Forensic Economics

Eric Zitzewitz*

A new meta-field of "forensic economics" has begun to emerge, uncovering evidence of hidden behavior in a variety of domains. Examples include teachers cheating on exams, road builders skimping on materials, violations of U.N. sanctions, unnecessary heart surgeries, and racial biases in employment decisions, traffic stops, auto retailing, and even sports judging. In each case, part of the contribution of economic analysis is in uncovering evidence of wrongdoing. Although research questions differ, forensic economic work shares commonalities in approaches and limitations. This article seeks to draw out the common threads, with the hope of stimulating further research across fields. (JEL K13)

1. Introduction

In many subfields of economics, understanding behavior that agents would prefer to conceal has become a central component of the research agenda. In development, government corruption is viewed as one of the most important impediments to growth. In finance and health care, understanding how agents resolve conflicts between fiduciary duties and competing interests is crucial to many questions. In industrial organization, the welfare implications of industrial structures or auction designs often depend crucially on the extent of collusion.[1] Many important topics in labor, public, and education economics, such as racial discrimination, tax evasion, and teacher shirking, involve behaviors that agents are unlikely to confess.

These research questions have prompted academic economists to engage in what this survey will refer to as "forensic economics." Traditionally, forensic economics has referred to the application of economics to the detection and quantification of harm from behavior that has become the subject of litigation, and has been practiced by experts who are paid by the court or one of the parties.[2] The academic forensic economics reviewed here also applies economics to

[1] See, for example, Klemperer (2002), who argues that auction designs that produce the most revenue when bidders behave noncooperatively are the least robust to collusion.

[2] This is the focus of the National Association of Forensic Economics (www.nafe.org), which publishes the *Journal of Forensic Economics*.

detecting and quantifying behavior, but does so with a different motivation: to advance the general understanding of behavior that is important to the functioning (or disfunction) of the economy. In some forensic economics work, the academic economist generates the first evidence of a particular behavior. More commonly though, the economist sizes the extent of an activity about which there had been only anecdotal evidence and provides insight into where it is more prevalent and why.

Forensic economics is usually necessary because the agents engaged in a particular behavior prefer to keep it hidden. Motivations for hiding behavior vary—in many of the examples discussed below, the behavior in question is very likely illegal. In other cases, behavior may be a violation of contract terms. In still others, it may be a violation of ethical norms. In each case, there may be significant controversy as to which side of the line on which behavior falls. In selecting papers to discuss, I choose those dealing with economically important hidden behavior that I judged would be generally considered to be at least unethical. Ethical behavior is often hidden for various reasons (e.g., a desire for privacy), but I will leave it, as well as behavior that is generally readily admitted, outside the scope of the review.[3]

Forensic economics is sometimes work that could be done by regulators or other third parties. The need or opportunity for economists to get involved arises for different reasons. In some cases, the subjects of forensic economic work are corrupt government officials or captured regulators themselves, agents who obviously have limited incentive to self-investigate. In other cases, corruption, capture or simply conflict avoidance by those in regulatory roles may help explain the space left for outside investigation. In still other cases, the activity uncovered by forensic economics is not illegal, and perhaps not even of regulatory interest, but is nevertheless economically important.

Beyond regulatory incentive problems, in many cases, economists get involved because economics gives them a comparative advantage. Wrongdoers may cover their tracks well enough to fool traditional forensic investigators, but they leave distortions in the data that economic analysis can detect.

The work by Heron and Lie (2007) and others on executive stock option backdating provides a compelling example. Executive stock options grants receive more favorable tax and accounting treatment when they are granted with strike prices that are at-the-money (i.e., equal to the current share price). Obviously, options with low strike prices (those that provide the executive with the option to purchase company stock at a lower price) are more valuable to executives. Thus the temptation may exist for executives to arrange grants with low strike prices but to represent the grants as having been made at-the-money at the time of the grant. This can be done by choosing a day from the recent past on which the stock price was low and claiming that an option grant was made on that day.

Indeed, the typical stock option grant appeared quite fortunately timed from the perspective of the executive—granting company stock prices tended to drop immediately before the grant and rise afterwards. Market

---

[3] Inevitably, there are some borderline cases. Notwithstanding other motivations, hiding behavior can sometimes be an indication that an agent expects others to view their actions as unethical. In my experience, production workers generally admit that they would work harder under a piece rate than an hourly wage (as Lazear 2000 finds), but teachers would not admit to the levels of absenteeism uncovered by Duflo, Hanna, and Ryan's (2008) work, discussed below. Likewise, many teachers would admit teaching the subjects covered by a test, but not cheating on the test (as documented by Jacob and Levitt 2003a and 2003b). Sales staff would likely admit working hard to generate sales before a quarter-end in response to an incentive nonlinearity, but not offering large discounts to get nonincremental business shifted forward in time (as documented by Larkin 2007).

efficiency requires that postgrant stock returns be uncorrelated with information that was public on grant day, assuming that was really when the decision was made. The correlation between grants and future returns suggests two possibilities: (1) executives were granting options when their private information suggested their stock was undervalued, or (2) executives were choosing dates after the fact that turned out to be stock price lows. While the first possibility might be considered insider trading, misrepresenting grant dates in order to characterize option grants as at-the-money was more serious, as it constituted both tax and accounting fraud.[4] Consistent with the second possibility, Heron and Lie found that the correlation between grants and future stock returns declined sharply after the Securities and Exchange Commission (SEC) adopted a rule requiring grants to be disclosed within two days, and the remaining positive correlation came mainly from firms that filed the required disclosure late. Newspaper and regulatory investigations of the most egregious cases subsequently confirmed that several firms had indeed backdated their option grants.

Heron and Lie's work relies on the theory of efficient financial markets for the null hypothesis that allows it to identify backdating. In their study of racial profiling by police, Knowles, Persico, and Todd (2001) likewise use economic theory to develop a null hypothesis to distinguish between taste-based and statistical discrimination. The key insight from their economic model is that so long as every "type" of car is searched with probability less than one (where type includes the race of the driver and other characteristics observable only to the officer), then the marginal return to searching a given type of car must equal the average return. Thus if we assume that police seek to maximize felony drug arrests, and the average police search of a black motorist's car is less likely to find drugs than the average police search of a white motorist's car, it implies that the police are over searching black motorists relative to whites, perhaps due to a taste for doing so. Knowles, Persico, and Todd actually find that searches of black motorists' cars are slightly more likely to find drugs, implying that the higher search rates for black motorists would be due to statistical discrimination, rather than tastes.[5] While this might not be of much immediate comfort to the innocent black motorists who have to endure higher search rates, understanding motivations for discrimination is crucial to designing policies to alleviate its effects.

These examples illustrate a typical relationship between economics, forensic economics, and forensics. Economics provides a null hypothesis whose violation suggests hidden activity. Forensic economics documents the violation (or its absence) and provides evidence on potential alternative explanations. Traditional forensic investigation then gathers the specific evidence needed to convict or exonerate specific individuals. Along the way, the field of economics makes a contribution to society and learns, via a clinical study, about how agents balance the temptations and consequences of the hidden behavior in question.

While the questions addressed by forensic economics differ in different subfields, the techniques and underlying economics are actually quite similar. This is all the more surprising given that forensic economics seems to be developing almost

[4]The initial work on the correlation between option grants and subsequent stock returns focused on executives' private information as the explanation (e.g., Yermack 1997; Aboody and Kasznik 2000; Chauvin and Shenoy 2001). Like Heron and Lie, Lie (2005) and Narayanan and Seyhun (2008) focus on a backdating explanation.

[5]In contrast, Anwar and Fang (2006) find higher search rates and lower success rates for searches of black motorists by the Florida State Police, suggesting a role for taste-based discrimination.

independently in different subfields (at least, based on an informal analysis of citation patterns). Given that existing reviews of forensic economics papers have typically focused on a small number of papers in a single subfield or topic, a contribution of this paper will be to review a broad range of papers from different fields in a way that highlights their similarities in how economics is applied.[6]

The next section presents a taxonomy of forensic economics papers, discussing over 100 different studies. The following section reviews some of the common themes from the substantive conclusions of the studies and then turns to some higher-level questions. What does economics learn from forensic economics? Why has forensic economics had dramatic policy impact in some cases, but limited impact in others? How can policy either facilitate or frustrate forensic economics? To preview, a key question is whether policymakers view themselves as the principals cheated by hidden behavior, or as the allies of the agents doing the cheating. When policymakers have viewed themselves as principals, they have used forensic economics approaches and results to change agents' incentives, through policy and enforcement thereof, and they have engaged forensic economists by forcing the disclosure of data needed to produce the next round of results. In contrast, where policymakers have taken the opposite view, forensic economics results have had limited impact, and data availability has even been curtailed.

---

[6]Past reviews that have included forensic economics papers include Ritter (2008) and Macey and O'Hara (2009) for finance, Svensson (2005) and Fisman and Miguel (2008a and 2008b) for developing country corruption, Porter (2005) for collusion in auctions, Healy and Whalen (1999) for earnings management, and Slemrod (2007) for tax evasion.

2. *A Taxonomy*

Most forensic economics work identifies hidden behavior by testing data against a null hypothesis, which is often derived from economic theory. We can classify analyses by the source of identification and discuss the work in five groups.

The first, and simplest, approach to detecting hidden behavior is simply to observe it directly. Some economists have found direct evidence of wrongdoing in data collected by others, others have constructed the measures themselves, while still others have run audit studies or field experiments to detect it.

A second, related, approach is to identify hidden behavior by constructing two measures that capture the same economic activity, but are affected differently by hidden behavior. As with the direct observation studies, in some cases the economist finds two measures of the same outcome in official data, while in other cases he or she generates the second measure. In these studies, the null hypothesis is that hidden behavior is the main or only reason the two measures should differ, and concern about alternative explanations focuses on potential other reasons for differences.

A third approach is to examine how a single outcome, that is potentially the product of both honest and hidden behavior, varies with the incentives for hidden behavior. Here the null hypothesis is that hidden behavior is the only reason the outcome would vary with these incentives. Ruling out potential omitted variables is a crucial issue with these studies. One strategy is to exploit discontinuous changes in the incentives for hidden behavior, resulting, for instance, from sudden changes in enforcement. Another strategy is to generate the variation through a field experiment. In some settings, the economist feels comfortable arguing that controlling for observables is sufficient to address these issues.

A fourth approach is to build a model of honest behavior and test the data for deviations from that model. For example, work that tests for collusion in product markets or auctions often begins with a model of noncooperative behavior and checks for deviations that are consistent with collusion. Models are not always formal or even economic. A number of studies identify hidden behavior using statistical models that find hidden behavior by looking for outliers, such as odd patterns on test answer sheets in particular classrooms or very high open heart surgery rates in particular towns.

A fifth approach exploits the predictions of efficient market or price theory. One strategy exploits the fact that future asset returns in an efficient financial market should be difficult to forecast using public information, so if a decision (e.g., to grant stock options or place a mutual fund trade) supposedly taken at time t is correlated with returns from t to $t + s$ that make the decision more profitable, this is good indication that it was really taken at time $t + s$. Another strategy exploits settings in which efficient markets aggregate information and provide an estimate of the scope of hidden behavior.

The above taxonomy is one of techniques, not papers. Many studies combine multiple techniques, such as by using one to establish a main result and another to examine alternative explanations. For instance, direct observation and measurement consistency studies often also test whether apparent hidden behavior is correlated with incentives for it. In settings where the enforcement environment changed suddenly (sometimes due to the circulation of a paper's first draft), studies often check whether the evidence of hidden behavior was simultaneously reduced.

2.1 Direct Observation

In a few settings, economists have not needed to use economics to help detect hidden behavior because they have been able to observe it directly. Examples include studies where the researchers have found their evidence in preexisting datasets, by collecting data themselves, or by uncovering hidden behavior through an audit study or field experiment.

Christie and Schultz (1994 and 1995) provide an example of finding hidden behavior that was "hiding in plain sight" in official data. Using data on trades and quotes provided them by Nasdaq, Christie and Schultz find that Nasdaq market makers avoided quoting certain stocks in odd-eighths. They interpret this as evidence of collusion to maintain a higher minimum tick size (an important determinant of market maker profit). Support for this conclusion comes from that fact that the incidence of odd-eighth quoting increased sharply on the exact day their original study was publicized, which Christie, Harris, and Schultz (1994) argues was consistent with the publicity causing the collusive agreement to collapse.

Another direct observation example is provided by Fisman and Miguel (2007), who obtained data from New York City on parking ticket nonpayment by United Nations diplomats. Fisman and Miguel find more nonpayment by diplomats from countries with reputations for more corruption, as reflected in cross-country surveys. This both provides an independent validation for these corruption perception surveys and also suggests that arguably petty unethical behavior may be correlated with more serious unethical behavior. Levitt (2006) also studies petty unethical behavior, using data from an entrepreneur who sells bagels and donuts in offices using a lockbox and an honor system. The honor system works reasonably well, with an 11 percent theft rate. Theft declined sharply after September 11, 2001, is higher after price increases, and is 33 percentage points (!) higher at nonprofits. Company fixed effects have significant explanatory power. Levitt argues that these

results suggest that perceived fairness affects decisions to engage in unethical behavior. Both the parking ticket and the bagel results hint that certain subpopulations may develop a sense of entitlement to at least small-scale unethical behavior, and that this may be correlated with larger misdeeds.

Edelman and Larkin (2009) provide directly observed evidence of misbehavior by another category of nonprofit employees: academic researchers. The Social Science Research Network (SSRN) tracks downloads of academic papers and compiles lists of the most downloaded papers in a given topic area in the last sixty days. Placement on these lists presumably increases the visibility of the paper, which might be especially important to less-known researchers and those approaching tenure reviews. Download statistics are also used by some institutions in evaluating their scholars. Initially SSRN counted all downloads, including multiple downloads by the same Internet Protocol (IP) address, so researchers could easily inflate their papers' statistics. Edelman and Larkin analyze the frequency of suspicious downloads (24 percent of the total!) and find that they vary with demographics such as field (economists make fewer suspicious downloads than those in finance, law, management, and accounting). Edelman and Larkin find more suspicious downloads for papers that are near the boundary of top-ten-list qualification, where incentives for inflation may be higher. Perhaps surprisingly though, they do not find evidence that suspicious downloading declines after scholars are tenured, when incentives for it presumably decline. One possibly is that this form of unethical behavior is habit-forming, and thus persists even when incentives for it are weakened.

In three other examples, researchers generated their own direct observations of illegal behavior. Olken and Barron (2009) hired surveyors to ride along with truckers in Indonesia and observe over 6,000 bribe payments to police officers, soldiers, and weigh station attendants. They find that bribe demands reflect more economic sophistication than one might have supposed. When the Indonesian army was withdrawn from half of a major trucking route, bribe demands on the other half rose, but total bribe demands on the entire route declined, consistent with the checkpoints operating in a decentralized manner and taking double marginalization concerns into account. Levitt and Venkatesh (2000 and 2007) analyzed drug gang finances and street prostitution, respectively, using data they collected directly from participants. Apart from the low compensation for engaging in such risky behavior, one of the striking findings is the level of complicity of the Chicago police, who prostitutes "are more likely to have sex with . . . than get officially arrested by" (abstract of the latter paper).

In a final set of direct observation studies, forensic economists run audit study experiments to directly observe hidden behavior in the field.[7] One of the oldest examples are audit studies of racial discrimination by realtors (e.g., Yinger 1986). Ondrich, Ross, and Yinger (2003) examine the data from one such study in which matched pairs of white and black couples approached real estate agents asking to see a specific house. The authors find that agents were less likely to show a requested house to all customers when it is in an integrated suburban neighborhood but were more likely to show blacks these houses than whites. Agents are also more likely to steer blacks away from their initially requested house, except when the house has visible problems. Agent marketing efforts increase with the initially requested house's asking price for whites but not blacks. The authors argue that agent behavior could be consistent with either taste-based or

---

[7] See List (2006) for a general overview of field experiment methodology.

statistical discrimination (potentially based on misinformed preconceptions) about average preferences of races for certain house characteristics and a desire to maximize transaction probabilities.

Audit studies have also tested for labor market discrimination on race (see Fix and Turner 1998 for a review) or gender (Neumark 1996) or discrimination in commercial settings (Ayres and Siegelman 1995; Yinger 1998; Riach and Rich 2002; List 2004). In general, these studies do find disparities in the treatment of different groups, but these findings and their interpretation have been critiqued on several grounds. One criticism of audit studies involving in-person visits is that, despite the training that auditors undergo, the auditors may differ in ways unobservable to the researcher that the auditee believes are correlated with productivity (Heckman and Siegelman 1993). A second, related, criticism is that the auditors know the nature of the study in which they are undertaking, and may consciously or unconsciously vary their effort levels or behave in ways that seem artificial to different degrees. Audit studies involving resumes (Jowell and Prescott-Clarke 1970; Bertrand and Mullainathan 2004; Banerjee et al. 2009) can make resumes identical except for a name, which is chosen to signal race.[8] Even that approach has been critiqued on the grounds that distinctively black names may signal more than just race, such as a disadvantaged background (Fryer and Levitt 2004).[9] While discrimination against black-sounding names may in practice disadvantage many of the same individuals as racial discrimination, as discussed above, understanding mechanisms is important to policy implications. A third critique of audit studies is that they may capture disparities in treatment during the search process, but these disparities may only result in small impacts on ultimate outcomes, such as employment or wages (e.g., Heckman 1998).[10] This highlights an important caveat for many forensic economic, and indeed many nonforensic, results: welfare implications of findings are not always as straightforward as they may first appear.

2.2 *Measure Comparison*

In settings where economists are not able to directly observe hidden behavior, they need to devise a statistical test to detect it. One of the simplest null hypotheses in forensic economics is that two measures of the same economic activity should yield similar results, at least in the absence of activity that is hidden from one of the measures. In this section, I review several versions of this approach: studies that compare official data from two sources, those that compare data self-reported by agents with data from a second source, those that compare inputs and outputs, and those that compare official data with a second measure generated by the economist.

---

[8] Hanna and Leigh Linden (2009) conduct a related experiment in which they attach false cover sheets with randomly generated demographic characteristics (age, gender, caste) to exams in India. They find grading discriminates against low-caste "students," and, interestingly, that the discrimination comes only from teachers who are themselves low-caste.

[9] Bertrand and Mullainathan (2004) address this critique by separately examining the effects of high-education and low-education black and white names. They find that names typically used by high-education black mothers still receive fewer callbacks than those used by a lower-education white mother. At least if one assumes that employers infer race and education from names equally well, this suggests employer preferences are over race rather than mothers' education.

[10] Heckman's critique is illustrated by Goldberg (1996), who finds that mean and median markups paid for cars in actual transactions do not vary by race and gender, although minority purchase prices have higher dispersion. She suggests that some minorities may have high reservation prices (due to high information or search costs), causing all minorities to therefore face statistical discrimination in the initial offers they receive. But the absence of a difference in mean transaction prices suggests that minorities with lower reservation prices improve on these offers through bargaining.

2.2.1 *Two Official Sources*

As an example of a paper comparing official data from two sources, Fisman and Wei (2004) compare Hong Kong's reported exports to China and China's reported imports from Hong Kong. They treat the difference, shipments reported as exports to Hong Kong but not as imports to China, as an "evasion gap." One might be tempted to dismiss this gap as a statistical discrepancy, but the authors find that it is larger for products where Chinese tariffs are higher and smaller for products with high tariffs on closely related products. These correlations suggest that tariffs are evaded in part by misclassifying products into lower tariff categories; it is difficult to imagine why a statistical discrepancy from another source would be correlated in exactly this manner. Topalova, Mishra, and Subramanian (2007) conduct a related analysis using variation across both products and time in Indian tariffs (before and after large reductions in the 1990s). They likewise find more "missing imports" when tariffs are higher.

Fisman and Wei (2007) conduct an analogous test for "missing exports" of cultural artifacts to the United States; shipments that are reported to the United States but not to their country of origin. They find more missing exports for countries regarded as more corrupt on international surveys, and that a similar correlation does not exist for toys. This last "placebo" test suggests that smuggling is behind the original result, as opposed to a general undercounting of exports in corrupt countries.

In the Fisman and Wei papers, the differences between import and export statistics arose from differing incentives for underreporting to the two different countries. Like Topalova, Mishra, and Subramanian, Ljungqvist, Malloy, and Marston (2009) examine a setting in which incentives for misreporting change over time. Specifically, they compare the 2002 and 2004 versions of Thomson Financial's I/B/E/S stock analyst recommendation data (in both cases, the data covering the 1993–2002 time period). The 2004 version followed regulatory investigations of conflicts of interest and optimism bias in recommendations during the dotcom era; these investigations understandably led certain analysts and their employers to wish that some past bullish recommendations be forgotten about. Since these brokerages are both sources and customers of Thomson Financial, Thomson might have had an incentive to favor their interests over those of other consumers of the data. The authors find many instances of recommendations being changed, anonymized, added, and deleted. Deleted recommendations were disproportionately strong buys; added recommendations were disproportionately holds and sells. Changes were disproportionately from buy to sell. Anonymizations were disproportionately strong buy recommendations of stocks that subsequently underperformed; anonymizations were also more likely for analysts who remained in the industry than for those who exited. After Ljungqvist, Malloy, and Marston circulated a working paper, Thomson Financial reversed many of the changes made to the 2004 version of the data, attributing them to a series of computer errors (see the published version of the paper and Ritter [2008] for more details).

Zinman and Zitzewitz (2009) provide an example of a study that compares self-reported and official data and that exploits even higher-frequency variation in incentives for misreporting. We compare "snow reports" of new natural snowfall issued by ski resorts with data from surrounding government weather stations. The resorts report more snow than the government, which could of course be due to the resorts being located on especially snowy mountains. We find though that this gap is much larger on

weekends, especially for resorts that cater to experts and are within driving distance of major cities. Given the absence of a weekend effect in true precipitation, this result suggests that resorts engage in deceptive advertising, especially when the returns to it are high.

Snyder and Zidar's (2009) study of resume padding by economists provides another example of comparing self-reported with official data. Snyder and Zidar compare publications listed in vitas posted on the web with journals' tables of contents. They find that while outright fabrication of publications is extremely rare, more subtle forms of inflation are more common (e.g., mischaracterizing a Papers and Proceedings article as a regular *American Economic Review* article; including invited articles among peer reviewed).

2.2.2 *Inputs and Outputs*

Turning to studies that compare inputs and outputs, Sukhtankar (2011) provides the most literal example in his study of corruption in Indian sugar mills. Sugar production is very close to a fixed proportions technology; if sugar cane is crushed but sugar is not produced, it is very likely that output was diverted. Sukhtankar finds that sugar mills suffer declines in their output–input ratio in election years, especially for those mills controlled by politicians who are contesting the election. He argues that this is consistent with the politicians diverting resources from the mills to finance their campaigns, and he finds evidence that the farmer-members are compensated by receiving higher prices the next year if the politician in question wins.

Another example of identifying hidden behavior by comparing outputs and inputs is the "return gap" measure of Kacperczyk, Sialm, and Zheng (2008). The authors compare the returns of a mutual fund (with fees and expenses added back) to the returns of its most recently disclosed prior holdings. This gap is negative for the average fund, which helps explain why the average equity mutual funds underperforms the stock market even after adjusting for expenses. The return gap has many potential explanations, since it would capture the profitability of trades done since the most recent holdings disclosure and might be negative on average due to commissions and other transactions costs. The authors show however that return gaps are more persistent than the returns themselves, which they argue is consistent with the pro- or anti-investor nature of a fund's hidden actions being more persistent than the returns of its holdings. They also find that differences in return gaps are consistent with the incentives for cross-subsidization faced by fund families (Gaspar, Massa, and Matos 2006). Small, young funds, with low expenses and strong recent and inflows performance have future inflows that are much more sensitive to $1 of extra return than larger, older, high-expense, and poorly performing funds. If fund families face choices about how to allocate shares in hot initial public offerings (IPOs) or which fund trades a given stock first, they face a temptation to favor funds with more performance-sensitive inflows in ways that would produce the patterns found by Kacperczyk, Sialm, and Zheng.

Morey and O'Neal (2006) also compare mutual fund returns with the returns of their prior disclosed holdings, but with a different objective. Morey and O'Neal find that bond mutual funds' returns are more correlated with lower-credit-quality bond returns than one would expect from their disclosed holdings. They also find that bond fund returns become more correlated with safer bonds around holdings disclosure dates. They conclude that bond funds engage in "window dressing," altering their holdings around disclosure dates "to present a safer portfolio to shareholders."

Baicker and Staiger (2005) provide a fiscal example of an input–output gap. They

test for diversion of matching funds distributed to county hospitals under Medicaid's Disproportionate Share Hospital (DSH) program. DSH was a matching funds program, intended to motivate higher state spending at hospitals that served the poor, but states could game the program by reversing transfers of funds once those funds were matched by the federal government. For example, Baicker and Staiger cite several examples of such transfers from a Government Accountability Office report, such as a Michigan nursing facility that received a $277 million DSH payment from the state (that was 50 percent federally reimbursed) and wired $271 million back to the state the same day (348). They find that counties that receive an extra $1 under DSH report net intergovernmental transfers in the *Survey of Governmental Finances* that are only 57 cents higher. This suggests that the remaining 43 cents was either transferred to other governments or offset by a decline in other transfers to the county in question, suggesting that transfers of matched funds back to states are quite common. The authors find that their estimate of nondiverted DSH funds are much more positively associated with patient outcomes than their estimate of diverted DSH funds, helping to explain why previous analyses had found limited benefits from DSH funds.

2.2.3 *Researcher-Created Second Measures*

In a final set of studies, second measures were unavailable and they needed to be created as part of the research. Clotfelter (1983) uses data from a project in which the U.S. Internal Revenue Services (IRS) randomly selected personal income tax returns for audit and estimates what he calls "the sum of evasion and favorable errors of interpretation" (365). He finds more underreporting for returns with high marginal tax rates (controlling for income) and returns with more income that was not reported to the IRS by another party (e.g., self employment income, as opposed to wages, interest, or dividends).

Olken (2006) estimates rice received from a Indonesian government transfer program from a survey of recipients and compares this with administrative data on rice distributed. He estimates that 18 percent of the rice was stolen, and that the missing share was higher in sparsely populated and ethnically heterogeneous areas. Reinikka and Svensson (2004 and 2005) compare administrative data on Ugandan central government school appropriations with estimates of money that actually reached the schools, finding that only 13 percent of funds did in 1991–95, but that this rose to 91 percent after a newspaper campaign and other reforms. Larger and better connected schools suffered less resource diversion in the earlier period.

Jin and Kato (2006) uncover smaller scale fraud by purchasing baseball cards on EBay, obtaining independent ratings of the quality of these cards, and comparing these with self-reported quality. They find that sellers who self-report quality tend to over claim, relative to sellers who invest in third-party verification. Buyer's bidding behavior reveals an excessive willingness to pay for self-reported, as opposed to verified, quality. Jin and Kato also find that sellers who self-report exaggerated qualities are also more likely to fail to deliver a card entirely, providing another example of a correlation in small and larger-scale unethical behavior.

In summary, measure consistency studies identify a gap between two alternative measures of the same outcome. In the studies mentioned above, differences in the measures arise due to either lying (e.g., about income, imports, snowfall, bond portfolio risk, publications, or baseball card quality) or resource diversion (e.g., of sugar, rice, or Medicare payments). Differences in the measures could also have more innocent explanations, so the studies usually then examine the correlation of gaps with the profitability

or feasibility of hidden behavior to determine its likely explanation. Alternative explanations for these correlations may exist, but their potential role is reduced by the fact that the dependent variable is a gap between two related measures, netting out otherwise problematic confounding factors.

2.3 *Correlations with Incentives*

In a larger group of forensic economics studies, researchers have only one outcome measure, which reflects both honest activity and potential hidden behavior. The researcher then tests whether this measure varies with the profitability or feasibility of hidden behavior.

These studies face an identification problem common to much other empirical work, namely distinguishing between causation and correlation through other mechanisms. Three broad approaches are to: (1) exploit exisiting discontinuities in incentives, (2) to generate variation in incentives using an experiment, and (3) to identify situations where confounding factors can be convincingly ruled out, and thus correlations are sufficient evidence.

2.3.1 *Discontinuities*

One approach is to identify cases where incentives for hidden behavior vary discontinuously, but other reasons for correlations vary continuously.[11] Sudden changes in policies provide one such discontinuity. As mentioned above, many forensic economics studies bolster their main evidence by examining changes in agent behavior when regulatory investigations into the practice in question are announced. The studies reviewed in this subsection use discontinuities in incentives as their main means of identifying hidden behavior. These discontinuities

can arise from changes in government policy, technology, or other external events, or can be discontinuities built into existing incentive schemes themselves.

Changes in policy, technology, or other external factors often create discontinuous changes in the incentives for hidden behavior. Jin and Leslie (2003) study a policy change in Los Angeles that required restaurants to post a report card disclosing their hygiene inspection score. They find that hygiene scores improve and that hospitalizations for food-borne illnesses decline sharply, suggesting an improvement in true hygiene accompanied the improved scores. In a later paper, Jin and Leslie (2009) find that pre-mandatory-disclosure hygiene was poorer (and improvements due to disclosure larger) at franchisee-owned chain outlets, especially those with low repeat business. When a franchisee owns an outlet with low (outlet-level) repeat business, he or she does not internalize the effects of poor cleanliness on the reputation of the rest of the chain.

Di Tella and Schargrodsky (2003) examine the effects of auditing on government procurement. They find that the prices paid by hospitals in Buenos Aires to private-sector suppliers fall by 15 percent during a crackdown on corruption. They find that the crackdown has a greater effect in reducing the prices paid by higher-wage officials. This is consistent with an efficiency-wage theory of official honesty due to Becker and Stigler (1974), in which higher-wage officials engage in less corruption to avoid risking being fired.

Bandiera, Prat, and Valletti (2009) analyze dispersion in prices paid for similar items by different Italian public bodies. They use the addition and deletion of items from a centrally negotiated price list, which public bodies can but are not required to utilize, to distinguish between overpayment for active reasons (e.g., receiving bribes from the supplier) and passive reasons (e.g., lacking the time or motivation to price shop). They find

---

[11] This empirical technique, often called regression discontinuity, is used in many nonforensic applications (see Imbens and Lemieux 2008 for a recent review).

that the likelihood of buying via the central agreement is higher when the ratio of the agreement price to the outside price is lower, which they argue is more consistent with passive waste, since officials appear to use the central price list to avoid overpayment when it is made easier for them to do so.

Shocks to effective enforcement can be generated by a change in the identity of the regulator. In Zitzewitz (2002), I examine the effect of changes in SEC chairman on compliance with Regulation Fair Disclosure. I find that a proxy for compliance increases sharply when Regulation Fair Disclosure becomes effective but that most of this increase is reversed when Arthur Levitt (a proponent of Regulation Fair Disclosure) is replaced as SEC chairman by Laura Unger (an opponent). Here the research is informative about the behavior of both the regulated agents and the regulator itself.[12]

Shocks can also be generated by technologies that facilitate hidden behavior. Lin, Qian, and Liu (2008) examine the effect of a relaxation of abortion restrictions in Taiwan that coincided with a sharp growth in demand for ultrasound machines capable of determining fetus' sex. The authors find a sharp increase in the male share of births when abortion was relaxed, especially for older mothers and for third and subsequent children, who might have stronger incentives to sex-select in response to a son preference. They also find evidence of a partly offsetting relative improvement in the one-month mortality of third-born girls. They conclude that parents substituted prenatal selection for postnatal selection and that "ten more female infants

survived for every one hundred that were aborted" (abstract). Technology and a legal change facilitated one type of hidden behavior (sex-selective abortion, which is illegal in Taiwan) but reduced demand for a substitute (female infanticide or neglect). Qian (2008) provides another example of economic incentives affecting sex-selection. She finds that post-Mao agricultural reforms that increased the return to cash crops increased the male ratio in areas that produced fruit (at which men have a comparative advantage) but decreased the ratio in areas that produced tea (at which women have a comparative advantage).

Krueger and Mas (2004) and Mas (2008) examine the effect of shocks to worker–management relations on product quality. The former paper finds that defective Bridgestone/Firestone tires were disproportionately produced during times that concessions were being demanded or when the plant was employing both permanent and replacement workers. The latter paper finds that Caterpillar construction equipment produced during contract disputes trades at a discount on the used market and is more likely to be resold, suggesting quality problems. The Firestone example is especially serious, given that defective tires were linked to tread separations that caused 271 fatalities and over 800 injuries (254). The identification approach of these papers is to assume that other factors affecting quality vary continuously, but that incentives or desire to "punish" one's employer (and, ultimately, one's customers) by producing defective products increases discontinuously at times of unrest. A second approach examines discontinuities that have been built into incentives by the incentives scheme's designers, perhaps mistakenly. Oyer (1998) notes that many salespeople have incentive pay with discontinuities at fiscal year ends and finds a similar discontinuity in firm sales around fiscal year ends (as compared with other firms

---

[12] In a related study, Zitzewitz (2009) examines the recovery-to-harm ratio for settlements of mutual fund trading cases. In cases negotiated jointly by the SEC and the New York Attorney General, regulators achieved a ratio of 77 percent; in cases negotiated by the SEC without New York, the ratio was only 7 percent. The results provide additional evidence that the identity and disposition of the regulator has a significant effect on outcomes.

in the same industry). This is consistent with either increased effort before a year-end or, more problematically, with collusion between salespeople and customers to book next-year orders this year, distorting accounting revenue and earnings figures. Larkin (2007) finds salespeople at an enterprise software firm have a nonlinear compensation scheme that rewards concentrating sales in a single quarter and that salespeople not only concentrate their orders but appear to offer significant discounts to customers in exchange.

Keys et al. (2010) test whether securitization, and specifically the fact that mortgage lenders were not exposed to a loan's long-term credit risk, led to relaxed underwriting standards. The authors exploit the fact that while credit risk decreases continuously with borrower credit scores, the ease of securitizing a loan increases discontinuously at certain credit scores thresholds. The authors find that default probabilities are higher for loans that just above score thresholds that were important for a loan's securitizability. This is consistent with lenders using more relaxed underwriting criteria on non-credit-score signals of borrower risk when a loan was more readily securitized, and thus its risk more readily transferred to a third party. Bubb and Kaufman (2009) have challenged this interpretation, showing that discontinuities in default rates exist at score thresholds even for loan categories where there is no discontinuity in securitization rates. They argue that higher default rates above a score threshold may be due to lenders using credit score thresholds in choosing which loans to screen extensively. Another potential explanation is that a borrower with scores just above a threshold may be more likely to have managed their score to reach that threshold and may otherwise be more likely to be less creditworthy than their loan file suggests.

Slemrod (1985) notes that tax tables, in which tax owed is a staircase-shaped function of taxable income, create (small) incentives for (small amounts of) tax evasion. In the year studied (1977), tax liability jumped discreetly at the boundaries of $50 tax brackets for all but the highest income taxpayers. By underreporting taxable income by $1, a tax payer who would otherwise have been at the very bottom of one bracket could save $7–$21 in tax. Slemrod finds that 23.5 of taxpayers with marginal rates of 30–42 percent (the highest for which the staircase function applied) report an income in the top $10 of the $50 bracket, compared with 20 percent if incomes were distributed continuously, suggesting that 3.5 percent of returns in this income category engage in manipulation. The corresponding percentage for returns that do not use the tax tables is 19.7 percent, providing a nice falsification test. The percentage in the top $10 is increasing in marginal tax rate and is higher for tax returns for which taxable income is "fungible" (e.g., those with itemized deductions or self-employment income), consistent with evasion increasing with incentives and opportunities for it. Saez (2010) finds more recent evidence of clustering of incomes at kinks (i.e., discontinuities in the first derivative) in the tax schedule, again especially among the self-employed.

Work on earnings manipulation has also looked for evidence of incentive discontinuities that arise from the salience of certain earnings thresholds to investors with limited attention. Burgstahler and Dichev (1997) argue that while firms may always benefit from reporting higher earnings, these benefits should be discontinuous at certain thresholds: zero and last year's earnings. Degeorge, Patel, and Zeckhauser (1999) add analyst's median expectation of earnings as another focal threshold. Both papers find that firms are especially likely to report earnings that just beat these thresholds and unlikely to report earnings that just miss these thresholds, suggesting that firms manipulate earnings by transferring profits from one quarter to the next. Burgstahler and Dichev also find

that firms that just beat thresholds rely disproportionately on components of earnings that are known to be manipulable. Bhojraj et al. (2009) confirm this in more recent data and find that just beating a threshold with low earning quality is followed by positive short-term returns, insider selling and equity issuance, and negative longer-term returns. Grundfest and Malenko (2009) likewise find that firms that are especially unlikely to report earnings that imply just missing getting rounded up to the next penny per share are more likely to subsequently restate earnings or be sued for accounting fraud. Bollen and Pool (2009) make a similar finding for hedge funds: hedge funds are much more likely to earning barely positive than barely negative returns, and the former group of funds has lower future returns. Interestingly, the incentives for manipulation are reinforced by the fact that missing a threshold takes on a special significance in an environment where other firms manipulate to avoid missing them, missing causes investors to infer that a firm has exhausted its stock of deferred earnings.[13]

Other examples of incentive discontinuities come from sports.[14] Duggan and Levitt (2002) study Japanese sumo wrestlers, who face a strong incentive to win at least eight out of the fifteen bouts in a tournament. They find that a wrestler who needs to win his last match to reach the threshold is very likely to do so, and then very likely to lose his next match against the same opponent in a later tournament. This suggests that wrestlers trade wins in response to the nonlinear incentives.

College basketball teams who collude with gamblers face a similar incentive discontinuity. Most betting on college basketball is "point spread" betting, in which one wagers on whether a favored team will win a game by at least a certain number of points. A heavily favored teams may collude with gamblers who have bet against it and "point shave," or reduce its effort level late in a game to ensure that it wins by less than the point spread. Wolfers (2006) finds an asymmetric distribution of outcomes: heavily favored college basketball teams are much more likely to win by just less than the point spread than by just more. He concludes that "six percent of strong favorites have been willing to manipulate their performances" (283).[15]

Bernhardt and Heston (2010) challenge Wolfers's interpretation of his result, and ironically they do so by examining other incentive discontinuities. First, they point out that Gibbs (2007) finds a similar distribution of outcomes for professional basketball games, and they argue that bribery of professional players by gamblers is much less likely

---

[13] Other work on earnings manipulation and incentives includes Healy (1985), who finds that accrual policies that are consistent with attempts to maximize bonus payments and Bergstresser and Philippon (2006) who find greater use of discretionary accruals to manipulate earnings when CEOs have large stock and option holdings. Bergstresser and Philippon also find that CEOs and other insiders exercise more options and sell more stock when accruals are high. Bergstresser, Desai, and Rauh (2006) discuss cycles of earnings manipulation via assumed rates of return in defined benefit pension plans and executive stock option issuance and exercise.

[14] The study of sports is controversial, as some economists regard the lessons as unlikely to generalize to other settings. A more optimistic view is that studies of sports are like lab experiments that someone else goes to the trouble and expense of organizing and that have larger sample sizes and arguably more invested participants. Just as the best experimental economics studies are carefully designed to have as much external validity as possible, the best sports economics studies carefully choose settings in which we are likely to learn about economic behavior off the playing field.

[15] Snyder (2007) conducts a related test for Congressional elections, finding that incumbents in Congressional elections are much more likely to win than lose by a small margin. Snyder's setting differs from Wolfers' in that whereas favored basketball teams have no reason to care whether they cover the spread (absent side payments from gamblers), politicians have strong incentives to care whether they win. Snyder's results could potentially be explained by incumbents having better polling data and thus being more able to determine the precise amount of effort needed to win, although arguably the discontinuity Snyder finds is too large to be explained in this manner.

given high player salaries (Bernhardt and Heston, footnote 4). Second, they estimate point spreads for college basketball games on which bookmakers did not offer betting, and find a similarly asymmetric distribution of outcomes. Rather than finding that a proxy for hidden action does change with the incentives for it, they find that it does not. They argue that Wolfers's result may have a more innocent explanation, such as favored teams reducing their effort level with a safe lead. While Bernhardt and Heston's failure to find a particular cross-sectional result may be due to lacking statistical power, their paper provides an example of the importance of documenting both positive and negative results.

2.3.2 *Field Experiments*

An alternative approach to identifying hidden behavior is for the researcher to experimentally vary incentives for it. This approach has been most commonly used in studies of tax evasion. Slemrod, Blumenthal, and Christian (2001) examine the results of a Minnesota Department of Revenue experiment that sent some taxpayers a letter warning them that they would be audited. They found that low and middle-income sole proprietors who received the letter reported more income, but there was less evidence of an impact on wage earners, who as discussed above have their income reported to the tax authorities by third parties. This is consistent with the sole-proprietors who were not warned of an audit under reporting their income. Kleven et al. (2010) find consistent results from a similar experiment in Denmark. Pomeranz (2010) finds that audit threat letters sent to one firm increase the value added tax (VAT) payments of that firm's suppliers, consistent with the purported self-enforcement benefits of the VAT.

Olken (2007) applies a method similar to the tax auditing experiments to detect skimming in road building projects in Indonesia.

In his experimental intervention, government audits were announced in advance of a randomly selected set of road projects. Skimming on these projects and a control group of projects was then measured afterwards by independent audits of road quality—the materials that were purchased for a road project were compared with an estimate of materials used by independent engineers. The independent audits revealed that preannouncing a government audit reduced skimming by eight percentage points. The two types of audits play different roles in the context of our taxonomy: the preannounced audits provide experimental variation in expected enforcement, while the independent audits create multiple measures for the construction of an estimate of hidden behavior.

Duflo, Hanna, and Ryan (2008) provide an example of uncovering shirking by experimentally changing the level of monitoring. They evaluated a program that paid teachers bonuses if they attended school each day and took a picture of themselves with their students with a camera with a tamper-proof timestamp. Teacher absenteeism dropped to 21 percent (compared with 42 percent in control schools) and student test scores increased 0.17 standard deviations. The response of attendance and test scores to the auditing suggests that, absent a special program, teacher attendance is poor, existing incentives for attendance are weak, and low teacher attendance severely harms student performance.

2.3.3 *Continuous or Binary Variation in Incentives*

A third category of studies examines natural variation in incentives for hidden behavior, where confounding factors are a potential concern. One broad area where this issue arises is in testing for whether expert opinion is biased by (usually endogenously formed) political or commercial relationships or by

shared membership in an ethnic or other demographic group. In these cases, causality is often established more through a preponderance of evidence rather than a single, "silver bullet," statistical test.

For example, Khwaja and Mian (2005a) find that firms with politicians as directors are able to borrow 45 percent more from banks and are 50 percent more likely to default and that all of this extra lending and default is accounted for by government-owned banks. Apart from favoritism, the authors consider "social lending" as a possible alternative explanation: firms with politician directors may be more likely to be engaged in "social" activities that further government policy goals. They cast doubt on this alternative in a variety of ways, including showing that banks with explicit social lending goals actually lend less to politically connected firms. Faccio, Masulis, and McConnell (2006) also find evidence of favoritism towards politically connected firms, specifically that they are more likely to receive government bailouts that are ultimately financed by the World Bank or International Monetary Fund.

Turning to commercial relationships, Lin and McNichols (1998) and Michaely and Womack (1999) find that more positive stock recommendations are issued by analysts who work for brokerages that are affiliated with investment banks, especially those providing underwriting services to the firm in question. At least before the regulators investigations into analysts' conflicts, one might have wondered whether this correlation was partly driven by underwriters (who bear financial risk from failed offerings) seeking clients about which they have positive opinions. Michaely and Womack took an unusual approach to distinguishing between the bias and selection hypotheses, surveying business school graduates working in the industry. When provided with a summary of Michaely and Womack's results, 88 percent of the surveyed alumni chose "strategic conflict of interest" over "selection bias" as the primary explanation. While one cannot always count on the subjects of forensic economics research to confess their motives in anonymous surveys, when they do, it helps rule out alternative hypotheses.

Other papers have examined the impact of commercial relationships on the allocation of underpriced (or "hot") IPO shares in the 1990s. Liu and Ritter (2010) examine the behavior of CEOs who were revealed by a regulatory investigation to have received allocations of other firms' hot IPOs. They find that these CEOs were more likely to have underpriced their own IPOs and that they were much less likely to switch their next offering away from the underwriters who provided them the allocation. The authors interpret this as evidence of a quid pro quo in which underwriters use underpriced IPO shares to "bribe" other firms' CEOs to win their underwriting business. Reuter (2006) finds that mutual funds who do more brokerage business with an underwriting firm receive larger allocations of hot IPOs, but similarly sized allocations of cold IPOs. Subsequent work (Gaspar, Massa, and Matos 2006; Cici, Gibson, and Moussawi 2010; Massa, Reuter, and Zitzewitz 2010) has suggested that these underpriced IPO allocations are not allocated pro rata into the portfolios whose trading generated them. There are alternative reasons an underwriter might allocate to investors with whom it has commercial relationships; these investors might be less likely to sell cold offerings, insuring the underwriter against having to offer price support. This alternative explanation is inconsistent, however, with the allocation–relationship correlation being strongest for the hottest IPOs.

Similar issues arise with studies of advertising and media bias. Reuter (2009) finds that *Wine Spectator* provides higher ratings (relative to other publications) to the wines of its advertisers. One might wonder

whether this arises from tastes, e.g., from *Wine Spectator*'s editors and readers preferring certain types of wine and from the makers of these wines advertising to reach those readers. Reuter provides evidence that advertisers' wines are more likely to be retasted and have low initial scores adjusted upward, which is more consistent with a pro-advertiser bias rather than advertiser self-selection based on tastes. Reuter and Zitzewitz (2006) find that personal finance magazines are more likely to recommend the funds of their advertisers. Perhaps less so than with wine, personal finance magazine readers may have tastes for specific types of funds, and firms offering these funds may be more likely to both advertise in the magazines and get mentioned. Given that Reuter and Zitzewitz find that the advertising–content correlation persists after controlling for funds' objective characteristics (e.g., past returns, expenses), the overall level of advertising, and past media mentions, a tastes-based explanation would have to be specific to individual personal finance magazines and the tastes in question would have to be uncorrelated with observable characteristics and prior coverage. In more recent work, Rinallo and Basuroy (2009) provide evidence of a strong advertisement–content correlation in fashion magazines, and Di Tella and Franceschelli (2009) find that the Argentinean government reduces advertising in newspapers that cover government scandals more heavily.

Tastes as an alternative explanation arise in other studies of judging bias. In Zitzewitz (2006a), I find evidence of nationalistic bias in ski jumping and figure skating judging. This bias could be the result of country *A*'s judges favoring country *A*'s style of skating or jumping, but I also find that judges evaluate athletes from third countries more favorably when they are represented on the judging panel. This helps distinguish tastes from strategic bias: one might expect country *A*'s judges to have a taste for country *B*'s style,

but it is harder to explain why these tastes should vary with whether country *B* has a panelist (with whom to potentially exchange votes).

Tastes are also a potential alternative explanation for Price and Wolfers's (2010) finding of own-race biases in foul calls by NBA referees. One concern is that referees of a certain race may favor a style of play that is more common among their own race. For example, black players on average make more successful steals, and white players block more shots. Each play involves risking committing a specific type of foul (reaching in and hacking, respectively). If black referees monitor for hacking or white referees monitor for reaching in more aggressively, it could lead to referees calling more fouls on other-race players. Price and Wolfers find that their results are robust to controlling for players' positions and average statistics (including steals and blocks), however, casting doubt on style of play as the main explanation.[16]

Further doubt on a tastes explanation is cast by findings of own-race bias in ball–strike calls by baseball umpires (Parsons et al. 2011; Chen 2007). Ball–strike calls are one of the more objective judgments in sports judging; indeed, both papers control for the actual location of the pitch as measured by the Questec system of cameras. Parsons et al. find evidence of smaller or no bias when umpires are well-monitored (by Questec, a large crowd, or when the call could lead to a walk or strikeout), implying that umpires are adjusting bias, consciously or unconsciously, in response to monitoring.

---

[16] Price, Remer, and Stone (2012) find that NBA referees' discretionary fouls appear biased toward achieving game outcomes that would increase league profits (favoring home teams, keeping games close, extending playoff series). Their results could also reflect referees having tastes for these outcomes, although, here and in other cases, allowing tastes to affect judgments might be itself considered a form of bias.

In other studies, the variation in incentives is so large that it can reasonably be expected to swamp confounding factors. Levitt and Syverson (2008) find that houses owned by real estate agents sell for 3.7 percent more (controlling for observable characteristics) and take ten days more to sell. When a real estate agent works for a client, she typically keeps 1.5 percent of the selling price; when selling her own house, she keeps just under 100 percent. Thus an agent has a much stronger incentive to price her own house to obtain its full value, even if it requires more selling effort. While there may be alternative explanations for agents being more patient (perhaps they are less credit constrained than a typical seller), the differences in incentives are so large that they necessarily swamp any other economic reason for different seller behavior.

Another setting in which alternative explanations, while conceivable, are unlikely to be important in practice are studies of school behavior around the dates of high-stakes tests. Figlio (2006) finds that schools are more likely to suspend poorly performing students on test days, while Figlio and Winicki (2005) find that they serve higher-calorie food on testing days. This sort of high-frequency variation of policies in an apparent effort to game test results may influence our beliefs about a gaming motivation for other findings, such as an increased likelihood for marginal students to be held back after high-stakes tests are introduced (Figlio and Getzler 2002; Jacob 2005; Cullen and Reback 2006).

Cici, Gibson, and Moussawi (2010) study another such large incentive difference, comparing mutual funds run by firms that also offer hedge funds with those that do not. Hedge fund managers typically earn incentive fees of 20 percent of funds performance (above some threshold), while incentive fees for mutual funds are rare and comparatively tiny when they exist (Elton, Gruber, and Blake 2003). Managers who manage hedge and mutual funds side-by-side face a temptation to favor their hedge funds through legal or illegal means: when allocating shares in "hot" (i.e., underpriced) IPOs, when deciding which portfolio trades first, or when allocating trades among portfolios. Cici, Gibson, and Moussawi find evidence that side-by-side mutual funds underperform matched peer funds by about 1.2 percent per year. About 40 percent of this underperformance is accounted for by differences in the above-discussed return gap, which Kacperczyk, Sialm, and Zheng (2008) argue captures funds unobserved actions, including cross-portfolio favoritism. Cici, Gibson, and Moussawi also find that the side-by-side funds receive the same allocations of cold IPOs, but 40 percent smaller allocations of hot IPOs.[17] Of course, there can be differences other than incentives between the side-by-side firms and their peers that contribute to these results: managers of side-by-side funds may be less skilled traders, or less well connected and thus less able to get allocations of hot IPOs. The differences in outcomes are arguably too large to be explained by manager skill though, and it is far from obvious that mutual fund firms that also manage hedge funds should attract *less* skilled managers than their competitors.

A related approach is to test whether hidden behavior varies with opportunity, as opposed to motive. As discussed above, the self-employed have more scope for underreporting income on their tax returns. Pissarides and Weber (1989) and Feldman and Slemrod (2007) test whether the self-employed spend the same amount on food and charitable contributions, respectively, as employees with similar *reported* incomes.

---

[17] Cici, Gibson, and Moussawi follow Reuter (2006) in inferring IPO allocations from the first holdings disclosure following the IPO; Reuter helps validate this approach by showing that the correlation between commissions and IPO holdings is stronger the closer to the IPO date that holdings are disclosed.

The identifying assumption in the Feldman–Slemrod tests is that "the source of one's income is unrelated to one's charitable inclinations and that the ratio of true income to taxable income does not vary by income source" (abstract); Pissarides and Weber make an analogous assumption about food expenditures (excluding business-related meals). The studies find large differences: food consumption by the U.K. self-employed and charitable giving by the U.S. self-employed is consistent with underreporting by factors of 1.55 and 1.54, respectively. While one might wonder whether the self-employed have greater networking gains from involvement in charity (or work up a greater appetite), the differences the authors find are so large that they likely primarily reflect underreporting.

Studies of bias and incentives sometimes yield counterintuitive results. Hubbard (1998) finds that privately owned inspection stations are less likely to fail cars (controlling for observable car characteristics) than state-owned stations, despite the incentive created by the opportunity to do repair work after a failed inspection. The explanation appears to be that the incentive to both "diagnose and treat" is so salient to customers, so many customers assume foul play when failed, take the repair and future inspection business elsewhere, and spread negative word-of-mouth. Consistent with this, Hubbard (2002) finds that stations are more likely to return to a station when they passed last year and that stations with higher pass rates do more inspections in the future.[18] These

results, which suggest an optimistic view of consumer sophistication, generally contrast with those from a literature in health economics on physician-induced demand (see McGuire 2000, for a review). One example is Gruber and Owings (1996), who find that obstetricians in states with more rapidly declining birth rates performed cesarean deliveries at a higher rate, which they argue is consistent with physicians "exploit[ing] their agency relationship with patients by providing excessive care" (abstract).

Incentives that solve one agency problem can sometimes create hidden behavior as a side effect (Kerr 1975; Holmstrom and Milgrom 1991). Malpractice lawsuits provide incentives to doctors to practice carefully, but may also induce costly and potentially harmful defensive medicine (Kessler and McClellan 1996; Currie and MacLeod 2008). Chevalier and Ellison (1997) find that investor demand for mutual funds increases in performance but that the relationship is convex; demand is especially high for funds that finish at the very top of the league table. Chevalier and Ellison find that funds alter their portfolios late in the year in response to these incentives, increasing risk when their incentives become more convex. Additional examples are provided by the above mentioned studies of bonus management, earnings management, and manipulation of sales timing and SSRN downloads.

2.4 Model-Based Inference

In forensic economic studies that exploit variation in incentives, the identifying assumption is usually that incentives for hidden behavior are uncorrelated with other factors affecting the outcome variables (at least once observables are controlled for). In these studies, we have a (simple) model of how agents interact in the absence of the hidden behavior of interest, and identify hidden behavior by looking for deviations. In

---

[18] Gino and Pierce (2010) find that certain mechanics have persistent differential pass rates for luxury brand cars, which they attribute to wealth envy. Pierce and Snyder (2008) identify firm-effects in emission test pass rates, which they attribute to "ethical spillovers" within firms. Identification in these papers involves assuming that unobservable car characteristics are not correlated within firm or within mechanic-brand combinations in ways that would produce these results.

the next section, we turn to other approaches that involve testing for deviation from other models of behavior. These models include economic models, which derive predictions by assuming honest behavior by profit-maximizing agents that can be potentially falsified by the data. The models can also include statistical models, in which hidden behavior is identified by finding outliers.

2.4.1 *Economic Models*

Auctions are one commonly studied setting that lends itself to economic modeling. Most of the literature on collusion in auctions studies settings in which collusion is known to have occurred, but some of these papers propose methods that can be used to test for collusion prospectively. Porter and Zona (1993) examine a specific form of collusion in government procurement auctions: "complementary bidding" where a cartel of bidders agree on a auction winner and other cartel members submit higher bids in order to make the winning bid appear competitive. Their method involves estimating a logistic model to predict the lowest bidder in an auction, estimating an ordered logit model to predict the ranking of all bidders, and then testing if the coefficients in the two models are equal. This test exploits the fact that firms should play mixed strategies in auctions, and thus their bids should be unpredictable (by other bidders and by the econometrician) conditional on observable variables.[19] If firms' bidding behavior appears related to their costs differently in auctions they win than in auctions they do not, it suggests

something observable to the firm, but not to the econometrician, that predicts whether one will win an auction. A collusive scheme is one, and arguably the leading, candidate for such an unobservable.

Porter and Zona (1999) find a different deviation from the competitive model in their study of a cartel in public school milk procurement auctions in Ohio. They exploit the fact that milk has relatively high shipping costs and thus spatial concerns are important. They find that Cincinnati firms (who were found by regulators to have colluded) had made lower bids for distant contracts and higher bids for local contracts. This behavior "is consistent with local monopoly power, but local monopoly power in Cincinnati is consistent only with collusion" (287). Pesendorfer (2000) analyzes two contemporaneous milk cartels: one that arranged for the low cost producer to win auctions and shared rents with side payments (in Florida) and without side payments (in Texas). He finds the bidding behavior to be consistent with the low cost producer winning in Florida (which higher cost players could tolerate if they were compensated with side payments) and with contracts being shared, sometimes with higher cost providers, in Texas.[20]

A bidding ring often conducts an internal "knockout" auction to determine which member will bid seriously in the main auction (and what side payments they will make to the other members in exchange). Asker (2010) analyzes data from the knockout auctions conducted by a stamp bidding ring that was investigated by the New York Attorney General. In knockout auctions, a bidder who does not expect to win has an incentive to overbid her valuation in order to increase

---

[19] The econometrician is assumed to have observed everything the firms have and have correctly specified the functional form. Porter and Zona assume that bidder values are private (i.e., there is no winner's curse) and independent up to a single auction fixed effect. If cartel firms had especially correlated costs (e.g., due to being located near each other), this could produce bidding that appears complementary.

[20] Baldwin, Marshall, and Richard (1997) and Bajari and Ye (2003) present frameworks in which one can test whether bidding behavior is more consistent with competitive bidding or a specific collusive model.

side payments received. Asker finds that two ring members rarely won competitive knock-out auctions and appeared to be participating in the ring mainly to receive these payments. Ideally, the ring would like to exclude these participants, but in practice it may be difficult to given that they may then face an incentive to confess in exchange for leniency. Asker also estimates that the incentive to overbid in the knockout auction in some cases may have caused the knockout winner to bid above her valuation in the final auction. This contrasts with Kwoka's (1997) study of a real estate knockout cartel, where he found that collusion significantly depressed prices.

Beyond auctions, a common approach to studying collusion in product markets is to estimate a so-called conduct parameter. The conduct parameter captures the extent to which firms include the profits of their competitors in their objective function. A common formulation is to assume that firms satisfied the following first-order condition:

$$c = P + \theta Q P'(Q),$$

where c was the firm's marginal cost, Q was industry-wide quantity, and $P(Q)$ was an industry-wide price level. The parameter θ captures the firm's conduct: $\theta = 0$ represents pricing at marginal cost, θ equal to the firm's market share represents Cournot behavior, and $\theta = 1$ represents the fully collusive case (where the firm acts to maximize industry profit).

The parameter θ can also be interpreted as the firm's conjecture about the elasticity of industry output with respect to its own output, so this approach is also called "conjectural variations." A $\theta = 1$ would imply that firms behaved as if they expected a given percentage change in its own output to be matched by all competitors. As the conjectural variations framing emphasizes, a high θ could result from tacit as well as explicit collusion. Charging a high price or restraining

output in the belief that competitors will respond in kind is likely commonplace, and is not illegal if there is no communication.

Genesove and Mullin (2001) provide an example and validation of the conduct parameter approach.[21] They study the sugar refining industry in the 1890–1914 period, which included episodes of entry and price wars, the establishment of a cartel, and the gradual erosion of cartel market share through the expansion of a competitive fringe. The authors find that their estimate of θ parameter was highest during periods of cartel stability, was lowest during the price wars, and declined slowly as the competitive fringe grew. It also matched the evidence from price–cost margins, which could be estimated readily in this industry due to its simple production technology. Wolfram (1999) conducts a similar exercise for the British electric power industry, also finding that margin-based and conduct parameter measures of market power are in rough accord.

These methods have been critiqued as sensitive to the correct specification of the model. Porter (2005) notes that tests for collusion depend on correctly specifying the collusive model, since bidders could always modify their behavior to simulate the behavior a given model would expect from competitive firms with high costs. Corts (1999) has critiqued the conduct parameter approach as sensitive to assumptions about the model of competition and provides simulations of repeated-game collusion in which a conjectural variations approach badly overestimates the intensity of competition. All empirical approaches are potentially sensitive to their maintained assumptions. A cost of more complicated methods is that these sensitivities are often less transparent

---

[21] Bresnahan (1989) provides a review of this approach; Corts (1999) cites many subsequent papers from the 1990s.

than when examining simple correlations of behavior and incentives.

Model-based inference for forensic purposes is conducted outside industrial organization. As discussed above, Knowles, Persico, and Todd (2001) test whether racial disparities in search rates for motor vehicles can be rationalized given racial differences in the likelihood of success (e.g., in finding drugs). Gordon (2009) takes a related approach to testing for partisan bias in public corruption prosecutions. He finds that, conditional on being prosecuted and found guilty, defendants from the prosecutors' party receive heavier average sentences. Assuming sentencing judges do not have partisan biases that are negatively correlated with the prosecutor's, this suggests that prosecutors prosecute only the most serious cases against their co-partisans. In both studies, the authors infer a bias if the average returns to searching or prosecuting one group differ, with the assumption being that equal average returns implies equal marginal returns. Knowles, Persico, and Todd discuss this issue at length (212–15), arguing that so long as no "type" of motorist is searched with probability one, the assumption is valid since all types are marginal.[22]

The economic models in model-based inference are usually, but not always, formal. My own work on the stale-price arbitrage of

mutual funds provides an example of identifying hidden behavior based on deviations from an informal model of honest behavior. U.S.-based mutual funds used to calculate the net asset values using closing prices, even when those prices were from exchanges located in other time zones. This, together with the fact that funds allowed trading until 4 PM U.S. time, created an arbitrage opportunity that was being heavily exploited. In Zitzewitz (2003), I calculated that international equity funds lost an average of 1.1 percent of their assets in 2001 to stale price arbitrage. A sizable number of hedge funds were organized with "mutual fund timing" as their main strategy (Boudoukh et al. 2002, footnote 7).

This trading activity is legal for the trader, so the forensic component of the paper was understanding why mutual funds tolerated it in such large scale. In the paper, I noted that while the SEC was encouraging mutual funds to update foreign closing prices using a process known as "fair value pricing," mutual funds were largely attempting to address arbitrage trading through short-term trading fees and trading restrictions. I noted that the former method was cheaper than the latter and had the advantage of simultaneously denying the arbitrage to everyone, whereas trading fees and restrictions could be evaded in numerous ways. I argued that a preference for the latter methods could be rationalized if the firms were receiving side payments from certain arbitrageurs, and thus valued the ability to selectively deny opportunities for arbitrage. Subsequent regulatory investigations revealed that at least 20 percent of the industry (by assets; Zitzewitz 2009, table 8) had indeed either taken side payments from arbitrageurs or had allowed senior managers to arbitrage trade in their own funds. While there was no formal economic model in this paper, the basic approach is the same. The evidence of hidden behavior

---

[22] Other work has studied police profiling by examining variation in officers' incentive or ability to profile. Grogger and Ridgeway (2006) compare the racial mix of drivers stopped just before and after sunset (using seasonal variation in sunset time to control for clock time) in Oakland and find it is identical, suggesting that Oakland police do not condition traffic stops on driver race (for either statistical or taste reasons). In contrast, Antonovics and Knight (2009) find that Boston police officers are more likely to stop opposite-race drivers, which they argue is likely consistent with racial differences in officer tastes. Makowsky and Stratmann (2009) find that Massachusetts police are much more likely to let local drivers off with a warning. They do not find large racial differences in ticketing rates, but do find that female drivers, especially young females, are much more likely to be let off with a warning.

came from actions that appeared inconsistent with honest agents' incentives.

2.4.2 *Statistical Models*

The other broad model-based approach is a statistical approach of looking for events that should be rare assuming a given model of honest behavior. Jacob and Levitt (2003b) look for a specific form of cheating in which teachers selected a group of questions from the middle of a test and changed their students' answers to the answer the teacher deemed correct. This method of cheating seems at first glance to be quite clever, since the teachers avoid raising scores to suspicious levels by choosing a subset of questions and they economize on effort by correcting answers from the harder questions near the end of a test and the same questions across students.

The cheating teachers' economizing on effort turns out to facilitate their detection though, since the authors can identify cheating by finding cases of "suspicious answer strings." In particular, they look for blocks of questions in which many students in the same classroom had identical (and not necessarily correct) answers. The authors also looked for cases where student answers were highly correlated within classrooms, where the degree of correlation was higher for some questions than others, and where the students were more likely to correctly answer easy questions than hard ones. The authors validated their "suspicious answer strings" measure of cheating by showing that students in these classrooms had large score improvements from prior years and that these improvements were much more likely to mean revert the next year. They also found correlations between measured cheating and proxies for the costs and benefits of cheating. Measured cheating is higher for classrooms with low prior performance, but lower for classrooms with multiple grades

(which would require cheating on two different tests).[23]

The Jacob and Levitt study provides an example of testing for fraud by looking for patterns that are very unlikely under an assumed model of honest behavior. Cramton and Schwartz (2000) find that bidders of analysis of collusion in the FCC spectrum auctions signaled desired area codes via the last three digits of opening round bids; these area codes appeared far more often than if the last digits of bids were random (as they should be since bidders should play mixed strategies). In related work in political science, researchers have tested for election fraud by exploiting biases people display when "making up numbers"—overusing certain digits, using repeated digits too rarely, and using consecutive digits too often. Beber and Scacco (2008) find that the last two digits in results from the 2003 Nigerian and 2009 Iranian elections display these biases, but those from the 2002 Swedish and 2008 U.S. elections do not.[24] In statistics, more complicated techniques have been used to detect bank fraud, money laundering, telecommunications fraud, computer intrusions, medical insurance fraud, and even student plagiarism (Bolton and Hand 2002). In general, fraud detection in statistics takes a more agnostic approach to what honest and

---

[23] In a subsequent paper, Jacob and Levitt (2003a) convinced the Chicago Public Schools to conduct proctored reexaminations in a random sample of classrooms, including those of teachers suspected of having cheated. They found that proctoring reduced scores for suspected cheating teachers, but not for teachers whose scores had been anomalously high but did not show suspicious answer strings. The effect of proctoring on scores for suspected cheaters helps confirm the results of the original study, and provides another example of uncovering hidden behavior by experimentally varying monitoring (see section 2.3.2 above).

[24] The results for Iran are reported in Beber and Scacco. "The Devil is in the Digits," *Washington Post*, 6/20/2009 (available at http://www.washingtonpost.com/wp-dyn/content/article/2009/06/20/AR2009062000004.html, last accessed 2/2/2011).

hidden behavior one should expect and looks for rare events, while economists usually take approaches that take advantage of an understanding of the incentives for honest and hidden behavior. The former approach is often applied in settings with very large datasets, where the efficiency costs of the more agnostic approach are mitigated and thus the gains in robustness are worth the trade-off.

The *Dartmouth Atlas* (1998), which has created a dataset of medicine practices in different areas of the country, has provided data used to identify medical fraud through analysis of rare events. Data from the *Atlas* helped uncover overuse of open-heart bypass surgery in Redding, California, of angioplasties in Elyria, Ohio, and a general high utilization in McAllen, Texas.[25] In the Redding case, Tenet Healthcare paid a $54 million fine to settle charges of performing unnecessary procedures. Silverman and Skinner (2004) use data from the *Atlas* to examine the practice of "upcoding": intentionally misdiagnosing one condition for another that allows for a higher reimbursement from insurance or Medicare. The authors compare the ratio of high-reimbursement diagnoses to low-reimbursement, but closely related, diagnoses. They found that for-profits hospitals, especially those that had recently converted to that status, were much more likely to diagnose in the higher-reimbursement categories and their propensity to do so increased from 1989 to 1996 and then decreased sharply when an investigation was announced.

Ashenfelter, Harmon, and Oosterbeek (1999) propose a test of whether research results are affected by a "file drawer" bias in which results are selected for publication based on their statistical significance. They examine 96 estimates of the coefficient of wages on schooling from twenty-seven different studies, and test whether coefficients and standard errors are correlated. The intuition behind this test is that if all coefficients are estimates of the same true value and vary only due to differences in sample, then the point estimates should be uncorrelated with the standard errors. If the coefficients are instead selected based on the t-ratio, then we should expect a positive correlation. The authors find evidence of a reporting bias and find that meta-estimates of returns to schooling drop when bias corrected. Donohue and Wolfers (2005) make a similar finding for estimates of the deterrent effect of the death penalty. Publication bias has also received considerable attention in the medical journals; Dubben and Beck-Bornholdt (2005) even conduct an analysis of publication bias in studies of publication bias (they do not find any).

2.5 *Market Efficiency*

A final category of forensic economics studies identifies hidden behavior using null hypotheses derived from financial or price theory. In efficient financial markets, future asset returns should be uncorrelated with currently available public information and with decisions that are conditioned on this information.[26] The work on stock option backdating discussed above provides an example of a study that exploits this property to understand the timing at which decisions were made. The studies show that option grants represented as having occurred at time t are correlated with stock returns between t and $t + s$, which suggests that decisions were really made at $t + s$. As Heron and Lie

---

[25] For Redding, Elyria, and McAllen, see Pollack, Andrew, "California Patients Talk of Needless Heart Surgery," *New York Times*, 11/4/2002, p. C1 (http://www.nytimes.com/2002/11/04/business/california-patients-talk-of-needless-heart-surgery.html), Abelson, Reed, "Heart Procedure is Off the Charts in an Ohio City," *New York Times*, 8/18/2006, A1 (http://www.nytimes.com/2006/08/18/business/18stent.html), and Atul Gawande (2009), respectively (all last accessed 2/2/2011).

[26] See Schwert (1981) and MacKinlay (1997) for reviews of the event study methodology.

(2007) discuss, the fact that the correlation disappeared when the SEC began requiring timely disclosure of stock option grants helped rule out the inside information explanation in favor of the backdating explanation.

Zitzewitz's (2006b) work on late trading of mutual fund shares provides another example of the timing of a decision being misrepresented and of correlations with asset returns being used to establish the actual timing. Unlike the "market timing" discussed above, which involved trading mutual funds to exploit biases in valuations, late trading is illegal. Mutual fund late traders colluded with brokerage employees to make trading decisions after 4 PM Eastern Time but to receive pricing based on 4 PM asset values. One technique was for brokerage employees to timestamp trading tickets before 4 PM but then destroy the tickets if market moves after 4 PM made the trade turn out to be disadvantageous. The time-stamped trading tickets satisfied traditional audits, since the stamps were all before 4 PM. But the uncanceled trades were correlated with post-4 PM market movements, which was suspicious for exactly the same reason that positive returns after option grants were. As with option backdating, the main alternative explanation was that the apparent late trades were actually trades made by individuals who were able to predict future returns. Once again, the fact that the correlation disappeared after investigations into late trading were announced helped rule out informed trading as the likely explanation.[27]

A second application of financial market efficiency in forensic economics has been to use asset prices to measure investors' expectation of hidden behavior. A series of papers examine whether investors expected companies to benefit from connectedness to politicians who suddenly gained or lost power: Suharto (Fisman 2001), Richard Cheney (Fisman et al. 2006), George W. Bush (Knight 2006; Goldman, Rocholl, and So 2009), ministers or members of parliament in many countries (Faccio 2006), the majority party in the U.S. Senate (Jayachandran 2006), hometown politicians (Faccio and Parsley 2009), and Adolf Hitler (Ferguson and Voth 2008). Generally, these studies have found that equity prices increase in response to sudden increases in connectedness, suggesting that investors expect connected firms to benefit from favoritism. Another example is work finding that the stock prices of arms manufacturers rise when conflicts intensify in countries under a U.N. embargo (DellaVigna and La Ferrara 2010). These increases are particularly large for firms that produce small arms and ammunitions and for firms located in countries perceived as more corrupt, suggesting that investors expect the output of these firms to be more likely to find its way to the embargoed country. Both the arms embargo and the connectedness papers measure investor expectations about hidden activity, and so rely on the efficient market hypothesis twice—once in interpreting returns around events as reflecting changes in expected profits and again in assuming that investor expectations of profitable hidden behavior are rational.

A third application is to examine price changes and volume before public disclosures of news in order to understand the extent of insider trading on this news. Many studies have documented anticipatory returns and abnormal volume in advance of

---

[27] It should be noted that informed trading was a less plausible alternative hypothesis for late trading than for option backdating. The mutual fund trades were only correlated with future returns between 4 and 9 PM, which regulatory investigations had revealed was the latest time that brokerage employees were able to cancel trades.

Furthermore, the mutual fund trades were correlated with returns on general stock market indices. It is presumably more plausible that executives would have inside information about the future returns of their own companies than a mutual fund trader would have inside information about future general stock market returns.

corporate news announcements (e.g., Beaver 1968). More recent work has focused on a narrower population of agents. Ziobrowski et al. (2004) find that the stock trades of U.S. Senators were correlated with future returns. Cohen, Frazzini, and Malloy (2008 and 2010) find that mutual funds managers and analysts perform better when investing in or opining on the stocks of their classmates. Dube, Kaplan, and Naidu (2011) find positive returns and abnormal volume in advance of U.S.-authorized coups affecting companies at risk of nationalization, suggesting that even the most sensitive classified information is traded upon.[28] Khwaja and Mian (2005b) find that brokers in Pakistan appear to engage in a "pump and dump" scheme in which they take a position in a security, trade among themselves at gradually escalating prices, and then sell at inflated prices to return-chasing investors. Frieder and Zittrain (2008) find evidence of price and volume changes in lightly traded U.S. stocks around the dissemination of unsolicited emails that are also consistent with "pump and dump" schemes. Touted stocks appreciate significantly in advance of the unsolicited email, peak when the touting does and then decline sharply. In all these studies, the correlation of current activity with future asset price changes appears anomalous precisely because efficient market theory casts doubt on more innocent explanations.

A fourth application tests whether firm executives are acting in their shareholders' best interests by examining the short and long-run stock market response to decisions. Identification in these studies is complicated by the fact that stock market returns after a decision reflect both the impact of the

decision on firm value and changes in investors' beliefs about other aspects of the firm's situation. Settings must be chosen carefully for event returns to be informative about the causal effect of a decision on firm value. Yermack (2006) finds that companies that disclose the personal use of a company jet as a CEO perk experience negative announcement and subsequent returns. He also finds that this perk is often associated with a CEO playing golf at courses a long distance from company headquarters (as observed in the United States Golf Association handicap database). Malmendier and Tate (2009) find that firms with CEOs who win media awards suffer negative subsequent returns. They also find these CEOs are more highly compensated and engage in more outside activities (board memberships, book writing) and attribute this to an adverse "superstar" effect. In both examples, it seems unlikely that the firm's decision (buying the CEO a jet, lobbying for a CEO award) is a signal of the firm's prospects being poorer than previously disclosed, which increases our confidence that the sign of the effect on firm value is correctly identified.

Hsieh and Moretti (2006) use as their null hypothesis a prediction of efficient *product* market theory: namely, the law of one price. They examine the gap between the price charged by Iraq under its Oil for Food program and the price of free-market oil, adjusted for transportation costs. They calculate that Iraq underpriced oil sold through the program by $1.3 billion. The bulk of the underpricing occurred during time periods in which buyers were disproportionately individual traders rather than multinational oil companies; the authors argue that this is consistent with the traders being more willing to pay bribes. Once the United Nations assumed control of the oil price, underpricing disappeared, suggesting that it was not due to quality differences or other alternative explanations.

---

[28] Related work has examined the price impact of illegal insider trades identified by regulators (e.g., Meulbroek 1992) and has tested whether publicly disclosed trades by corporate insiders exploit inside information (Seyhun 1986).

3. *Common Themes*

Having organized the above review by methodological approach, in this section, I discuss some of the common themes from the studies' substantive conclusions. I will again focus on common themes that cut across subfields, as summaries of field-specific conclusions have been well covered in the reviews cited above. I then turn to some of the higher-level issues that the relative success of academic forensic economics work raises.

3.1 *What Do We Learn from Forensic Economics?*

I will highlight five common themes from the work reviewed above: the failure of government, the success of incentive theory, the persistence of racial and other in-group favoritism, the utility of using small-scale hidden behavior as a predictor of larger-scale hidden behavior, and the power of small interventions.

The fact that government failure emerges as a theme from forensic economics work is probably not a surprise, given that concern about corruption is a major motivation for the work in development. Perhaps more surprising is the evidence about the performance of developed country or international institutions, with examples ranging from the Chicago Police to the SEC, the United Nations, local governments in the United States, and even the governing bodies of several major sports.

The evidence of government failure is particularly worrisome given the evidence about the performance of the media, usually thought to be the primary check on government failure in modern democracies. An advertiser-supported media faces significant disincentives to raise issues that conflict with their advertisers' interests. This is true both when the government or state-owned firms are major advertisers or even owners,

or when both government and the media are potentially influenced by the same interest group. As discussed in the introduction, the imperfect performance of the government and media are precisely what creates the opportunity and, arguably, the need for academics to get involved in forensic work to begin with.

A second recurring theme from the review is the power of incentives in explaining hidden behavior. In Becker's (1968) seminal economic model of crime, potentially criminal agents trade off the returns to crime with the probability of and punishment from being caught. In many, if not most, of the studies discussed above, we observe hidden behavior increasing with the incentives for it, and declining sharply with increases in scrutiny or enforcement. Indeed, as discussed, testing for these changes is one of the major techniques used to identify or confirm the existence of hidden behavior in the first place.

With that said, it has long been recognized that a model of the decision making leading to crime or other unethical behavior has to allow for factors such as tastes or misperceptions, since many individuals routinely abstain from behavior that would otherwise be utility maximizing and some engage in criminal behavior that is difficult to rationalize, given the high probability of punishment. There is likewise some behavior observed in the surveyed studies that is difficult to rationalize. Given the wealth levels of the individuals involved, the clearest examples come from finance. Stock option backdating was engaged in by wealthy individuals for relatively small gains, despite the fact that it constituted both accounting and tax fraud. In Zitzewitz (2006b), I note the fact that late trading usually added little to the expected return to an already profitable (and legal) stale-price arbitrage trading strategy, yet about 60 percent of the dollars engaged in stale-price arbitrage were also engaged in late trading. The use

and acceptance of hot IPO shares as apparent bribes for investment banking business likewise involved relatively small financial gains for running fairly serious legal and reputational risks. But notwithstanding that caveat, the meta-field of forensic economics owes a significant intellectual debt to the economics of crime.

A third common theme is the persistence of racial and other forms of favoritism, in settings as varied as housing, hiring, and sports judging. Some of the evidence is plausibly of statistical discrimination, while other cases appear more likely taste-based. One approach taken to distinguish between the two was a version of the outcomes test advocated by Becker (1957). If black motorists are searched more frequently by police despite lower search success rates, as Anwar and Fang (2006) find in Florida, this suggests that tastes may contribute to the higher search rates. If instead searches of black motorists have higher success rates, as in Maryland, then, at least under the assumptions of Knowles, Persico, and Todd (2001), the higher search rate reflects statistical discrimination.[29]

A second approach taken to identify taste-based discrimination is to attempt to eliminate statistical discrimination by carefully controlling the information available. Audit studies test for discrimination by presenting agents with job candidates or real estate buyers with identical nonracial characteristics. Resume or test grading studies control even more carefully the information available. A third approach studies the relatively rare settings in which agents make objective evaluations of performance (e.g., ball and strike calls), the researcher observe the relevant facts as well as the agent, and thus the study can directly compare agents' evaluations with an objective measure (e.g., of pitch location).

An important caution about these studies is that disparate treatment may not lead directly to disparate outcomes. In particular, settings in which potentially discriminatory agents face competition (e.g., employers, real estate, and autos) at least offer consumers the ability to shop for less discrimination; such shopping should be most helpful when discrimination is taste-based, but tastes for it are not universal.[30] In contrast, settings in which the potentially discriminatory agent is a monopolist (e.g., judges, prosecutors, police officers, test graders, umpires) arguably deserve the most scrutiny.

A fourth common theme is the correlation of small-scale (and often legal) misdeeds with larger-scale malfeasance. Examples include diplomats' (legally) unpaid parking tickets being correlated with serious home-country corruption, rounding down on tax returns or up on corporate earnings being correlated with more serious fraud, and the presumably run-of-the-mill misbehavior that accompanies an industrial dispute being correlated with the making of defective tires that killed drivers in rollover accidents. The policy implications of these findings are not always straightforward, however, as a policy of auditing round-number tax returns may simply induce the abandonment of small-scale cheating by the large-scale cheats.

A fifth common theme is that relatively inexpensive interventions can often have large impacts. The simple health inspection

---

[29] Crucial assumptions in Knowles, Persico, and Todd's model are that every observable driver type is searched with probability less than one and that police equalize marginal success rates across driver types. If instead there is a type of black motorist that police search at every opportunity, then high success rates for this type could drive up average success rate for all black motorists, masking taste-based discrimination against other types of black motorists.

[30] Consistent with shopping partially alleviating discrimination, Charles and Guryan (2008) find that racial wage gaps are most negatively correlated with the prejudice of a state's *least* prejudiced members (specifically, with the 10th percentile of a survey-based prejudice).

score disclosure policy studied by Jin and Leslie (2003, 2009) had large health effects. The fairly inexpensive teacher attendance monitoring scheme studied by Duflo, Hanna, and Ryan (2008) had large effects on student performance. Olken's (2007) study of the effect of audits on materials theft is likewise encouraging. The scrutiny created by forensic economic papers themselves has likewise had an impact in some cases, which I turn to in the next subsection.

3.2 Uneven Policy Impact

One of the prime ways for forensic economics to make a social contribution is to affect policy for the better. While many of the studies discussed above do not have immediate policy implications, impact has varied widely for the ones that do. Macey and O'Hara (2009) discuss this variation, focusing on three cases in which policy reacted to forensic economics findings (Nasdaq market making, mutual fund trading, options backdating) and one where it initially did not (Senate insider trading). Ritter (2008) emphasizes the role of the press in disseminating findings and creating pressure for policy changes, while Macey and O'Hara acknowledge the role of the press but also emphasize regulators' incentives. Regulators other than the SEC became involved in the first three cases (the Department of Justice, the New York Attorney General, and the IRS, respectively), while Macey and O'Hara argue that the SEC might have had limited incentive to investigate the Senate, which appoints its commissioners and sets its budget.

Beyond these four studies, impact has been likewise mixed. Jin and Leslie's (2003) and Simon et al.'s (2005) evidence of the hygiene improvements and hospital admission reductions due to publicly displayed grading cards in Los Angeles helped motivate their adoption by a few states (North Carolina, South Carolina, and Georgia) and cities (San Francisco, St. Louis, New York).

This is a success story, but the fact that these did not become universal after their findings were disseminated is arguably a failure. In contrast, the dramatic increase in resources reaching schools in Uganda after the newspaper campaign publicizing the level of theft is impressive. One often-critiqued feature of the press is its limited attention span; if policy impact depends on the press, there is potential for backsliding once initial attention fades. Many of the reforms aimed at addressing broader investor–manager conflicts in mutual funds proposed by the SEC after the 2003 mutual fund scandal were subsequently weakened or dropped when media attention faded (e.g., compensation disclosure, independent chairs for fund boards). A follow-up study in Uganda to assess the permanency of the media's effect there would be useful.

Another form of policy impact from forensic economics is the arming of regulators and investigators with new tools. These tools are most useful when they are simple enough for nonexperts to implement and when the agents under study cannot modify their behavior in a way that foils detection. Methods that capture the full impact of the hidden behavior, rather than a symptom of that behavior, do better on this second dimension. For example, there is little incentive to do option backdating, mutual fund late trading, or insider trading in a way that avoids a correlation with stock returns, since it is this correlation that makes the activity profitable in the first place.

In contrast, using small-scale fraud as a predictor of larger-scale fraud may only work until the practitioners learn to avoid the small-scale fraud, as discussed above. Likewise, approaches involving multiple measures may not be as robust if the second measure can be manipulated. For example, if China tried to detect smuggling by comparing its import statistics with Hong Kong's export data, smugglers could presumably start misreporting to Hong Kong. Second

measures based on direct observation or auditing (e.g., of road thickness or of rice that actually reaches villages) may be harder to manipulate, but only if they remain under the control of incorruptible institutions. Studies based on incentive correlations or discontinuities might be likewise vulnerable to manipulation. Point shavers may start losing games by even more points, and agents might steal more even when incentives to do so are weak to avoid suspicious correlations. A specific model-based approach is often quite easy to fool given knowledge of the model being used, so this approach often depends in practice on keeping precise methods secret from agents. The exact approaches taken to detect bank fraud, for instance, are closely guarded secrets for this reason. Unfortunately, governments, and particularly democracies, are not always successful at keeping secrets. A race between the development of techniques by investigators and their discovery and evasion by agents may result, with the equilibrium effectiveness of forensic techniques depending on many factors, including the relative sophistication of the two sides.

3.3 *Reaction of the Research Subjects*

There is also significant heterogeneity in how the subjects of forensic economics have reacted to it. For many studies, especially those of rogue individuals, this issue is often moot, since these individuals are either anonymous (tax evaders) or already convicted (collusive dairies) by the time of the research. In other cases though, the subjects of the research are governments or large firms or institutions, often in the countries where the economists reside or conduct their field work.

One possible dimension of reaction is expanding or restricting one's data sharing policy. Forensic economics might create incentives for firms or governments to withhold data, or share it more selectively,

in ways that harm economic research more generally. On the other hand, it might encourage parties with an interest in limiting hidden behavior, such as the principals of the agents who might engage in it, to increase disclosure, essentially partnering with forensic economists in the process. An encouraging example is the Chicago Public Schools engaging Jacob and Levitt (2003a) to help them catch cheating teachers; the school system apparently correctly viewed themselves as one of the parties being cheated. A less encouraging example is the International Skating Union reacting to the vote-trading scandal at the 2002 Olympics by first deleting country identifiers from their score sheets and then by making judges' scores anonymous, a change which was accompanied by an increase in judging biases (Zitzewitz 2010). Another is Thomson Financial's new policy of refusing to share with academics the files that identify specific brokerage firms and analysts in their data following Ljungqvist, Malloy, and Marston's work and the decisions of Merrill Lynch and Lehman Brothers to ask Thomson to remove their data from the version of I/B/E/S distributed to academics.[31] In these cases, the principals with an interest in restraining hidden activity (consumers of figure skating or I/B/E/S data) are not well organized, and the parties that could intervene (TV networks, corporate subscribers to I/B/E/S) may interpret their self interest as not being aligned with transparency.

Subjects of forensic economics can react in other ways. One approach might be characterized by the Washington saying "admit nothing, deny everything, make counter

---

[31] Subscribers to I/B/E/S via Wharton Research Data Services can view notices of the I/B/E/S policy change at http://wrds.wharton.upenn.edu/ds/ibes/translation1.shtml and http://wrds.wharton.upenn.edu/ds/ibes/translation2.shtml. Merrill Lynch's and Lehman Brothers' decisions are discussed at http://wrds.wharton.upenn.edu/news/sideitem/user2007/analyst_data.pdf (p. 21).

accusations." A director at a large mutual fund company complained to the publisher of *Financial Analysts Journal* about the publication of Boudoukh et al. (2002).[32] The letter conceded that the article "may be 'technically correct'" but complained that the article had "ethical shortcomings" because it "outlines how investors can profit [from long-term mutual fund shareholders] by taking advantage of pricing differences in international mutual funds" and called the decision to publish the article "abhorrent." The letter also claimed that "the use of 'fair value pricing' has effectively closed the arbitrage," which was false at the time it was written (in late 2002) and indeed in late 2009 is still not completely true.[33] The publisher replied in part that publishing technically correct but potentially controversial articles was part of its mission and stated that "fostering open discussion of controversial issues encourages research that will benefit the profession and investors in general." The authors subsequently replied that "the real issue is what the mutual fund industry will do to curtail such activity" and noted that few if any funds were disclosing the issue to their investors.[34] The NBA responded to Price and Wolfers' results about racial biases by commissioning a report of their own, in which they reached opposite conclusions. The report was never released publicly, and Wolfers has claimed that it actually confirms the findings of the original study, if the results are properly interpreted.[35] Thomson Financial initially called the Ljungqvist, Malloy, and Marston (2009) study "a hatchet job" and claimed that the authors did not understand the I/B/E/S data, a claim inconsistent with the extensive corrections they subsequently made to their data to address the identified issues.[36] All else equal, researchers and journal editors would likely prefer to avoid fights, particularly with well-financed entities, and so these responses may have some of their intended chilling effect on forensic economics work.

More serious reactions are unfortunately conceivable. The willingness of researchers to publish results that are likely unpopular with governments in which they do their research is commendable, particularly in cases of governments that are less than respectful of the rule of law.

3.4 *Missed Opportunities and Next Steps*

This review discusses many instances where economists uncovered or quantified hidden behavior of both scholarly and societal importance. Of course, time spent on forensic economic research has an opportunity cost, and one might wonder whether academic forensic economics has reached the point of diminishing returns.

It may therefore be fitting to conclude by discussing failures—instances where forensic economics could have identified important hidden behavior, but failed to. One widely discussed example is Bernard Madoff's ponzi scheme. Madoff feeder funds such as the Fairfield Greenfield Sentry fund were in the major hedge fund datasets used by researchers and their returns were likely included as

[32]"From the Editor," *Financial Analysts Journal* 58(6), 17.

[33]In the letter writer's defense, he may not have had the econometric skills to know this, since it would have required running a regression to check (albeit a univariate regression using data easily downloaded from the Internet). It should also be mentioned that the letter writer's employer was not one the firms subsequently revealed to have made special arrangements with arbitrageurs.

[34]"'Stale Prices and Strategies for Trading Mutual Funds': Authors' Response," *Financial Analysts Journal* 59(1), 15.

[35]See Wolfers' comments quoted in Babay, Emily, "Study indicates racial bias in NBA refereeing," *Daily Pennsylvanian*, 5/11/2007 (http://www.dailypennsylvanian.com/node/53119, last accessed 2/2/2011).

[36]Alpert, Bill, "Technology Trader: Mysterious Changes in Key Wall Street Data," *Barron's*, 3/5/2007, p. 39 (http://online.barrons.com/article/SB117288153945425442.html, last accessed 2/2/2011).

data points in studies of manager skill. These returns look very anomalous. From 1990 to 2006, the Sentry fund's annual Sharpe ratio was consistently high, with a mean of 3.23 and a standard deviation of 0.62, for a ratio of 5.17. For the 850 hedge funds with at least ten years' data in the *Hedge Fund Research* dataset during that time period, this ratio (the Sharpe ratio of the Sharpe ratio) averaged 0.82 (SD 0.51), implying a *z*-score of 8.5 for the Madoff fund. All six funds with a *z*-score above 3.2 have since been found to either be Madoff feeder funds or involved in unrelated frauds. In this vein, recent studies by Brown et al. (2009), Bollen and Pool (forthcoming), and Dimmock and Gerken (2012) have shown a broader set of ex ante indicators are predictive of eventual hedge fund fraud.

The excessive heart surgeries in Redding provide another example. The *Dartmouth Atlas* data was very helpful to investigators in confirming Redding's outlier status once whistleblowers came forward, but *Atlas* data on the extremely high surgery rate in Redding had been available for years, and so conceivably the behavior could have been uncovered and halted sooner. Since the Redding case, the *Atlas* data has been used more prospectively to uncover specific instances of waste, or worse.

These examples are, of course, much easier to find with hindsight. The point in mentioning these missed findings is to illustrate that while past investment in forensic economics work may or may not have reached the point where marginal benefits and marginal costs are equalized, it has clearly not reached the point where the marginal benefits are zero. More and earlier investment might have caught these examples we now know about, along with other examples we do not.

Another class of missing findings are cases where the data exists, but researchers have been unable to access it. Favoritism across investment portfolios provides one example. We have results that are suggestive of favoritism, such as return gaps or hot IPO holdings that are correlated with manager's incentive to favor portfolios. But much more direct tests could be conducted with transaction-level data from the managers themselves. The SEC could access this data as part of its regular examinations and conduct tests for favoritism. If it does so, a public report on the extent of portfolio favoritism would be valuable. Alternatively, if the SEC does not have the analytical resources needed, the data could simply be required to be disclosed, and a community of researchers would likely conduct the analysis for free.

This is one example of a technique policymakers could apply more widely. In areas where there are suspicions, but no systematic evidence, of hidden behavior, policymakers could confer with researchers about the data required to test for this behavior, and then require its disclosure. For important types of hidden behavior, the academic community would likely conduct the analysis for free, and the combination of competition among researchers and widespread data availability would help ensure that the results obtained were accurate. One area where this technique is already being applied is in government procurement, where mandatory disclosure provided data for several studies discussed above.

Yet another class of missed findings are negative findings. This is, of course, a common problem with many lines of inquiry. We have a few examples above of academic work challenging forensic results—the work on backdating challenging prior work suggesting that option grants exploited inside information, work on statistical discrimination providing alternative explanations for racial differences in treatment, and work challenging conclusions of point shaving. There were also statements by academics challenging Christie and Schultz's interpretation of odd-eighths avoidance as evidence

of collusion.[37] While the low apparent false positive rate for forensic economics may be genuine, one might question whether there are adequate incentives to produce work that overturns forensic results.

4. *Conclusion*

Understanding hidden behavior has become a central component of the research agenda in many subfields of economics. This survey has reviewed a variety of techniques economists have adapted or developed for detecting such behavior. While forensic economics has been developing relatively independently in many fields, there is considerable overlap in the techniques used, as well as some common themes in the substantive findings.

It is sometimes argued that economics missed forecasting the current financial crisis. If we accept a common view of the financial crisis as an unexpectedly toxic combination of known components, it is worth noting that many of these components were known, or at least better known, through forensic economics. Examples include misleading accounting, manager–investor conflicts of interest in asset management, a race for yield (and window dressing of the risk), and lax or arguably industry-captured financial regulation. Postcrisis forensic work has also begun investigating lender–investor conflicts of interest in securitization. If hidden behavior played a role in the crisis, then forensic economics will likely play a role in understanding exactly what happened.

A lack of interesting and economically important hidden behavior to study is not likely to be what constrains the future of forensic economics, however. Most forensic economics work relies on data created and disclosed by others. If disclosure decisions are delegated to those with an interest in opacity, this will constrain the growth of knowledge, forensic and otherwise. In contrast, if policy harnesses the talents of economists and other researchers, the future of forensic economics may be bright indeed.

REFERENCES

Aboody, David, and Ron Kasznik. 2000. "CEO Stock Option Awards and the Timing of Corporate Voluntary Disclosures." *Journal of Accounting and Economics* 29 (1): 73–100.

Antonovics, Kate, and Brian G. Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *Review of Economics and Statistics* 91 (1): 163–77.

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (1): 127–51.

Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek. 1999. "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* 6 (4): 453–70.

Asker, John. 2010. "A Study of the Internal Organization of a Bidding Cartel." *American Economic Review* 100 (3): 724–62.

Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *American Economic Review* 85 (3): 304–21.

Baicker, Katherine, and Douglas Staiger. 2005. "Fiscal Shenanigans, Targeted Federal Health Care Funds, and Patient Mortality." *Quarterly Journal of Economics* 120 (1): 345–86.

Bajari, Patrick, and Lixin Ye. 2003. "Deciding between Competition and Collusion." *Review of Economics and Statistics* 85 (4): 971–89.

Baldwin, Laura H., Robert C. Marshall, and Jean-Francois Richard. 1997. "Bidder Collusion at Forest Service Timber Sales." *Journal of Political Economy* 105 (4): 657–99.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." *American Economic Review* 99 (4): 1278–1308.

Banerjee, Abhijit, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan. 2009. "Labor Market Discrimination in Delhi: Evidence from a Field Experiment." *Journal of Comparative Economics* 37 (1): 14–27.

Beaver, William H. 1968. "The Information Content of Annual Earnings Announcements." *Journal of Accounting Research* 6: 67–92.

Beber, Bernd, and Alexandra Scacco. 2008. "What the

---

[37] For example, see the Financial Economists Roundtable's "Statement on the Structure of the Nasdaq Stock Market," available at http://www.luc.edu/orgs/finroundtable/statement95a.html (last accessed 11/19/2009).

Numbers Say: A Digit-Based Test for Election Fraud Using New Data from Nigeria." http://homepages.nyu.edu/~bb89/files/Beber_Scacco_ElectionFraud.pdf.

Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago and London: University of Chicago Press.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (2): 169–217.

Becker, Gary S., and George J. Stigler. 1974. "Law Enforcement, Malfeasance, and Compensation of Enforcers." *Journal of Legal Studies* 3 (1): 1–18.

Bergstresser, Daniel, Mihir Desai, and Joshua Rauh. 2006. "Earnings Manipulation, Pension Assumptions, and Managerial Investment Decisions." *Quarterly Journal of Economics* 121 (1): 157–95.

Bergstresser, Daniel, and Thomas Philippon. 2006. "CEO Incentives and Earnings Management." *Journal of Financial Economics* 80 (3): 511–29.

Bernhardt, Dan, and Steven Heston. 2010. "Point Shaving in College Basketball: A Cautionary Tale for Forensic Economics." *Economic Inquiry* 48 (1): 14–25.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. 2007. "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption." *Quarterly Journal of Economics* 122 (4): 1639–76.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Bhojraj, Sanjeev, Paul Hribar, Marc Picconi, and John McInnis. 2009. "Making Sense of Cents: An Examination of Firms That Marginally Miss or Beat Analyst Forecasts." *Journal of Finance* 64 (5): 2360–88.

Bollen, Nicolas P. B., and Veronika K. Pool. 2009. "Do Hedge Fund Managers Misreport Returns? Evidence from the Pooled Distribution." *Journal of Finance* 64 (5): 2257–88.

Bollen, Nicolas P. B., and Veronika K. Pool. Forthcoming. "Suspicious Patterns in Hedge Fund Returns and the Risk of Fraud." *Review of Financial Studies*.

Bolton, Richard J., and David J. Hand. 2002. "Statistical Fraud Detection: A Review." *Statistical Science* 17 (3): 235–49.

Boudoukh, Jacob, Matthew Richardson, Marti Subrahmanyam, and Robert F. Whitelaw. 2002. "Stale Prices and Strategies for Trading Mutual Funds." *Financial Analysts Journal* 58 (4): 53–71.

Bresnahan, Timothy F. 1989. "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization, Volume 2*, edited by Richard Schmalensee and Robert D. Willig, 1011–57. Amsterdam; Oxford and Tokyo: North-Holland.

Brown, Stephen, William Goetzmann, Bing Liang, and Christopher Schwarz. 2009. "Estimating Operational Risk for Hedge Funds: The ω-Score." *Financial Analysts Journal* 65 (1): 43–53.

Bubb, Ryan, and Alex Kaufman. 2009. "Securitization and Moral Hazard: Evidence from a Lender Cutoff Rule." Federal Reserve Bank of Boston Public Policy Discussion Paper 09-5.

Burgstahler, David, and Ilia Dichev. 1997. "Earnings Management to Avoid Earnings Decreases and Losses." *Journal of Accounting and Economics* 24 (1): 99–126.

Charles, Kerwin Kofi, and Jonathan Guryan. 2008. "Prejudice and Wages: An Empirical Assessment of Becker's the Economics of Discrimination." *Journal of Political Economy* 116 (5): 773–809.

Chauvin, Keith W., and Catherine Shenoy. 2001. "Stock Price Decreases Prior to Executive Stock Option Grants." *Journal of Corporate Finance* 7 (1): 53–76.

Chen, Jowei. 2007. "What Does Baseball Teach Us about Reducing Racial Discrimination? Evidence from Two Natural Experiments." Unpublished.

Chevalier, Judith, and Glenn Ellison. 1997. "Risk Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy* 105 (6): 1167–1200.

Christie, William G., Jeffrey H. Harris, and Paul H. Schultz. 1994. "Why Did Nasdaq Market Makers Stop Avoiding Odd-Eighth Quotes?" *Journal of Finance* 49 (5): 1841–60.

Christie, William G., and Paul H. Schultz. 1994. "Why Do Nasdaq Market Makers Avoid Odd-Eighth Quotes?" *Journal of Finance* 49 (5): 1813–40.

Christie, William G., and Paul H. Schultz. 1995. "Policy Watch: Did Nasdaq Market Makers Implicitly Collude?" *Journal of Economic Perspectives* 9 (3): 199–208.

Cici, Gjergji, Scott Gibson, and Rabih Moussawi. 2010. "Mutual Fund Performance When Parent Firms Simultaneously Manage Hedge Funds." *Journal of Financial Intermediation* 19 (2): 169–87.

Clotfelter, Charles T. 1983. "Tax Evasion and Tax Rates: An Analysis of Individual Returns." *Review of Economics and Statistics* 65 (3): 363–73.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy. 2008. "The Small World of Investing: Board Connections and Mutual Fund Returns." *Journal of Political Economy* 116 (5): 951–79.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy. 2010. "Sell-Side School Ties." *Journal of Finance* 65 (4): 1409–37.

Corts, Kenneth S. 1999. "Conduct Parameters and the Measurement of Market Power." *Journal of Econometrics* 88 (2): 227–50.

Cramton, Peter, and Jesse A. Schwartz. 2000. "Collusive Bidding: Lessons from the FCC Spectrum Auctions." *Journal of Regulatory Economics* 17 (3): 229–52.

Cullen, Julie Berry, and Randall Reback. 2006. "Tinkering toward Accolades: School Gaming Under a Performance Accountability System." In *Improving School Accountability: Check-Ups or Choice*, edited by Timothy J. Gronberg and Dennis W. Jansen, 1–34. Amsterdam and San Diego: Elsevier, JAI Press.

Currie, Janet, and W. Bentley MacLeod. 2008. "First Do No Harm? Tort Reform and Birth Outcomes."

Quarterly Journal of Economics 123 (2): 795–830.

Dartmouth Medical School. 1998. *The Dartmouth Atlas of Health Care in the United States*. Chicago: American Hospital Publishing.

Degeorge, Francois, Jayendu Patel, and Richard Zeckhauser. 1999. "Earnings Management to Exceed Thresholds." *Journal of Business* 72 (1): 1–33.

DellaVigna, Stefano, and Eliana La Ferrara. 2010. "Detecting Illegal Arms Trade." *American Economic Journal: Economic Policy* 2 (4): 26–57.

Dimmock, Stephen G., and William C. Gerken. 2012. "Predicting Fraud by Investment Managers." *Journal of Financial Economics* 105 (1): 153–73.

Di Tella, Rafael, and Ignacio Franceschelli. 2009. "Government Advertising and Media Coverage of Corruption Scandals." National Bureau of Economic Research Working Paper 15402.

Di Tella, Rafael, and Ernesto Schargrodsky. 2003. "The Role of Wages and Auditing During a Crackdown on Corruption in the City of Buenos Aires." *Journal of Law and Economics* 46 (1): 269–92.

Donohue, John J., and Justin Wolfers. 2005. "Uses and Abuses of Empirical Evidence in the Death Penalty Debate." *Stanford Law Review* 58 (3): 791–845.

Dubben, Hans-Hermann, and Hans-Peter Beck-Bornholdt. 2005. "Systematic Review of Publication Bias in Studies on Publication Bias." *British Medical Journal* 331: 433–34.

Dube, Arindrajit, Ethan Kaplan, and Suresh Naidu. 2011. "Coups, Corporations, and Classified Information." *Quarterly Journal of Economics* 126 (3): 1375–1409.

Duflo, Esther, Rema Hanna, and Stephen Ryan. 2008. "Monitoring Works: Getting Teachers to Come to School." Center for Economic and Policy Research Discussion Paper 6682.

Duggan, Mark, and Steven D. Levitt. 2002. "Winning Isn't Everything: Corruption in Sumo Wrestling." *American Economic Review* 92 (5): 1594–1605.

Edelman, Benjamin, and Ian Larkin. 2009. "Demographics, Career Concerns or Social Comparison: Who Games SSRN Download Counts?" Harvard Business School Working Paper 09-096.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake. 2003. "Incentive Fees and Mutual Funds." *Journal of Finance* 58 (2): 779–804.

Faccio, Mara. 2006. "Politically Connected Firms." *American Economic Review* 96 (1): 369–86.

Faccio, Mara, Ronald W. Masulis, and John J. McConnell. 2006. "Political Connections and Corporate Bailouts." *Journal of Finance* 61 (6): 2597–2635.

Faccio, Mara, and David C. Parsley. 2009. "Sudden Deaths: Taking Stock of Geographic Ties." *Journal of Financial and Quantitative Analysis* 44 (3): 683–718.

Feldman, Naomi E., and Joel Slemrod. 2007. "Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns." *Economic Journal* 117 (518): 327–52.

Ferguson, Thomas, and Hans-Joachim Voth. 2008. "Betting on Hitler—the Value of Political Connections in Nazi Germany." *Quarterly Journal of Economics* 123 (1): 101–37.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123 (2): 703–45.

Figlio, David N. 2006. "Testing, Crime and Punishment." *Journal of Public Economics* 90 (4–5): 837–51.

Figlio, David N., and Lawrence S. Getzler. 2002. "Accountability, Ability and Disability: Gaming the System." National Bureau of Economic Research Working Paper 9307.

Figlio, David N., and Joshua Winicki. 2005. "Food for Thought: The Effects of School Accountability Plans on School Nutrition." *Journal of Public Economics* 89 (2–3): 381–94.

Fisman, David, Raymond Fisman, Julia Galef, and Rakesh Khurana. 2006. "Estimating the Value of Connections to Vice-President Cheney." http://fairmodel.econ.yale.edu/ec483/ffgk.pdf.

Fisman, Raymond. 2001. "Estimating the Value of Political Connections." *American Economic Review* 91 (4): 1095–1102.

Fisman, Raymond, and Edward Miguel. 2007. "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets." *Journal of Political Economy* 115 (6): 1020–48.

Fisman, Raymond, and Edward Miguel. 2008a. *Economic Gangsters: Corruption, Violence, and the Poverty of Nations*. Princeton and Oxford: Princeton University Press.

Fisman, Raymond, and Edward Miguel. 2008b. "How Economics Can Defeat Corruption." *Foreign Policy* 168: 66–75.

Fisman, Raymond, and Shang-Jin Wei. 2004. "Tax Rates and Tax Evasion: Evidence From 'Missing Imports' in China." *Journal of Political Economy* 112 (2): 471–96.

Fisman, Raymond, and Shang-Jin Wei. 2007. "The Smuggling of Art, and the Art of Smuggling: Uncovering the Illicit Trade in Cultural Property and Antiques." National Bureau of Economic Research Working Paper 13446.

Fix, Michael, and Margery Austin Turner, ed. 1998. *A National Report Card on Discrimination in America: The Role of Testing*. Washington, D.C.: Urban Institute.

Frieder, Laura L., and Jonathan Zittrain. 2008. "Spam Works: Evidence from Stock Touts and Corresponding Market Activity." *Hastings Communication and Entertainment Law Journal* 30: 479–510.

Fryer, Roland G., and Steven D. Levitt. 2004. "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics* 119 (3): 767–805.

Gaspar, Jose-Miguel, Massimo Massa, and Pedro Matos. 2006. "Favoritism in Mutual Fund Families: Evidence on Strategic Cross-Fund Subsidization." *Journal of Finance* 61 (1): 73–104.

Gawande, Atul. 2009. "The Cost Conundrum: What a Texas Town Can Teach Us about Health Care." *The New Yorker*, June 1.

Genesove, David, and Wallace P. Mullin. 2001. "Rules,

Communication, and Collusion: Narrative Evidence from the Sugar Institute Case." *American Economic Review* 91 (3): 379–98.

Gibbs, Jonathan. 2007. "Point Shaving in the NBA: An Economic Analysis of the National Basketball Association's Point Spread Betting Market." http://economics.stanford.edu/files/Theses/Theses_2007/Gibbs2007.pdf.

Gino, Francesca, and Lamar Pierce. 2010. "Robin Hood under the Hood: Wealth-Based Discrimination in Illicit Customer Help." *Organization Science* 21 (6): 1176–94.

Goldberg, Pinelopi Koujianou. 1996. "Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* 104 (3): 622–54.

Goldman, Eitan, Jorg Rocholl, and Jongil So. 2009. "Do Politically Connected Boards Affect Firm Value?" *Review of Financial Studies* 22 (6): 2331–60.

Gordon, Sanford C. 2009. "Assessing Partisan Bias in Federal Public Corruption Prosecutions." *American Political Science Review* 103 (4): 534–54.

Grogger, Jeffrey, and Greg Ridgeway. 2006. "Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness." *Journal of the American Statistical Association* 101 (475): 878–87.

Gruber, Jonathan, and Maria Owings. 1996. "Physician Financial Incentives and Cesarean Section Delivery." *RAND Journal of Economics* 27 (1): 99–123.

Grundfest, Joseph A., and Nadya Malenko. 2009. "Quadrophobia: Strategic Rounding of EPS Data." Stanford Law School John M. Olin Program in Law and Economics Working Paper 388.

Hanna, Rema, and Leigh Linden. 2009. "Measuring Discrimination in Education." National Bureau of Economic Research Working Paper 15057.

Healy, Paul M. 1985. "The Effect of Bonus Schemes on Accounting Decisions." *Journal of Accounting and Economics* 7 (1–3): 85–107.

Healy, Paul M., and James M. Wahlen. 1999. "A Review of the Earnings Management Literature and Its Implications for Standard Setting." *Accounting Horizons* 13 (4): 365–83.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12 (2): 101–16.

Heckman, James J., and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael Fix and Raymond J. Struyk, 187–258. Washington, D.C.: Urban Institute Press.

Heron, Randall A., and Erik Lie. 2007. "Does Backdating Explain the Stock Price Pattern around Executive Stock Option Grants?" *Journal of Financial Economics* 83 (2): 271–95.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7: 24–52.

Hsieh, Chang-Tai, and Enrico Moretti. 2006. "Did Iraq Cheat the United Nations? Underpricing, Bribes,
and the Oil for Food Program." *Quarterly Journal of Economics* 121 (4): 1211–48.

Hubbard, Thomas N. 1998. "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market." *RAND Journal of Economics* 29 (2): 406–26.

Hubbard, Thomas N. 2002. "How Do Consumers Motivate Experts? Reputational Incentives in an Auto Repair Market." *Journal of Law and Economics* 45 (2): 437–68.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

Jacob, Brian A. 2005. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89 (5–6): 761–96.

Jacob, Brian A., and Steven D. Levitt. 2003a. "Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory." *Brookings–Wharton Papers on Urban Affairs* 1: 185–209.

Jacob, Brian A., and Steven D. Levitt. 2003b. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118 (3): 843–77.

Jayachandran, Seema. 2006. "The Jeffords Effect." *Journal of Law and Economics* 49 (2): 397–425.

Jin, Ginger Zhe, and Andrew Kato. 2006. "Price, Quality, and Reputation: Evidence from an Online Field Experiment." *RAND Journal of Economics* 37 (4): 983–1005.

Jin, Ginger Zhe, and Phillip Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *Quarterly Journal of Economics* 118 (2): 409–51.

Jin, Ginger Zhe, and Phillip Leslie. 2009. "Reputational Incentives for Restaurant Hygiene." *American Economic Journal: Microeconomics* 1 (1): 237–67.

Jowell, Roger, and Patricia Prescott-Clarke. 1970. "Racial Discrimination and White-Collar Workers in Britain." *Race & Class* 11 (4): 397–417.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng. 2008. "Unobserved Actions of Mutual Funds." *Review of Financial Studies* 21 (6): 2379–2416.

Kerr, Steven. 1975. "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Journal* 18 (4): 769–83.

Kessler, Daniel, and Mark McClellan. 1996. "Do Doctors Practice Defensive Medicine?" *Quarterly Journal of Economics* 111 (2): 353–90.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *Quarterly Journal of Economics* 125 (1): 307–62.

Khwaja, Asim Ijaz, and Atif Mian. 2005a. "Do Lenders Favor Politically Connected Firms? Rent Provision in an Emerging Financial Market." *Quarterly Journal of Economics* 120 (4): 1371–1411.

Khwaja, Asim Ijaz, and Atif Mian. 2005b. "Unchecked Intermediaries: Price Manipulation in an Emerging Stock Market." *Journal of Financial Economics* 78 (1): 203–41.

Klemperer, Paul. 2002. "What Really Matters in Auction Design." *Journal of Economic Perspectives* 16 (1): 169–89.

Kleven, Henrik J., Martin B. Knudsen, Claus T. Kreiner, Soren Pedersen, and Emmanuel Saez. 2010. "Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark." National Bureau of Economic Research Working Paper 15769.

Knight, Brian G. 2006. "Are Policy Platforms Capitalized into Equity Prices? Evidence from the Bush/Gore 2000 Presidential Election." *Journal of Public Economics* 90 (4–5): 751–73.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109 (1): 203–29.

Krueger, Alan B., and Alexandre Mas. 2004. "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires." *Journal of Political Economy* 112 (2): 253–89.

Kwoka, John E. 1997. "The Price Effects of Bidding Conspiracies: Evidence from Real Estate Auction 'Knockouts.'" *Antitrust Bulletin* 42 (2): 503–16.

Larkin, Ian. 2007. "The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software Sales." http://www.gsb.stanford.edu/facseminars/events/applied_microecon/documents/ame_11_07_larkin.pdf.

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review* 90 (5): 1346–61.

Levitt, Steven D. 2006. "White-Collar Crime Writ Small: A Case Study of Bagels, Donuts, and the Honor System." *American Economic Review* 96 (2): 290–94.

Levitt, Steven D., and Chad Syverson. 2008. "Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions." *Review of Economics and Statistics* 90 (4): 599–611.

Levitt, Steven D., and Sudhir Alladi Venkatesh. 2000. "An Economic Analysis of a Drug-Selling Gang's Finances." *Quarterly Journal of Economics* 115 (3): 755–89.

Levitt, Steven D., and Sudhir Alladi Venkatesh. 2007. "An Empirical Analysis of Street-Level Prostitution." http://economics.uchicago.edu/pdf/Prostitution%205.pdf.

Lie, Erik. 2005. "On the Timing of CEO Stock Option Awards." *Management Science* 51 (5): 802–12.

Lin, Hsiou-wei, and Maureen F. McNichols. 1998. "Underwriting Relationships, Analysts' Earnings Forecasts and Investment Recommendations." *Journal of Accounting and Economics* 25 (1): 101–27.

Lin, Ming-Jen, Nancy Qian, and Jin-Tan Liu. 2008. "More Women Missing, Fewer Girls Dying: The Impact of Abortion on Sex Ratios at Birth and Excess Female Mortality in Taiwan." National Bureau of Economic Research Working Paper 14541.

List, John A. 2004. "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field." *Quarterly Journal of Economics* 119 (1): 49–89.

List, John A. 2006. "Field Experiments: A Bridge between Law and Naturally Occurring Data." *B.E. Journal of Economic Analysis and Policy* 6 (2).

Liu, Xiaoding, and Jay R. Ritter. 2010. "The Economic Consequences of IPO Spinning." *Review of Financial Studies* 23 (5): 2024–59.

Ljungqvist, Alexander, Christopher Malloy, and Felicia Marston. 2009. "Rewriting History." *Journal of Finance* 64 (4): 1935–60.

Macey, Jonathan R., and Maureen O'Hara. 2009. "Regulation and Scholarship: Constant Companions or Occasional Bedfellows?" *Yale Journal on Regulation* 26 (1): 89–116.

MacKinlay, A. Craig. 1997. "Event Studies in Economics and Finance." *Journal of Economic Literature* 35 (1): 13–39.

Makowsky, Michael D., and Thomas Stratmann. 2009. "Political Economy at Any Speed: What Determines Traffic Citations?" *American Economic Review* 99 (1): 509–27.

Malmendier, Ulrike, and Geoffrey Tate. 2009. "Superstar CEOs." *Quarterly Journal of Economics* 124 (4): 1593–1638.

Mas, Alexandre. 2008. "Labour Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market." *Review of Economic Studies* 75 (1): 229–58.

Massa, Massimo, Jonathan Reuter, and Eric Zitzewitz. 2010. "When Should Firms Share Credit with Employees? Evidence from Anonymously Managed Mutual Funds." *Journal of Financial Economics* 95 (3): 400–424.

McGuire, Thomas G. 2000. "Physician Agency." In *Handbook of Health Economics, Volume 1A*, edited by Anthony J. Culyer and Joseph P. Newhouse, 461–536. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.

Meulbroek, Lisa K. 1992. "An Empirical Analysis of Illegal Insider Trading." *Journal of Finance* 47 (5): 1661–99.

Michaely, Roni, and Kent L. Womack. 1999. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." *Review of Financial Studies* 12 (4): 653–86.

Morey, Matthew R., and Edward S. O'Neal. 2006. "Window Dressing in Bond Mutual Funds." *Journal of Financial Research* 29 (3): 325–47.

Narayanan, M. P., and H. Nejat Seyhun. 2008. "The Dating Game: Do Managers Designate Option Grant Dates to Increase Their Compensation?" *Review of Financial Studies* 21 (5): 1907–45.

Neumark, David. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics* 111 (3): 915–41.

Olken, Benjamin A. 2006. "Corruption and the Costs of Redistribution: Micro Evidence from Indonesia." *Journal of Public Economics* 90 (4–5): 853–70.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.

Olken, Benjamin A., and Patrick Barron. 2009. "The

Simple Economics of Extortion: Evidence from Trucking in ACEH." *Journal of Political Economy* 117 (3): 417–52.

Ondrich, Jan, Stephen Ross, and John Yinger. 2003. "Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Available Houses from Black Customers?" *Review of Economics and Statistics* 85 (4): 854–73.

Oyer, Paul. 1998. "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality." *Quarterly Journal of Economics* 113 (1): 149–85.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation." *American Economic Review* 101 (4): 1410–35.

Pesendorfer, Martin. 2000. "A Study of Collusion in First-Price Auctions." *Review of Economic Studies* 67 (3): 381–411.

Pierce, Lamar, and Jason Alan Snyder. 2008. "Ethical Spillovers in Firms: Evidence from Vehicle Emissions Testing." *Management Science* 54 (11): 1891–1903.

Pissarides, Christopher A., and Guglielmo Weber. 1989. "An Expenditure-Based Estimate of Britain's Black Economy." *Journal of Public Economics* 39 (1): 17–32.

Pomeranz, Dina. 2010. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." http://www.people.fas.harvard.edu/~pomeranz/Pomeranz%20Job%20Market%20Paper.pdf.

Porter, Robert H. 2005. "Detecting Collusion." *Review of Industrial Organization* 26 (2): 147–67.

Porter, Robert H., and J. Douglas Zona. 1993. "Detection of Bid Rigging in Procurement Auctions." *Journal of Political Economy* 101 (3): 518–38.

Porter, Robert H., and J. Douglas Zona. 1999. "Ohio School Milk Markets: An Analysis of Bidding." *RAND Journal of Economics* 30 (2): 263–88.

Price, Joseph, Marc Remer, and Daniel F. Stone. 2012. "Sub-perfect Game: Profitable Biases of NBA Referees." *Journal of Economics and Management Strategy* 21 (1): 271–300.

Price, Joseph, and Justin Wolfers. 2010. "Racial Discrimination among NBA Referees." *Quarterly Journal of Economics* 125 (4): 1859–87.

Qian, Nancy. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance." *Quarterly Journal of Economics* 123 (3): 1251–85.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *Quarterly Journal of Economics* 119 (2): 679–705.

Reinikka, Ritva, and Jakob Svensson. 2005. "Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda." *Journal of the European Economic Association* 3 (2–3): 259–67.

Reuter, Jonathan. 2006. "Are IPO Allocations for Sale? Evidence from Mutual Funds." *Journal of Finance* 61 (5): 2289–2324.

Reuter, Jonathan. 2009. "Does Advertising Bias Product Reviews? An Analysis of Wine Ratings." *Journal of Wine Economics* 4 (2): 125–51.

Reuter, Jonathan, and Eric Zitzewitz. 2006. "Do Ads Influence Editors? Advertising and Bias in the Financial Media." *Quarterly Journal of Economics* 121 (1): 197–227.

Riach, Peter A., and Judy Rich. 2002. "Field Experiments of Discrimination in the Market Place." *Economic Journal* 112 (483): F480–518.

Rinallo, Diego, and Suman Basuroy. 2009. "Does Advertising Spending Influence Media Coverage of the Advertiser?" *Journal of Marketing* 73 (6): 33–46.

Ritter, Jay R. 2008. "Forensic Finance." *Journal of Economic Perspectives* 22 (3): 127–47.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.

Schwert, G. William. 1981. "Using Financial Data to Measure Effects of Regulation." *Journal of Law and Economics* 24 (1): 121–58.

Seyhun, H. Nejat. 1986. "Insiders' Profits, Costs of Trading, and Market Efficiency." *Journal of Financial Economics* 16 (2): 189–212.

Silverman, Elaine, and Jonathan Skinner. 2004. "Medicare Upcoding and Hospital Ownership." *Journal of Health Economics* 23 (2): 369–89.

Simon, Paul A., Phillip Leslie, Grace Run, Ginger Zhe Jin, Roshan Reporter, Arturo Aguirre, and Jonathan E. Fielding. 2005 "Impact of Restaurant Hygiene Grade Cards on Foodborne-Disease Hospitalizations in Los Angeles County." *Journal of Environmental Health* 67 (7): 32–36.

Slemrod, Joel. 1985. "An Empirical Test for Tax Evasion." *Review of Economics and Statistics* 67 (2): 232–38.

Slemrod, Joel. 2007. "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives* 21 (1): 25–48.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian. 2001. "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota." *Journal of Public Economics* 79 (3): 455–83.

Snyder, Christopher, and Owen Zidar. 2009. "Resume Padding among Economists." Unpublished.

Snyder, Jason Alan. 2007. "Discontinuous Electoral Distributions." Unpublished.

Sukhtankar, Sandip. 2011. "Sweetening the Deal? Political Connections and Sugar Mills in India." http://www.dartmouth.edu/~sandip/political_connections_sugar_mills.pdf.

Svensson, Jakob. 2005. "Eight Questions about Corruption." *Journal of Economic Perspectives* 19 (3): 19–42.

Topalova, Petia, Prachi Mishra, and Arvind Subramanian. 2007. "Policies, Enforcement, and Customs Evasion: Evidence from India." International Monetary Fund Working Paper 07/60.

Wolfers, Justin. 2006. "Point Shaving: Corruption in NCAA Basketball." *American Economic Review* 96

(2): 279–83.

Wolfram, Catherine D. 1999. "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review* 89 (4): 805–26.

Yermack, David. 1997. "Good Timing: CEO Stock Option Awards and Company News Announcements." *Journal of Finance* 52 (2): 449–76.

Yermack, David. 2006. "Flights of Fancy: Corporate Jets, CEO Perquisites, and Inferior Shareholder Returns." *Journal of Financial Economics* 80 (1): 211–42.

Yinger, John. 1986. "Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act." *American Economic Review* 76 (5): 881–93.

Yinger, John. 1998. "Evidence on Discrimination in Consumer Markets." *Journal of Economic Perspectives* 12 (2): 23–40.

Zinman, Jonathan, and Eric Zitzewitz. 2009. "Wintertime for Deceptive Advertising." http://www.dartmouth.edu/~ericz/wintertime.pdf.

Ziobrowski, Alan J., Ping Cheng, James W. Boyd, and Brigitte J. Ziobrowski. 2004. "Abnormal Returns from the Common Stock Investments of the U.S.

Senate." *Journal of Financial and Quantitative Analysis* 39 (4): 661–76.

Zitzewitz, Eric. 2002. "Regulation Fair Disclosure and the Private Information of Analysts." http://www.dartmouth.edu/~ericz/regfd.pdf.

Zitzewitz, Eric. 2003. "Who Cares About Shareholders? Arbitrage-Proofing Mutual Funds." *Journal of Law, Economics, and Organization* 19 (2): 245–80.

Zitzewitz, Eric. 2006a. "How Widespread Was Late Trading in Mutual Funds?" *American Economic Review* 96 (2): 284–89.

Zitzewitz, Eric. 2006b. "Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making." *Journal of Economics and Management Strategy* 15 (1): 67–99.

Zitzewitz, Eric. 2009. "Prosecutorial Discretion in Mutual Fund Settlement Negotiations, 2003–7." *B.E. Journal of Economic Analysis and Policy* 9 (1).

Zitzewitz, Eric. 2010. "Does Transparency Really Increase Corruption? Evidence from the 'Reform' of Figure Skating Judging." http://www.dartmouth.edu/~ericz/transparency.pdf.