these scenarios, indicating that participants reported leaving bigger tips for smaller bills.

Crossley, T. F., & Winter, J. K. (2012). Asking households about expenditures: What have we learned? In *Improving the measurement of consumer expenditures*, National Bureau of Economic Research.

METHODOLOGY: Article reviews literature examining the benefits and drawbacks of different methods of collecting household expenditure data through surveys. There is little evidence to suggest the superiority of single survey modes (face-to-face interviews, telephone interviews, self-administered questionnaires); while self-administered questionnaires may increase response rates and quality by allowing respondents time to recall their expenditure patterns and reduce confidentiality relative to modes requiring an immediate response to the interviewer, interviewers may be able to provide more assistance to respondents who have issues with question comprehension. Recall surveys may lead to downward biases in reported expenditure due to poor

recall relative to diaries as well the inclusion of expenditures from before the survey reference period, but diaries may lead to respondent attrition and a decline in the accuracy of responses as time passes due to the greater imposition on respondents. This may lead to a downward bias in expenditure estimates in diaries versus recall surveys, and has been found to be problematic in the case of expenditures on food. Expenditure data collected from diaries with short time frames may also show greater variance due to the fact that respondents report expenditures as they are made, and there may be a large degree of variance in expenditures in short time periods, particularly with respect to infrequent expenditure categories. The keeping of diaries may also influence respondents' expenditure patterns, resulting in biased estimates of population expenditure patterns. Diary respondents may also tend to aggregate different expenditures when they are made at the same time.

The format of survey questions has also been found to have an effect on data quality; open-ended formats lead to rounding of responses, while closed formats may lead respondents to choose categories that they perceive as reflecting their relative expenditures (high spender, high-spending bin) as opposed to their true expenditure. Aggregated expenditure categories tend to lead to lower total expenditure estimates, perhaps due to an inability of respondents to recall every type of expenditure. On the other hand, more disaggregated expenditure categories may put a greater burden on respondents and thus lead to lower quality (less accurate) responses. Using single respondents to solicit information on household expenditures may lead to lower-quality estimates, but using multiple respondents per household may place a greater burden on the household and consequently result in lower response rates. Incentives for completing the survey or diary may increase both response rates and data quality. Data quality can also be improved by asking respondents to reassess their expenditure estimates when they are inconsistent with previously given information, such as total budget.

Davis, S. F., Schrader, B., Richardson, T. R., Kring, J. P., & Kieffer J. C. (1998). Restaurant servers influence tipping behavior. *Psychological Reports*, 83, 223–226.

DESIGN OVERVIEW: Twenty-eight servers from a pair of restaurants (one in a small Midwestern town, 12 servers; and another in an urban area, 16 servers) recorded their tips for a four-week period while alternating whether they stood or squatted by tables in order to determine if that increased tip size. Aside from varying the squat/standing procedure, other descriptive measures including whether the meal was for lunch or dinner and what the gender of the server was were maintained for analysis. Of the 12 servers in the rural area, 7 were women and 5 were men, and there was an even 8/8 split in the urban area. Servers maintained all of the recordings, including the dollar amount of

the meal and the tip that they received. Possible issues with this study are that there is no mention of an incentive for the servers to maintain accurate record-keeping and that they might be misreporting their tips as a whole.

GEOGRAPHY: The study found that people from urban areas tipped significantly more than those from rural areas, but because the servers were not able to determine any kind of socioeconomic variables such as income or education, this might be a spurious effect caused by other variables.

GENDER: Female servers received significantly greater tips than male servers (15.6% compared with 14.1%, though this was the smallest of the significant findings).

Even, W. E., & Macpherson, D. A. (in press). The effect of the tipped minimum wage on employees in the U.S. restaurant industry. *Southern Economic Journal*.

DESIGN OVERVIEW: Two sets of regression analyses were run (specifically, the regressions were a version of “difference-in-difference estimation”—additional details and citations about this regression method can be found in the article): one using data from the Quarterly Census of Employment and Wages (QCEW) and the other using data from the Census Bureau’s Current Population Survey from 1990 through 2011. The regression equation controlled for changes due to season and various demographic variables that would change earnings in the industry, and accounted for both the federal minimum wage and the subminimum wage, among other factors.

Both data sources have their advantages. The QCEW data is pulled from unemployment insurance reports, ensuring essentially mandatory compliance for reporting. However, this data does not provide work hours for workers, nor does it give characteristics of the workers. CPS data, on the other hand, provides such characteristics, but because of methodology the sample for certain industries and states can be quite small and introduce the possibility of error. Both data sets were acknowledged to have specific strengths and weaknesses for their analysis.

NATIONAL AVERAGE TIPPING RATES: Findings from analyses of both data sources indicate that the salary of tipped workers does increase along with minimum wage increase, though the QCEW data points out that this only occurs among full-service restaurants and is not seen among limited service restaurants. Further findings indicate that increases in the minimum wage for tipped employees has a negative influence on the employment of this population and that raises in this minimum wage lead to reduced hours worked per week in addition to higher wages.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26, 132–139.

METHODOLOGY: Article reviews literature addressing factors that affect web response rates. Factors related to survey content include: the sponsor of the survey, with response rates being higher when the survey's sponsor is an academic or government agency; the content of the survey, with surveys asking questions concerning highly salient issues typically receiving higher response rates than those whose subject is less relevant to potential respondents; the length of the survey, with longer surveys having lower response rates. Sample design and contact methods also influence response rates: web panel designs typically yield higher response rates than single-shot surveys, while email-based contact can result in low response rates because of spam filters. However, the use of personalized messages, prenotifications, and reminders can raise response rates. Empirical work examining the influence of incentives (such as an electronically mailed gift certificate) on response rates has generally found small (or even negative) effects on participation. The survey frame also affects response rates, with surveys of the general population generally yielding lower response rates than surveys of specific populations such as employees, though top managers are less likely to respond than lower-level managers/employees. Populations with low socioeconomic status are less likely to respond because of limited Internet access, though this effect persists even after controlling for such access. The personalities of potential respondents also influence response rates, with more conscientious individuals having a greater propensity to respond.

Feinberg, R.A. (1986). Credit cards as spending facilitating stimuli: A conditioning interpretation. *Journal of Consumer Research*, 13(3), 348–356.

DESIGN OVERVIEW: One hundred and thirty-five customers were observed at random intervals over a one-week span at a local restaurant. Servers recorded party size, check amount, mode of payment, and amount of tip. The author also conducted four experiments investigating characteristics of credit card spending, but none of them dealt with tipping or the service industry.

CASH VERSUS CREDIT: A 2 (payment method) x 4 (check size divided into quartiles) ANOVA revealed that when credit card stimuli were present, customers left a significantly higher tip. For each quartile of check size, customers paying with credit cards provided higher tips. Credit card-paying customers, on average, left a tip that was 16.95% of the total bill, while cash-paying customers left a tip that was 14.95% of the total bill.

Fernandez, G. A. (2004). The tipping point—gratuities, culture, and politics. *Cornell Hospitality Quarterly*, 45(1), 48–51.

DESIGN OVERVIEW: Discussion about knowledge of tipping behavior, how the knowledge is passed on, and a discussion about what underlies the racial differences in tipping. Some topics that are

discussed are underlying psychological issues that might be at work within the Black community, including how the segregation of service in restaurants in the past might be the cause of certain behaviors in the present. The author calls for a national study to look at this subject, with enough of a sample to investigate racial differences across different areas with the sufficient detail needed to draw concrete conclusions.

Filion, K., & Allegretto, S. A. (2011). *Waiting for change: The \$2.13 Federal subminimum wage* (Briefing Paper No. 297). Economic Policy Institute and Center on Wage and Employment Dynamics.

DESIGN OVERVIEW: Analysis was conducted using the Census Bureau's Current Population Survey from 2008–2009. Descriptive results of reported wages were split by several demographic groups, including worker gender, race, age, education, and across various states with differing levels of wages for tipped employees.

NATIONAL AVERAGE TIPPING RATES: Overall, it was found that states with higher levels of subminimum wages had higher reported hourly wages for waiters and tipped workers than states with lower tipped minimum wages for tipped workers. However, it is worth noting that the median wage of workers was higher in those states overall, indicating that the relative affluence of those states are driving these changes.

GENDER: Demographic splits indicate that while females constitute the majority of tipped workers and waiters (72.9% and 72.4%, respectively) they earn less on average than male workers, particularly among waiters (\$9.04 for females and \$9.87 for males).

Frankel, L. L., & Hillygus, D. S. (2013). *Looking beyond demographics: Panel attrition in the ANES and GSS. Political Analysis. Advance online publication. doi:10.1093/pan/mpt020*

METHODOLOGY: Article examines the determinants of respondent attrition in the American National Election Studies (ANES), an online panel survey, and the General Social Survey (GSS), a face-to-face interview panel survey using logit regression. Both respondent demographics and survey experience characteristics are included as predictors of attrition. Among the demographic characteristics, age, education, and employment were negatively associated with attrition in the ANES, while non-English preferences and the number of young children were positively associated with attrition. Age and education had a statistically significant negative association in GSS, while foreign born and single member household status were positively associated with the probability of attrition. Among the survey experience characteristics, respondents to the ANES who reported a monetary motivation, had a negative experience, and/or took a long time to complete the survey were more likely to attrite, as were those who refused to answer the survey in the first wave. For the GSS, interviewer

experience was found to be negatively associated with the probability of attrition, and respondents who were interviewed by females were less likely to attrite.

Frash, R. E, Jr. (2012). Eat, drink, and tip: Exploring economic opportunities for full-service restaurants. *Journal of Foodservice Business Research*, 15, 176–194.

DESIGN OVERVIEW: The author pooled point-of-sale (POS) processed guest checks and their associated credit card checks from two restaurants (one fine dining establishment and one casual-theme full-service restaurant). One hundred and fifty checks were randomly selected from each restaurant's weekly pool and each check had to meet several conditions, namely that the checks had to include both food and alcoholic beverages, be from restaurants' dining rooms (i.e., no checks from the bar), be tendered after 5:00 p.m., paid by only one party, and not include any promotional or employee discounting. From the guest and credit card checks, the author recorded reliably accurate information for the guest check dollar amount, percentage of the guest check spent on alcoholic beverage purchases, server's gender, dollar tip amount, and tip percentage. Time the guest check was rendered and day of the week were also recorded.

INDUSTRY/SERVICE: Two hundred and ninety-seven guest checks were included in the final analysis from the two restaurants. The median percentage of the guest check that was attributable to alcoholic beverages was 26.8%, the median guest check was \$40.67, and the median tip percentage was 20.6%. A multiple regression was performed to predict the tip percentage from percentage of the guest check used on alcoholic beverages. A positive relationship was found between the percentage of guest check attributable to alcoholic beverages and the tip percentage of the whole guest check.

Garrity, K., & Degelman, D. (1990). Effect of server introduction on restaurant tipping. *Journal of Applied Social Psychology*, 20(2), 168–172.

DESIGN OVERVIEW: Forty-two, 2-person parties that ordered a Sunday brunch at a restaurant were randomly assigned into two interaction conditions. In one condition, the server greets the customer while introducing herself; in the other condition, the server just greets the customer.

CASH VERSUS CREDIT: Customers that used a credit card as a form of payment left, on average, larger tips than those using cash (22.6% versus 15.9%).

Green, L., Myerson, J., & Schneider, R. (2003). Is there a magnitude effect in tipping? *Psychonomic Bulletin & Review*, 10(2), 381–386.

DESIGN OVERVIEW: In order to determine if there is a magnitude effect in tipping (i.e., as bill size increases, percentage tipped decreases), researchers had two taxicab drivers, four restaurant servers (from two restaurants), and four hair stylists (from two salons) record the total bill size and the amount of the tip for each customer over several months. This amounted to nearly 1,000 service encounters.

INDUSTRY/SERVICE AND BILL SIZE: The author's regressed percentage tipped on the total amount of the bill for all bills less than \$100. The regression slopes were negative in each of the six cases (two taxicabs, two hair salons, and two restaurants), indicating a magnitude effect. Linear regression results for each of the six establishments demonstrate that as the total bill amounts get even larger, the slope of the regression line becomes less negative, approaching zero.

Greenberg, A. E. (2014). On the complementarity of prosocial norms: The case of restaurant tipping during the holidays. *Journal of Economic Behavior & Organization*, 97, 103–112.

DESIGN OVERVIEW: Data was pulled from all credit card transactions from a restaurant chain in upstate New York over the course of one year. All transactions required both a correct bill and tip amount, so that situations when no tip was left on the credit card were dropped from the analysis (because those situations likely included a cash tip since it was reported that instances of complete “stiffing” among credit card customers were quite rare).

For their analysis, the “holiday period” was determined to be the weeks prior and post-Christmas Day. Furthermore, other holiday days were added into the regression equation as a separate variable. Customers were restricted in the analysis to those who were observed as having dined at least once during the holidays and during the non-holiday period.

GEOGRAPHY: Forthcoming paper looking at whether prosocial behaviors (tipping behavior in general and generosity during the holidays) compete with one another, leading to no change in tipping behavior during the holidays, or whether they would complement one another such that people would tip at higher rates during the holidays. Overall findings were that people tipped higher during the holidays, but when the population was split, it was determined that this finding was skewed and that while bad tippers tipped better, “good” tippers tipped even more.

Findings were that tips during the holiday period were 3.7% higher than in the non-holiday period (24.3% overall).

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009–1055.

METHODOLOGY: Article discusses the use of field experiments in economic research. In contrast to traditional means of collecting data for the purpose of economic research—such as the use of naturally occurring data, where treatment and control status are not assigned at random, or laboratory experiments, where treatment status is randomly assigned but the setting is artificial—field experiments feature the use of randomly assigned treatment status but in a natural setting. They thus potentially allow the researcher to make causal inferences while simultaneously mitigating issues of external validity that are prevalent in laboratory experiments. The article briefly discusses findings from three types of field experiments that allow for varying degrees of external validity: artificial field experiments, where the subjects are aware of the experiment and the activity that they undertake does not directly correspond to naturally occurring activities, but where the subject pool represents a naturally occurring population of interest; frame field experiments, where, like artificial experiments, the subjects are aware that they are participating in an experiment but where the subject’s activity in the experiment more closely corresponds to naturally occurring phenomena; and natural field experiments, where the activity induced by the experiment is something the subjects would do naturally and they are simultaneously unaware that they are participating in an experiment, maximizing the chances that observed responses to the treatment would hold outside of the context of the experiment.

Hill, D. J., & King, M. F. (1993). An exploratory investigation into consumer knowledge of tipping etiquette: Accuracy, antecedents and consequences. In W. Darden & R. Lusch (Eds.), *Proceedings of the symposium on patronage behavior and retail strategy: Cutting edge III* (pp. 121–135).

DESIGN OVERVIEW: Sample was roughly 150 business majors ages 20 to 42 used for the analysis. They were asked to provide responses to what the appropriate tipping levels were for various services (not listed by the author, though the articles that they based these “correct” answers on were listed). They created a battery of tipping-related items and used a factor analysis to determine that there were five factors concerning tipping knowledge. Respondents were also asked a series of 27 developed questions on variables that were determined to influence tipping behavior from literature reviews and one-on-one interviews on this subject. The 27 questions were determined to have five useful factors: (1) social tipping orientation (their belief in the “social value of tipping”), (2) tipping experience, (3) tipping confidence (their belief that their knowledge of tipping behavior was correct), (4) tipping response (belief that poor service should receive poor tips, etc.), and (5) parental influence.

TIPPING KNOWLEDGE: Ultimately, most of the factors were not found to be correlated to correct tipping knowledge. The only two that were related were the parental influence (such that those who

learned more from their parents had more correct knowledge) and the age they first tipped (which also makes sense given that the earlier they started tipping, the more guidance they likely got from their parents and practice they have with tipping behavior).

Jargon, J. (2013, September 4). IRS rule leads restaurants to rethink automatic tips. *The Wall Street Journal*. Retrieved from

<http://online.wsj.com/news/articles/SB10001424127887323893004579055224175110910>.

SERVICE CHARGE: Article reporting on the change in how the IRS counts tips automatically added to the bill for large parties and the change that will occur starting in 2014. Under the new rules, restaurants will have to take those automatic tips and add it to the servers' actual wage at the end of the pay cycle and withhold taxes from it. This means that servers will have to wait for that money, as opposed to getting it at the end of the night, to ensure taxes are filed properly (which could mean less income for servers), and cause more paperwork and costs to restaurants to manage additional records.

This article was later cited by other websites, including NPR and the Consumerist (see Neuman, 2013; and Morran, 2013, citations).

Kerr, P. M., Domazlicky, B. R., Kerr, A. P., & Knittel, J. R. (2006). An objective measure of service and its effect on tipping. *The Journal of Economics*, 32(2), 61–69.

DESIGN OVERVIEW: Author investigated how service quality, measured by the amount of time it took to deliver the meal, influenced the tip size. Other variables included in the analysis were gender, race (White vs. all others), and income of the served location. Some information was added to the analysis based on census information, particularly the income variable. Two delivery drivers from the same restaurant measured all data in this study aside from "income," which was added based on census information on the location of the delivered food. The type of payment and the magnitude of the bill were also considered in the analysis.

However, it is worth noting that this article does not specify how many observations are being analyzed, or provide any information about the drivers other than state that the "personal attributes of the drivers were quite similar."

INCOME: Higher-income areas were more likely to leave better tips than lower-income areas.

GENDER: Males were found to tip marginally better than females.

CASH VERSUS CREDIT: Cash-paying customers were actually found to tip better than credit card customers, but this effect was nonsignificant when the magnitude of the bill was considered as part of the regression equation.

Klee, E. (2004). *How people pay: Evidence from grocery store data*. Federal Reserve Board.

Retrieved from http://www.newyorkfed.org/research/conference/2006/Econ_Payments/Klee_b.pdf

DESIGN OVERVIEW: Examination of household data from the Survey of Consumer Finances from 1995, 1998, and 2001. Findings indicate that the share of credit card and debit card usage has increased over the years, while the usage of checks has decreased. However, these market shares and usage rates will not apply to many tipping situations, and should only be considered for demographic groups that have credit or debit cards.

AGE: Credit card usage differed somewhat by age, such that very young heads of households and those over the age of 75 have lower credit card usage than other age groups, while debit card usage differed significantly. Debit card usage was highest among the youngest cohort and decreased as age increased.

INCOME: For both credit and debit cards, usage rates increased along with rising income brackets, indicating that more wealthy individuals are more likely to have credit and/or debit cards.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692.

METHODOLOGY: Article reports results from a field experiment conducted on Danish tax filers where tax filers were initially randomly assigned to one of two groups, where one group is subject to rigorous audits while the other is not. Subjects are then randomly assigned to three groups, where one group does not receive a notice of a future audit while the other two groups receive notices that they will be audited with different probabilities (50% or 100%). Subjects in different treatment groups are compared based on the difference in the amount of income that they report and baseline audit data, with income broken down into that income that is subject to third-party reporting (i.e., there are records kept by employers, etc., against which self-reported income can be checked) and income that is purely self-reported. The authors hypothesize that only self-reported income should be affected by past audits and the threats of future audits. Consistent with the hypothesis, the effect of the enforcement treatments on evasion is close to zero for income subject to third-party reports, but having been audited in the past and the prospect of future audits reduces evasion for self-reported income. Evasion was generally substantially higher for self-reported income. Higher marginal tax rates were found to increase evasion, though the effect was relatively small. The authors argue that the results support the importance of enforcement through third-party reporting in explaining why compliance is generally high in developing countries despite low audit probabilities and fines.

Koku, P. S. (2005). Is there a difference in tipping in restaurant versus non-restaurant service encounters, and do ethnicity and gender matter? *Journal of Services Marketing*, 19(7), 445–452.

DESIGN OVERVIEW: Thirty-five participants were randomly selected for seven different service sector businesses (245 total participants) that they indicated they had patronized within the past three months. Service sector business included restaurants, barbershops/hair salons, spas, golf club shops, auto detailing shops, auto mechanics' shops, and valet parking. Participants were provided a questionnaire that asked them if they tipped 15% or more of the total bill, less than 15%, or did not tip at all. They were also given a space to provide a reason for their tipping decision.

INDUSTRY/SERVICE: For analysis purposes, the researchers combined all non-restaurant services to compare against restaurant tipping. They also combined all those who said they tipped less than 15% and those who did not tip at all. Using a chi-square test, the researchers determined that there is a difference between the reasons people tip in the restaurant industry and outside of it.

RACE/ETHNICITY: The researchers also compared White versus non-White respondents on tipping tendencies outside the restaurant industry, and failed to find any difference.

GENDER: They only found a marginal difference between men and women in tipping outside the restaurant industry.

Koku, P. S. (2007). Some significant factors that influence tipping in service encounters outside the restaurant industry in the United States. *Services Marketing Quarterly*, 29(1), 23–45.

DESIGN OVERVIEW: The sample included 12 MBA students (6 male, 6 female) who indicated that they had used another service-sector business in addition to the restaurant industry in the past 3 months. Other service-sector businesses included spas/body massage, barbershop/hair salons, auto mechanics' shops, plumbing services, auto detailing shops, valet parking, and lawn care services. There were two sessions. All participants met in the first session for two hours and were asked about service encounters in which they tipped in the past month and what led them to do so, as well as service encounters in which they did not tip and why. The second session included 30-minute individual sessions.

INDUSTRY/SERVICE: Using the framework of transaction cost analysis (TCA), the authors propose several factors that influence a consumer's tip in other service-sector businesses (i.e., service industries other than restaurants). From information gleaned in interviews, the authors propose that the customer's decision to tip is influenced by (1) quality of service, (2) the length of time to be served or have his or her issue resolved in an emergency situation, (3) the likelihood of repeat purchase (which is influenced by service quality), and (4) budgetary constraints.

Lynn, M. (1988). The effects of alcohol consumption on restaurant tipping. *Personality and Social Psychology Bulletin*, 14(1), 87–91.

DESIGN OVERVIEW: The author became employed as a waiter at the restaurant where the study took place. For just over a month, he recorded information for 207 dining parties, including bill size, tip amount, whether alcohol was consumed and number of drinks, customer's gender, and payment method.

BILL SIZE: A regression of tip amount on bill size indicated that tipping is strongly, positively related to bill size. The resulting equation found a y-intercept of .32 (32 cents) with an additional tip of 11% of bill size; this accounted for 50% of the variance in tip amount.

INDUSTRY/SERVICE: After controlling for the relationship between bill size and tip amount and a host of other variables, a hierarchical multiple regression found a significant effect for alcohol. The results indicate that alcohol (but not number of drinks) consumption increases tipping.

Lynn, M. (2004). Black-White differences in tipping of various service providers. *Journal of Applied Social Psychology*, 34(11), 2261–2271.

DESIGN OVERVIEW: A randomized telephone-based survey was conducted to determine the difference in tipping behavior among various service industries. This data was acquired by Lynn in order to conduct follow-up analysis regarding tipping differences between Whites and Blacks. Waiters, bartenders, barbers, taxi drivers, food-delivery people, hotel maids, masseuses, bellhops, and ushers at theatres or sporting events were the occupations of interest. In the final analysis, 894 respondents (811 White and 83 Black respondents) were used. Respondents were asked, "If you received good service from ____ would you tip them a percent of the total cost of the service, tip them a flat amount, or not give them a tip?" Respondents were asked this question nine times for different service industries: waiter or waitress; bartender; barber, hair stylist, or cosmetician; cab or limousine driver; food-delivery person; hotel maid; skycap or bellhop; masseuse; and usher at theater, sporting events, etc. Respondents were then further questioned about the amount they would tip if they indicated that they would tip a percentage or flat amount.

INDUSTRY/SERVICE: Waiters received the most tips among Whites, though barbers also had a high tip percentage amount among both Whites and Blacks.

RACE/ETHNICITY: Blacks are less likely to base restaurant tips on bill size than are Whites. Black percentage tippers leave a smaller average percentage of the bill than do White percentage tippers across many service contexts. Finally, Black flat tippers leave larger average dollar tips than do White flat tippers across many service contexts (e.g., bartenders, barbers, hotel maids, and masseuses).

Lynn, M. (2004). **Ethnic differences in tipping: A matter of familiarity with tipping norms.** *Cornell Hospitality Quarterly*, 45(1), 12–22.

DESIGN OVERVIEW: The survey results were the same as reported in Lynn's 2004 article on tipping knowledge among various racial groups in which respondents were collected by random-digit-dialing (RDD) telephone methods. Respondents were asked how much it was customary to tip waiters and waitresses in the United States with "15% to 20%" considered to be the right answer. Roughly 1,000 total completes were gained, but only 99 were from Black respondents. It is also important to note that respondents were asked about customary practices rather than their own tipping behavior.

RACE/ETHNICITY: Overall, most Whites (over 70%) indicated that they knew the correct amount to tip a waiter or waitress, compared with only 37.4% of Black respondents. Furthermore, 12.1% of Black respondents reported that they did not know the correct amount compared with only 2.4% of White respondents.

Lynn, M. (2006). **Geodemographic differences in knowledge about the restaurant tipping norm.** *Journal of Applied Social Psychology*, 36(3), 740–750.

DESIGN OVERVIEW: A phone survey was conducted by Taylor Nelson Sofres using random-digit-dial sampling for a total sample of slightly over 1,000 respondents. The primary question of interest was how much people are expected to tip waiters and waitresses in comparison to how much they typically tip. The "correct" response was considered to be 15% to 20%.

RACE/ETHNICITY: Significantly more Whites (72%) have the correct knowledge of tipping conventions compared with Hispanics and Blacks (33% of both). These effects were still significant once other variables were controlled for.

AGE: Age was initially significant, such that respondents in their 40s to 60s had higher levels of knowledge compared with older and younger respondents, but once other factors such as race, sex, education, income, metro status, and region were controlled for, it became insignificant.

GEOGRAPHY: Metro status was marginally significant before controlling for other variables but non-significant after control variables were considered. That said, the Northeast region had higher levels of tipping knowledge compared with the South region, but there were no other significant differences between other regions.

INCOME: Higher levels of income were related to higher levels of knowledge of correct tipping norms even when controlling for other variables.

EDUCATION: Higher levels of education were related to higher levels of knowledge of correct tipping norms even when controlling for other variables.

GENDER: Knowledge did not vary by sex when directly compared with men, but when other variable were controlled for, it was found that women had a significantly higher level of tipping knowledge than men.

Lynn, M. (2006). Tipping in restaurants and around the globe: An interdisciplinary review. In M. Altman (Ed.), *Handbook of Contemporary Behavioral Economics: Foundations and Developments*, (pp. 626–643). M. E. Sharpe Publishers.

DESIGN OVERVIEW: Lynn examines results from the literature for anything related to tipping. This includes the determinants of restaurant tipping, including bill size, payment method, gender, and race/ethnicity. A meta-analytic review conducted by Lynn and McCall (2000) found that 69% of the variability of dollar tip amounts within a restaurant can be explained by the bill size. Several studies' results support a "magnitude effect" where dollar tip amount increases with bill size, but percentage tip decreases. Several studies have demonstrated that patrons paying by credit card tend to leave a larger tip than those paying with cash (Feinberg, 1986; Garrity & Degelman, 1990; Lynn & Latane, 1984). Furthermore, the presence of a credit card company insignia induces higher tip amounts (McCall & Belmont, 1996). There has been some support for men leaving larger tips than women and waitresses receiving larger tips than waiters. Results indicate that patrons are more likely to provide a higher tip for a server of the opposite sex. Black restaurant patrons are more likely than their White counterparts to tip a flat amount rather than a percentage and tip a lower percentage. Studies have shown these results even when controlling for education, income, and perceptions of service quality (Lynn & Thomas-Haysbert, 2003).

Lynn, M. (2011). Race differences in tipping: Testing the role of norm familiarity. *Cornell Hospitality Quarterly*, 51(1), 73-80.

DESIGN OVERVIEW: This study was a web-based survey from a consumer panel (Zoomerang.com) in which the aim was to test and determine if tipping knowledge mediates the relationship between race/ethnicity and tipping behavior, because no work up to this point had tested if this relationship existed. Multiple waves of invitations were sent until the desired demographics groups were gathered (100 respondents from both White and Black races, and with a separate split of those with and without a college education, 831 total observations in all).

As with previous studies, respondents were asked how much people in the United States are generally expected to tip waiters and waitresses, with 15% to 20% being considered the correct answer. Later in the survey, they were also asked about their tipping behavior for waiters and

waitresses that gave them good service in order to determine not just their knowledge about tipping behavior, but also their own behavior as well.

For other industries and services, such as hotel maids and bartenders, the respondent was simply asked if that industry was generally tipped at all. Respondents who indicated that the various other services were tipped were considered as having some knowledge of the norm for that occupation. As with waiters and waitresses, respondents were further asked how often they tipped members of the other professions.

RACE/ETHNICITY: Analyses indicate that tipping norm awareness did predict racial differences between Black and White tipping behavior for restaurant tips, both for tip type (whether a percentage of the bill was left versus a flat amount) and the percentage left.

No racial differences were found in the tipping/stiffing of hotel maids and luggage handlers, but racial differences were found for the other investigated services.

Finally, a moderated relationship for norm awareness was also tested for, but this was not found to be statistically significant.

INDUSTRY/SERVICE: A few significant differences were found for other professions. Specifically, they found that norm awareness mediated racial differences in stiffing behaviors for haircutters and pizza delivery, but not for bartenders, parking valets, or cab drivers.

Lynn, M. (2012). The contribution of norm familiarity to race differences in tipping: A replication and extension. *Journal of Hospitality & Tourism Research*. Advance online publication.

doi:10.1177/1096348012451463

DESIGN OVERVIEW: Web-based survey was sent out to members of a consumer survey panel. Response rates were not calculated as probability of panel selection is not captured. The final sample included 180 respondents after s cleaning the original data set for outlier responses (such as suggesting they gave tips over 100% of their bill) or improbable completion times...

Respondents were asked how much they would tip for one of two randomly assigned bill amounts, \$21.32 or \$46.23, if the service was determined to be unusually good, average, or unusually bad. Finally, respondents were also asked how much people in the United States are expected to tip a waiter for adequate to good service and given typical response options.

RACE/ETHNICITY: Controlling for age, sex, income, education, and bill size, Black and Hispanics were found to tip less, and were also less aware of the standard 15% to 20% tipping norm. Furthermore, it was found that tipping knowledge significantly affected tip size after controlling for race, indicating a partially mediated relationship.

Lynn, M. (2013). A comparison of Asians', Hispanics', and Whites' restaurant tipping. *Journal of Applied Social Psychology, 43*(4), 834–839.

DESIGN OVERVIEW: An online survey was conducted via a large multistate restaurant, yielding 1,274 final observations after 64 subjects who refused the race/ethnicity question were dropped from the analysis. The survey asked respondents about service and restaurant quality in addition to the size of their bill and tip size. Service quality was used as a control when observing the differences between the different racial groups.

This study asked respondent race/ethnicity as a single-item question, as opposed to how the U.S. Census Bureau asks two questions, one for race and one for ethnicity. In this setup, respondents could indicate that they were Hispanic or Black, but not both.

BILL SIZE: Flat-dollar tips increased along with bill size while percent tips decreased in the same span.

RACE/ETHNICITY: Hispanics tip significantly less than Whites but there are no differences between Asians and Whites. However, given the relatively low *N* of the Asian population (roughly 75 observations) the findings have to be taken with caution.

Lynn, M., & Gregor, R. (2001). Tipping and service: The case of hotel bellmen. *International Journal of Hospitality Management, 20*, 299–303.

DESIGN OVERVIEW: A hotel bellman interacted with 50 different customers while delivering one of two conditions of level of service, at a small luxury hotel. In the “limited” service condition, the bellman met customers at their cars with a cart and loaded their bags and then accompanied them to their hotel room after they checked in, opened the door, and brought the luggage to their room. They then asked guests if there was anything else they needed before collecting any tips and leaving the room. The “full” service condition included the same treatment as the “limited” condition, but the bellman also demonstrated how to use the television and thermostat, opened the blinds, and offered to get ice for the guest. The bellman recorded the guests' experimental condition, sex, apparent age, and tip following each interaction.

INDUSTRY/SERVICE: The hotel bellman received significantly higher tips for providing the “full” service condition (\$4.77) than the “limited” service condition (\$2.40). The effect of increases in tips based on service condition was similar among men, women, younger guests, and older guests.

Lynn, M., & Latane, B. (1984). The psychology of restaurant tipping. *Journal of Applied Social Psychology, 14*(6), 549–561.

DESIGN OVERVIEW: In the first study, 169 groups of customers were interviewed as they exited an IHOP. Only those who paid the bill were questioned, or if two or more paid the bill, their responses were combined. Participants were questioned about party size, restaurant atmosphere, food quality, service quality, bill size, tip size, and improvements for the restaurant; respondent gender was also recorded. All servers were female.

In the second study, 4 waiters and 5 waitresses collected data for 206 dining groups over a 1-week period. They recorded the number of people on the check, number of people at the table, number of checks at the table, bill size, gender of person(s) paying the check(s), method of payment, amount left as a tip, and server's level of effort spent serving the table. They recorded this information for parties of five people or less or larger parties without a reservation because of the automatic gratuity applied to larger parties.

BILL SIZE: In the first study, the average bill size was \$3.16 and the average tip per person was \$.42. Customers tipped an average of 15.6% of their bill size. A hierarchical, multiple linear regression of customer's gender, party size, number of separate checks, atmosphere, service, food ratings, and per-person bill size on percent tipped was performed. After controlling for other variables, per-person bill size predicted a significant amount of variance in percent tipped. The larger the per-person bill size, the smaller the percentage tip of the total check. In the second study, the average bill size per person was \$13.01 and the average tip per person was \$2.01. Customers tipped an average of 15.5% of their bill size. In the hierarchical, multiple linear regression, per-person bill size was unrelated to percent tip, which the authors speculate is due to the high price of the restaurant where the study was conducted compared with that of a café in the first study where some groups only ordered coffee or a snack.

PAYMENT METHOD: In the second study, a hierarchical, multiple linear regression of customer's gender, server's gender, party size, number of separate checks, effort ratings, per-person bill size, and payment method on percent tipped was performed. After controlling for other variables, payment method predicted a significant amount of variance in percent tipped. Customers paying their checks with credit cards tipped a larger percentage of the bill than cash-paying customers (16.9% versus 14.5%).

GENDER: In the first study, using the same hierarchical, multiple linear regression of customer's gender, party size, number of separate checks, atmosphere, service, food ratings, and per-person bill size on percent tipped, after controlling for other variables, gender predicted a significant amount of variance in percent tipped. Men tipped significantly more than women (17.4% versus 9.5%). For the second study, in the hierarchical, multiple linear regression of customer's gender, server's gender, party size, number of separate checks, effort ratings, per-person bill size, and payment method on

percent tipped, after controlling for other variables, customer's gender also predicted a significant amount of variance in percent tipped. Men tipped slightly more than women (15.7% versus 14.6%).

Lynn, M. & McCall, M. (2000). Gratitude and gratuity: A meta-analysis of research on the service-tipping relationship. *The Journal of Socio-Economics*, 29(2), 203–214.

DESIGN OVERVIEW: Meta-analysis conducted on a combination of published and unpublished studies that had variables concerning tipping behavior and service quality, yielding observations for 2,547 dining parties across 20 different restaurants. The unit of analysis used was the *N* of restaurants, as the authors argue that as tipping expectations and norms can vary by establishment, that is the most appropriate level for analysis. Some splits were done to determine the relationship between service quality and tipping behavior based on the metric used in the analysis and the person providing the data, as some of the relationships were based upon a server's estimation of the service quality rather than the customer's.

Ultimately, it was determined that there was a significant relationship between service quality and tips, but that it accounted for less than 2% of variance in tipping behavior. This value was stronger (almost 5%) among studies that had stronger measures of service quality. However, there was no such relationship found for measures that recorded the perceptions of servers, indicating that servers do not see a link between service quality and tipping behavior.

Lynn, M., & McCall, M. (2000). *Beyond gratitude and gratuity: A meta-analytic review of the predictors of restaurant tipping*. Working paper, School of Hotel Administration, Cornell University.

DESIGN OVERVIEW: The authors limited the meta-analysis to research concerned with the restaurant industry where the data were collected about an individual service encounter from one of three or more modes: (1) restaurant checks, charge receipts, and comment cards; (2) records kept each evening by restaurant servers; and/or (3) interviews with patrons as they departed restaurants. A total of 22 published studies and 14 unpublished studies were included in the meta-analysis. The authors meta-analyzed the relationships of tip size to bill size and of bill-adjusted tip size to 23 predictors from the tipping literature, including weather, payment method, and alcohol consumption.

BILL SIZE: The meta-analysis indicated that tip amounts were positively related to bill size. In fact, the authors found that bill size accounted for about two-thirds of the variability in tip amounts.

GEOGRAPHY: Meta-analysis results indicate that patrons left larger bill-adjusted tips when the weather was sunny.

CASH VERSUS CREDIT: Patrons left larger bill-adjusted tips when they used a credit card as their method of payment or when they received their bill on a tip tray embossed with a credit card company's insignia.

INDUSTRY/SERVICE: Alcohol consumption was not related to bill-adjusted tips.

Lynn, M., & Thomas-Haysbert, C. D. (2003). Ethnic differences in tipping: Evidence, explanations, and implications. *Journal of Applied Social Psychology, 33*(8), 1747–1772.

DESIGN OVERVIEW: A pair of studies were conducted to investigate racial differences in tipping. The first study was based on the data from the 1997 Speer article. The second study was based on a collection of data sets based on five tipping articles that either interviewed customers after they had left their restaurant or the servers after the customers had had their meal.

The first study used the data from Speer (1997), with an *N* of about 1,000 from a telephone survey and about 100 Black respondents. The combination of data sets in the second study resulted in an *N* of about 1,800 respondents, with 94 Black respondents, 149 Asian respondents, and 113 Hispanic respondents. All the restaurants in the five studies used in the second study came from in or near Houston, Texas.

RACE/ETHNICITY: The first study showed the same results as in previous studies in that Blacks tipped less than Whites, but additional mediating analyses were conducted. Age, income, education, and tip size were all found to be partial mediators of the race/ethnicity relationship.

The second study found that Whites left significantly higher tip sizes compared with both Blacks and Asians, but not Hispanics. Another finding of note was that Asians and Hispanics were more likely to tie the percent tip to service quality than Whites and Blacks.

Lynn, M., & Williams, J. (2012). Black-White differences in beliefs about the U.S. restaurant tipping norm: Moderated by socio-economic status? *International Journal of Hospitality Management, 31*(3), 1033–1035.

DESIGN OVERVIEW: A pair of phone surveys were used for the analysis that used separate, but very similar, questions. One survey asked, "Thinking about tipping overall, not your own practices, how much is it customary for people in U.S. to tip waiters and waitresses?" The other survey asked, "Thinking about restaurant tipping norms, how much are people in the U.S. expected to tip waiters and waitresses?" Both questions were open-ended and results were coded into predetermined response options, such as "15%–20%." Tipping knowledge was considered to be either partial (in

terms of knowing that it was customary to tip waiters and waitresses) or complete (that it was customary and that 15% to 20% was the correct amount).

A measure of Socio-Economic Status (SES) was crafted based on a pair of questions, one asking income and the other asking education background. These were standardized and then averaged together to form one scale.

RACE/ETHNICITY: The significant difference between White and Black tipping knowledge was mediated by SES for partial tipping knowledge but not for complete tipping knowledge. This would seem to indicate that all low SES individuals in general are unaware that tipping is customary in certain situations, but that the “correct” tipping amount is not influenced by SES, and still seems to involve a racial component.

Lynn, M., Zinkhan, G., & Harris, J. (1993). Consumer tipping: A cross-country study. *Journal of Consumer Research*, 20, 478–488.

DESIGN OVERVIEW: The authors used information about tipping in 33 service professions across 30 different countries (Star, 1988). Each service in each country was coded as either “tipped” or “not tipped” and aggregated. The authors obtained data for four different work-related indices that they posit are related to tipping differences across countries. The study used 116,000 questionnaire responses from industrial corporation employees in 50 different countries (Hofstede, 1983). The four indices are “power distance,” where a high score reflects an acceptance for hierarchical structure and a low score reflects the opposite; “uncertainty avoidance,” where a high score reflects a culture that is concerned with following the rules and a low score reflects one that is willing to take risks; “individualism,” where a high score is associated with a culture that is concerned with individuals’ independence and a low score reflects a culture of collectivism; and “masculinity” reflects a culture whose values are primarily masculine.

GEOGRAPHY: There was a correlation of .46 between the power distance index and number of services that get tipped, indicating a strong relationship between high power distance scores and the number of services that are tipped. There was a correlation of .55 for uncertainty avoidance and tipping, indicating that tipping occurred more often in countries that were less tolerant of uncertainty. There was a correlation of -.39 between the individualism index score and tipping, indicating that tipping was more common in collectivistic countries. There was a correlation of .47 for masculinity index and tipping, indicating that tipping occurred more often in countries with masculine values. Japan was an outlier in all four analyses and was omitted.

McCall, M., & Belmont, H. J. (1996). Credit card insignia and restaurant tipping: Evidence for an associative link. *Journal of Applied Psychology*, 81(5), 609–613.

DESIGN OVERVIEW: For the first experiment, data were collected from 77 paying customers at a family restaurant; men were most frequently the paying customer (59 men and 18 women). Patrons tended to be people vacationing at a nearby ski resort. The independent variable was what type of tip tray the diner received with the check, either a blank tip tray or a tip tray with the credit card insignia of a major credit card company in the center of the tray. Servers recorded the amount of the bill, number of patrons in the dining party, the sex of the individual paying the bill, the method of payment, and the total amount tipped.

For the second experiment, data were collected from 27 paying customers from a café in a separate town from Experiment 1, whose main clientele is university students. The sample included 13 men and 12 women, and two missing cases where gender was not recorded. The methodology of Experiment 2 replicated Experiment 1 except that the credit card insignia on the tip trays was from a different credit card company.

CASH VERSUS CREDIT: In the first experiment, an analysis of covariance (ANCOVA) revealed that a credit cue significantly affected percentage tipped. Specifically, individuals that were given the tip tray with a credit card insignia tipped a significantly higher percentage (19.77%) than those who received a blank tip tray (15.48%).

In the second experiment, all paying customers used cash. Data were analyzed the same way for Experiment 2 as they were in Experiment 1. Similar to Experiment 1, the ANCOVA demonstrated a significant effect of credit cue on percentage of the bill tipped, where the presence of a credit card insignia resulted in a tip percentage of 21.91% compared with those who received a blank tip tray (17.53%). While the following two experiments did not compare tipping by method of payment used, these were the basis for a lot of method-of-payment research in the future.

McCrohan, K. F., & Pearl, R. B. (1983, August). *Tipping practices of American households: Consumer based estimates for 1979*. 1983 Program and Abstracts: Joint Statistical Meetings, Toronto, CA.

DESIGN OVERVIEW: Diary population was recruited via telephone recruitment and auto-registration listings, creating an estimate of \$5.7 billion in tipping revenue. Demographic targets were based on census data. Two samples were used: 10,000 family households and an additional 1,500 nonfamily households. The sample populations were recruited via telephone recruitment and auto-registration listings. Reports were given on a quarterly basis over the course of the entire year. Families reported over a two-week span every quarter and were staggered such that there were diaries coming in from

some of the sample every week. However, the nonfamily sample only reported during one quarter in the entire year.

NATIONAL AVERAGE TIPPING BEHAVIOR: Of the \$72.7 billion that was spent on dining out in 1979, 31% was considered to be spent on tipping occasions and such occasions accounted for over half of all revenue. Of this revenue, tipping behavior constituted \$5.7 billion, or roughly 14.4% of tipping occasion behaviors.

After examining the data and determining what types of establishments should be classified as “tipping occasions,” they determined that the true stiffing rate was somewhere around 20%, though that included some situations where people ordered hasty meals or snacks.

GEOGRAPHY: Findings indicate that tipping was higher in the northeast region of the country compared with the middle parts of the country and that metro areas tipped at higher rates.

INCOME: Very small differences were found relating to income, such that the highest income group tipped at about 1% greater rate than the lowest income group.

CASH VERSUS CREDIT: Credit transactions tipped at a somewhat higher rate than cash transactions (1% difference), but at this point they were only used in less than 3% of all dining transactions.

McCrohan, K. F., & Pearl, R. B. (1991). An application of commercial panel data for public policy research: Estimates of tip earnings. *Journal of Economic and Social Measurement*, 17, 217–231.

DESIGN OVERVIEW: Authors expand on the analysis of the consumer diary data discussed in Pearl and McCrohan (1984). The diary panel of restaurant patrons now includes the years 1982, 1983, and 1984. The authors find that tipping occurs in only 29% of eating occasions, but that tipping occasions account for approximately half of all expenditures. A regression analysis was also undertaken to examine the determinants of the tipping rate (tip amount over total expenditure, for both tipping and non-tipping occasions) for a given occasion.

NATIONAL AVERAGE TIPPING RATES: Across all periods, tip rates averaged approximately 14.4% and that the average was relatively invariant across the types of eating establishments (inside, outside, or non-tipping), though stiffing behavior varied by type, with tipping type restaurants (family, atmosphere, and coffee shop) accounting for 90% of all tips.

CASH VERSUS CREDIT: Findings from the regression analysis indicate that tipping rates are higher when establishments accept credit cards.

INDUSTRY/SERVICE: Findings from the regression analysis indicate that tipping rates are higher when establishments serve alcohol.

GEOGRAPHY: Findings from the regression analysis indicate that tipping rates are higher when establishments are located in metropolitan areas.

Morran, C. (2013, September 5). Are these the final days of automatic 18% tips at restaurants? *Consumerist*. Retrieved from <http://consumerist.com/2013/09/05/are-these-the-final-days-of-automatic-18-tips-at-restaurants/>.

SERVICE CHARGE: Report on the change in how IRS considers the automatic 15% to 20% gratuity in restaurants, citing the piece by Jargon (2013) in *The Wall Street Journal*. Darden Restaurants, parent company of Olive Garden, Red Lobster, and LongHorn Steakhouse, has already reported that it was going to drop the automatic gratuity policy because of this issue.

Neuman, S. (2013, September 5). IRS to count automatic gratuities as wages, not tips. *NPR*. Retrieved from <http://www.npr.org/blogs/thetwo-way/2013/09/05/219290573/irs-to-count-automatic-gratuities-as-wages-not-tips>.

SERVICE CHARGE: Blog post on the IRS's change in how automatic gratuities are counted. The blog post covers an original *Wall Street Journal* article on this issue (see Jargon, 2013, for original report).

Noll, E., & Arnold, S. (2004). Racial differences in tipping: Evidence from the field, *Cornell Hospitality Quarterly*, 45, 23–29.

DESIGN OVERVIEW: Two unpublished studies, both of which were reported by servers from a large restaurant chain, were used. In the first study, approximately 100 servers were asked a variety of questions regarding supposed “tip predictors” such as race, alcohol use, and gender. The second study aimed to investigate whether servers were accurately reporting their tip sizes as that misreporting could significantly damage the results that were found in the first experiment. Two servers in the same restaurant chain (but in another state) agreed to note their tips over a two-week period. Overall, tips were recorded from 151 sets of customers.

RACE/ETHNICITY: Nearly all of the servers in the first study reported that they were aware of the differences in tipping by race. Three-quarters of the servers indicated that their Black customers were less likely to provide a tip, and when a tip was provided, more likely to tip below 15% than White customers. In the second study, the two reporting servers reported similar findings for differences between White and Black customers (though it is worth noting that outliers of tips over 26% were removed for both White and Black customers prior to analysis).

GENDER: In the first study, it was also found that male customers tipped more than female customers.

INDUSTRY/SERVICE: In the first study, it was reported that customers who consumed alcohol gave significantly higher tips than those who did not consume alcohol.

CREDIT/CASH: In the first study, servers reported that credit card customers tipped significantly more than customers who paid with cash. However, in the second study it was reported that customers who paid with cash gave marginally higher tips than those who paid with credit cards.

Papp, T. G., & Burkhammer, A. L. (2001, March). *An investigation of server posture and gender on restaurant tipping*. Paper presented at the 22nd Annual Industrial Organizational Psychology and Organizational Behavior Graduate Student Conference, Pennsylvania State University.

DESIGN OVERVIEW: Servers were recruited and asked to record information for 10 different dinners, alternating between squatting and standing in order to determine how this changed tipping behavior. Servers recorded bill size, tip size, and gender of the diners. Servers were recruited from campus and by sending out survey packets to various restaurants in the area, yielding a final sample of 107 observations across 12 different servers. Eight of the final servers were female and four were male. Each server was instructed to record five meals squatting and five meals standing, and to only record this for small dinners or two or fewer diners.

GENDER: The only effect found was a marginally significant difference between male and female servers such that male servers had more tips, but this was the only effect that was found.

Parker, J. A., Souleles, N. S., & Carroll, C. D. (2012). *The benefits of panel data in consumer expenditure surveys*. *National Bureau of Economic Research*.

METHODOLOGY: Article reviews the benefits of the panel nature of the Consume Expenditure (CE) Survey. The authors argue that repeating questions for individual respondents increases response accuracy by increasing familiarity and understanding of the survey. In addition, repeat interviews allow for respondents and interviewers to check the consistency of the responses, thus further mitigating measurement error. On the other hand, requiring repeat interviews increases the burden on respondents and thus potentially increases sample attrition and thus selection bias, though the authors argue that there is little evidence that those who drop out of the sample are different in a way that would influence expenditure. Repeat measures also help reduce noise in individual respondent expenditures that could result from irregular expenditures taking place in individual interview periods or measurement error that results from using long recall periods. When modeling expenditure, panel data allows researchers to control for individual-level unobserved fixed effects, allowing the researcher to potentially make causal inferences concerning the effect of some time-varying factors on individual expenditure. Controlling for unobserved individual fixed effects may also reduce variability in estimated effect sizes, increasing the precision of estimates. Panel data also allows the researcher to assess the dynamics of expenditure for a given household.

Paul, P., & Gardyn, R. (2001). The tricky topic of tipping. *American Demographics*, 23(5), 10–11.

DESIGN OVERVIEW: The article used the same data source that was mentioned in the Lynn piece on differences between Blacks and Whites among various service types (2004). Roughly 900 total phone numbers were randomly called to get the survey population. The professions that were listed in the article were waiters, bartenders, barbers, taxi drivers, food delivery workers, hotel maids, skycaps or bellhops, masseuses, and ushers at theater or sporting events.

INDUSTRY/SERVICE: Waiters were tipped far more often on a percentage basis than all other listed professions (74% were tipped a percentage compared with 22% who got a flat tip), and were also tipped the highest amount when tipped by percentage (along with barbers, both at 17%). Of all other professions, the percentage of respondents who said they were tipped a percentage was much lower than that for waiters, ranging between 5% for ushers to 31% for taxi drivers and food delivery workers.

NATIONAL AVERAGE TIPPING RATE: Waiters were also stiffed the least of all the professions, with only 2% reporting stiffing behaviors. Of the other professions, only masseuses (25%), hotel maids (26%), and ushers (70%) were stiffed at rates greater than 20%, while bellhops were stiffed the least of the other professions at 10%.

GEOGRAPHY: Various regional differences were discussed, such as respondents from the Northeast region gave higher tips to waitstaff and busboys (16% to 20%, respectively) compared with other regions, but they tipped cab drivers less than other regions (21% only gave a dollar or less for cab rides compared with 13% from the rest of the country).

Pearl, R. B. (1984). *A survey approach to estimating the tipping practices of consumers*. Special report on regression analysis to the Internal Revenue Service under contract TIR-81-21, Survey Research Laboratory, University of Illinois, Champaign, IL.

DESIGN OVERVIEW: Special analysis of the 1982 data using regression. Analyses were run using both a weighted and unweighted approach in order to examine both the propensity to tip and the tipping percentage on occasions where a tip was left. Regressions using scaled weights produced somewhat better regressions and were used in the final analysis. These analyses produced R^2 values of .20 for tipping behavior, but only .13 for regressions related to the actual tipping rate. Propensity to tip was mostly predicted by whether it was for full-scale restaurants or for snack places.

GEOGRAPHY: Metro areas tipped at higher levels than nonurban areas.

Pearl, R. B., & McCrohan, K. F. (1984). Estimates of tip income in eating places, 1982. *Statistics of Income Bulletin*, 3(4), 49–53.

METHODOLOGY: Authors attempt to improve upon prior attempts at estimating tipping income for restaurants through the analysis of a large ($N = 10,000$ households of two or more related persons + 2,800 households of one or two unrelated persons) diary panel of restaurant patrons for 1982. Respondents kept a diary where they recorded information about all eating occasions over the course of a two-week period in a given quarter. The large sample (weighted to be representative of the U.S. population in the given years) allowed for the more precise estimates, while querying customers rather than employees or managers of establishments on tipping behavior mitigated bias that may have resulted from the incentive of employees to underreport tipping income or managers to exaggerate tipping income in order to justify subminimum wages. The authors argue that the use of a diary as opposed to a survey increases the accuracy of the information provided, because details of dining occasions are recorded closer to the time of the meal. In addition, they maintain that the use of a diary lowers the probability that respondents will exaggerate the size of the tip in order to impress the interviewer.

NATIONAL AVERAGE TIPPING RATES: The results of their analysis of data from the diary imply that tips comprised approximately 7.4% of all expenditures and 14.3% of all expenditures on meals where tipping actually occurred.

INDUSTRY/SERVICE: Respondents were asked to categorize establishments in six types (family, atmosphere/specialty, coffee shop, cafeteria, fast-food and drive-in, and take-out) where the first three categories were classified by the authors as “tipping establishments.” Within the tipping establishments, sit-down and specialty establishments received tips on 60% of occasions. Within this group, tips made up 12.9% of all expenditures and 14.5% of all expenditures on occasions where a tip was actually given.

Pearl, R. B., & Sudman, S. (1983). *A survey approach to estimating the tipping practices of consumers*. Final report to the Internal Revenue Service under Contract TIR-81-21, Survey Research Laboratory, University of Illinois.

DESIGN OVERVIEW: Methodology was very similar to the previous report on 1979 tipping behavior that was conducted by NPD, with a sample of 10,000 families and an additional 2,800 households containing one or two unmarried people. The study was updated to include tipping behavior not only in restaurant situations, but also in other industries, including bars, hotels, barbershops, and taxi services. In this case, each household maintained records of tipping behavior at eating places during a one-week period each quarter, with half of the sample doing this in addition to a supplementary

two-week diary study over the course of two quarters that covered additional services that might get tipped (over 50 other industries were identified as having been tipped, but four of them accounted for 80% of such situations and were the primary focus in the report). They were also asked to provide some brief information about the type of establishment that they ate at to determine whether it was a situation that tipping was expected in order to determine a true stiffing rate.

In order to determine if there were sources of bias in the data, an additional phone survey was conducted with 935 households during the summer months to validate the data that was being obtained via the diary studies. The validation study reported somewhat lower tipping rates for each service, but they were within sampling error and might be due to the change in methodology between a recall-question telephone survey and a diary survey.

AVERAGE TIPPING BEHAVIOR: In restaurants, the tipping rate was 14.3% overall, though only one-fifth of responses came within the 14%–16% band, one-fifth of responses exceeded 20%, and another one-seventh of responses reported less than 10%. Tipping rates also decreased along with increasing household size.

The true stiffing rate was determined to be similar to levels reported in 1979, in that roughly 21.2% of tipping situations for restaurants were stiffed and about 10% of expenditures. As noted in the prior study, it is impossible to determine which purchase included snacks and small items that might not be considered to be tip-worthy. Stiffing rates were the lowest for credit card purchases.

GEOGRAPHY: The overall tipping rate was found to be somewhat higher in the Northeast region of the country and in metro areas.

INCOME: As noted in the previous study, as income levels increased, the tipping rate also increased somewhat with greater income, but not to a large degree.

CASH VERSUS CREDIT: Credit card users gave higher tips (14.9%) than cash users (14.3%).

INDUSTRY/SERVICE: The tipping estimates that were reported for other industries, notably bars, differed substantially from independent reports. Bars and taxi services reported receiving tips of 19% to 20% overall, while barbers received 11.6%. The average tip at hotels and motels could not be accurately assessed based on percentages, and the average tip amount was \$1.89, though this amount was still higher than that reported for the other services overall.

Stiffing rates are very difficult to assess for these other noted industries because hotels might be considered to be “stiffed” even if it was simply a one-night stop at a motel, as 70% of hotel instances did not get a tip.

Pearl, R. B., & Vidmar, J. (1988). *Tipping practices of American households in restaurants and other eating places: 1985–86*. Supplementary report to the Internal Revenue Service under Contract TIR 86-279, Survey Research Laboratory, University of Illinois, Champaign, IL.

DESIGN OVERVIEW: Report on tipping behavior from 1986, including some comparisons with previous years. It was found that roughly \$6.76 billion was spent on tipping in restaurants and other eating establishments compared with \$6.67 billion in 1985 and \$5.85 billion in 1979. However, the percentage of money spent at eating-style restaurants compared with all eating places dropped from 39% to 34% in 1986. As in previous reports, restaurants were separated into categories that were determined to be “tipping style” restaurants, though even when split in this manner the “stiffing rate” seemed higher than it should be at 30%. Given this, they were recategorized based on the main type of food in order to create a group of “high tipping-type restaurants.” This category was found to have tipping incidences of more than 80% on most occasions. They also note that the estimates of tipping revenue that they produce are lower than those provided by the U.S. Census Bureau and higher than those generated by the Bureau of Economic Analysis.

In addition to their standard analyses, a regression analysis was conducted specifically using the variables and information that might be available to the IRS in order to create a framework for future use and identification of tipping discrepancies. Scaled weights and a combination of scaled and expenditure weights were used in the analysis. The run with the expenditure weights was done to try to correct for some of the downward bias that occurs when bill size increases. The expenditure-based approach accounted for a higher R^2 than the scaled approach only (16.8% versus 13.1%). Predictions using the scaled weights alone also showed somewhat higher tipping rates than were accurate.

AGE: Middle-aged and older populations had higher rates of tipping incidence compared with younger groups.

GEOGRAPHY: Regional differences were found such that the Northeast area (which consisted of the New England and Middle Atlantic Census divisions) tipped at higher rates. Nonmetropolitan areas had one of the highest negative predictive values in the analysis. Metro areas had higher rates of tipping incidence than nonurban areas and their respective census regions. Metro areas were also significant in the regression analysis.

INCOME: As in previous studies, they found some differences in tipping behavior based on income levels. In this particular report, they found that tipping incidence was higher with higher socioeconomic statuses. The difference in tipping rate between the highest and lowest income group was only about 1%, so the range in this type of tipping behavior was not too great.

EDUCATION: Education had a similar effect on tipping incidence as did income, but had a great range of tipping rates. Tipping rates were also 1.5% higher among the highest education group compared with the lower groups.

CASH VERSUS CREDIT: Credit cards had the largest coefficient in the regression analysis, showing that credit card users had higher tip percentages than those who paid with cash.

INDUSTRY/SERVICE: Establishments that served alcohol were not found to be as important to the regression analysis as had been found in previous reports.

Rind, B. (1996). Effects of beliefs about weather conditions on tipping. *Journal of Applied Social Psychology*, 26(2), 137–147.

DESIGN OVERVIEW: In the first study, 266 adult hotel guests (181 males and 85 females) were put into four conditions. A room-service server reported one of the four weather conditions (sunny, partly sunny, cloudy, or rainy) when asked or volunteered the information (if the guests didn't ask) while delivering food or drinks. He always reported temperatures in the 50s. The windows of the hotel rooms were soundproof and dark-tinted that gave the impression it was cloudy even under sunny conditions.

In the second study, 205 adult hotel guests (115 males and 90 females) were randomly assigned to four conditions. A room-service server reported one of the four weather conditions (cold and rainy, cold and sunny, warm and rainy, warm and sunny) when asked or volunteered the information (if the guests didn't ask) while delivering food or drinks.

GEOGRAPHY: For the first study, a linear contrast analysis revealed a significant positive association between believed weather conditions and tipping. Tipping percentages improved as the conditions went from rainy (18%) to cloudy (24%) to partly sunny (26%) to sunny (29%).

For the second study, an ANOVA demonstrated that hotel guests in the sunny condition tipped significantly higher percentages than those who were told it was rainy. However, there was no effect for the temperature conditions.

Sanchez, A. (2002). The effect of alcohol consumption and patronage frequency on restaurant tipping. *Journal of Foodservice Business Research*, 5(3), 19–36.

DESIGN OVERVIEW: A waitress at a steakhouse restaurant collected data for 164 tables during dinnertime over a three-month period; however, only 138 tables (158 parties) were included in the analysis. The waitress recorded several variables of interest, including group ethnicity, group size, number of parties (number of checks), party size, customers' and paying patron's ages (ages estimated), customers' and paying patron's gender, number of alcoholic beverages (for the party and

for paying patron), food bill per party, bill size per party, tip amount per party, and payment method. Several other variables, including the number of children per party and patronage frequency, were recorded and analyzed.

RACE/ETHNICITY: For analysis purposes, customers were either identified as Caucasian or non-Caucasian. Ethnicity did not have any significant effect on tipping behavior. Caucasians tipped slightly less (\$7.42) than non-Caucasians (\$7.49).

AGE: Results indicated that estimated age of the paying patron by the server was a good predictor of tips. Older, paying patrons tipped more than those paying patrons judged to be younger.

CASH VERSUS CREDIT: Paying patrons' choice of payment method (i.e., cash, check, or credit) did not have any relationship with the total tip amount. Those patrons who paid with cash or check tipped slightly more (\$7.49) than those using credit (\$7.42).

INDUSTRY/SERVICE: Consumption of alcoholic beverages was found to significantly affect the tip amount. Tips from paying patrons who had one alcoholic drink (\$10.19) or more than one (\$9.52) tipped significantly more than those who did not drink an alcoholic beverage (\$6.44).

Schwer, R. K., & Daneshvary, R. (2000). Tipping participation and expenditures in beauty salons. *Applied Economics*, 32, 2023–2031.

DESIGN OVERVIEW: A stratified, convenience sample of 317 respondents was selected for this survey. This sample included a mix of respondents from banks, university staff and students, government employees, and customers of barbershops and beauty salons. Furthermore, the survey was conducted over time periods during the spring and summer of 1995. Questions on the survey dealt with patronage, what barbershops or beauty salons they go to, important qualities in the salon or barbershop they go to, and various demographic and socioeconomic questions.

Analyses were conducted using a combination of probit and Tobit regressions. Two Tobit regressions were used, a censored version as well as a truncated run. The truncated, two-step Tobit model showed the better fit of the Tobit models.

SERVICE/INDUSTRY: Overall, while Post (1992) recommended tipping 15% to 20% for hair salon/barbershops, it was determined that all customers tipped at 8% of their bill, and 9% when customers who left no tip were excluded.

INCOME: Income was included in the analyses, but significant findings were only discovered in the probit analysis. In both cases the dummy variables showed marginally significant findings.

AGE: Results from the truncated Tobit analysis indicate a marginally significant finding for tipping behavior, such that tipping rates decrease with the age of the respondent, though no such significant finding was discovered in any of the other data runs.

GENDER: Mixed findings regarding gender were found between the probit and truncated Tobit models. The probit model showed a marginally higher tipping total from women than men, but an opposite finding was reported in the truncated Tobit analysis.

RACE: White respondents were found to tip marginally more in only one of the three models (the censored Tobit) and race was generally found to be a nonsignificant variable.

Seiter, J. S., & Weger, H., Jr. (2013). Does a customer by any other name tip the same? The effect of forms of address and customers' age on gratuities given to food servers in the United States. *Journal of Applied Social Psychology, 43*, 1592–1598.

METHODOLOGY: A field experiment of diners ($N = 142$) at two Utah restaurants was conducted to examine the effects of differences in how servers addressed customers (first name, Mr./Mrs., etc.) on tip rate. A regression analysis was conducted that included form of address effects, customer age, and the interaction between age and form of address. Data was collected by three student/servers.

AGE: In the regression model without the interaction (i.e., just form of address and age), customer age had a negative association with tip amount, but the estimated relationship was not found to be statistically significant at the 5% level, but was at the 10% level ($p = .09$). This negative relationship was stronger when the customers were addressed by their first name.

Simpson, H. (1997). Tips and excluded workers: The New Orleans test. *Compensation and Working, Bureau of Labor Statistics, 32–36.*

DESIGN OVERVIEW: Data was gained from “BLS field economists” in face-to-face interviews for the most part. Of the 359 establishments that were sampled, 77% provided some data, but only 11 provided tipping data, indicating that the findings in this article are to be considered as preliminary without any significance testing. Besides information regarding the number of tipped workers at the establishment and the dollar amount of tips collected, the BLS workers also gave a rating for their confidence in the data that was provided. However, while the majority (82%) of the data for “hours worked” was determined to be good, only 55% of the tip data was considered to be good, and 27% was considered “poor.” This indicates that the data in the article might be flawed and underscores the difficulty of obtaining reliable tipping data.

SERVICE/INDUSTRY: Of the occupations that met publication criteria (certain number of workers from a certain number of establishments at least), waiters had the highest amount of average tips per hour (\$6.10), followed by hostesses (\$5.73), bussers (\$4.86), and bartenders (\$3.70).

GEOGRAPHY: The article reports that tipped employees would underreport their tips during the busy months and overreport during the slower months in order to balance things out for their bosses and create less hassle. Similarly, the months when data was collected (July–August) were considered to be slower tourist months, so the data might be skewed somewhat by that.

Speer, T. (1997). *The give and take of tipping. American Demographics, 19(2), 51–54.*

DESIGN OVERVIEW: Random telephone survey of roughly 1,000 adults in 1996. Respondents were asked what the largest determining factor was regarding their tipping behavior, and service was often claimed as the most important thing, though this percentage was smaller among non-restaurant services.

INDUSTRY/SERVICE: Roughly 28% of respondents indicated that they never tipped the individual in the hotel who replaces their towels and bed sheets. Also of note was that 36% of respondents indicated that they always carried their baggage at hotels and airports, and were thus unable to answer any questions about tipping this particular profession. Similarly, roughly half of adults reported that they don't use taxi cabs or limo drivers, so they were unable to answer any such questions about tipping behavior. Finally, 40% of respondents indicated that they are never served by bartenders.

Also worth noting is that this article has a chart that indicates the percentage of respondents who indicate specific tipping percentages for a number of different industries.

INCOME: Higher-income (\$50,000 or higher) individuals reported that the reason they tipped was that they tipped to help some individuals (notably parking valets, luggage handlers, and taxi drivers). Lower-income individuals were less likely to tip at all because they reported that the bill should reflect the full cost of the service, though this behavior does not extend to waiters.

GEOGRAPHY: Southerners were more likely to say they would never tip for some services, mostly taxi drivers, waitstaff, and barbers, while Midwesterners were the most likely to say that they would never tip parking valets, bartenders, maids, and luggage handlers. Northerners tipped the highest of the groups when split by region, or reported as much.

GENDER: In this study, women were reported as more likely to leave a tip than men, particularly when it comes to services other than taxis or waitstaff. Women are more likely to report that they tip based on the impact that it has on others when compared with men.

Star, N. (1988). *The international guide to tipping: When, where, and how much to tip in the U.S. and around the world.* New York, NY: Berkley Books.

DESIGN OVERVIEW: Star's book discusses cross-country differences in tipping. Specifically, the author describes expectations and norms for tipping across 38 professions in 34 different countries. The 38 professions cover a diverse set of service-related professions including restaurant jobs (e.g., servers, bartenders, hostesses, etc.), guides, hotel staff, and hair stylists. According to Lynn, Zinkhan, and Harris (1993) that had correspondence with Star, her tipping suggestions and summaries were primarily based on questionnaires sent to hotels, national railroads, resorts, restaurants, tour groups, and so on in each of the 34 countries.

Thomas-Haysbert, C. D. (2002). The effects of race, education, and income on tipping behavior. *Journal of Foodservice Business Research*, 5(2), 47–60.

DESIGN OVERVIEW: Phone surveys were conducted on a population of 1,005 respondents. The phone survey was conducted by Market Facts for *American Demographics* and methodology of the phone survey is discussed in greater detail in another article (Speer, 1997). Questions were asked regarding whether respondents tipped various service-industry workers such as servers, bartenders, taxi drivers, parking attendants, and luggage handlers, and why they tip or did not tip.

INDUSTRY/SERVICE: Luggage handlers were tipped the most (98% said they always tipped this group, followed by servers, parking attendants, taxi drivers, and bartenders.

INCOME: Income was found to significantly affect tipping behavior and when used as a dichotomous controlling variable it nullified the influence of race on tipping behavior.

EDUCATION: Same effect as income was found in that it is significantly related to tipping behavior and when used as a control it nullifies the effect of race on tipping behavior.

RACE/ETHNICITY: White respondents tipped every category of worker significantly more often than Black respondents, but this effect was nonsignificant once education and income levels were considered for all service workers except for taxi drivers. However, Black respondents were more likely to indicate that service quality was more important to them than White respondents and that they tipped more to ensure better service in the future. Blacks were also more likely than Whites to indicate that they did not tip because they felt that it should be included in the bill. Black respondents reported that they tipped less than Whites but this effect was nullified when income and education were incorporated into the model.

Appendix C – Search Engines and Search Terms

Table 5. Search Engines

Search Engine	Description
University Library System	Online database of journal articles maintained by local DC Metro University.
Google Scholar	Google search engine that produces links to both gated and ungated scholarly articles.
JSTOR	Archive of peer-reviewed articles published in academic journals.
Social Science Research Network (SSRN)	Archive of social science working papers.
Business Source Complete (EBSCOhost)	Database containing archived peer-reviewed articles published in business-related journals.
ABI/INFORM Complete (ProQuest)	Database containing peer-reviewed articles published in business-related economics, business, accounting, and marketing journals.
Accounting & Tax (ProQuest)	Database containing peer-reviewed articles published in high-impact accounting, auditing, tax management, and tax law journals, as well as trade publications.
PsycINFO	Database of peer-reviewed behavioral science and mental health articles.

Table 6. Search Terms

Search Term	Themes
Gratuity, tipping, tip giving, stiffing behavior, tip reporting	GENERAL TIPPING, NATIONAL AVERAGE TIPPING RATES
Internet, mail, and mixed-mode surveys: the tailored design method.	METHODOLOGY
Regional, urban versus rural, metropolitan tipping differences, holiday differences, seasonal effects, tourist tipping	GEOGRAPHY
Income, education, age, gender, SES, salary tipping differences/restaurant tipping differences	INCOME, EDUCATION, AGE, GENDER
Black-White/Asian/Hispanic/ racial tipping differences	RACE/ETHNICITY
Tipping knowledge, tipping norms	TIPPING KNOWLEDGE
Service charge law change, mandatory service charge, mandatory restaurant tips	SERVICE CHARGE
Tipping differences by industry, non-restaurant tipping, tipping in services industries, alcohol and tipping	INDUSTRY/SERVICE
Method of payment tipping, credit card/cash tipping, cash differential	CASH VERSUS CREDIT