Is Inventory’s Fiscal Year End Effect Caused by Sales Timing?  
A Test Using a Natural Experiment from Germany

Richard Lai  
Harvard Business School\(^1\)

First draft: August 5, 2006  
This draft: January 26, 2008

One explanation for firms’ inventory cycles is that they are tied to the calendar year, reflecting fundamental demand seasonality. But we find that these cycles are also tied to the fiscal year, an accounting artifact constructed by firms themselves. Specifically, inventory is lower at fiscal year end (FYE), a phenomenon we call inventory’s FYE effect. Among U.S. manufacturers, wholesalers, and retailers, we find inventory is 10% lower at FYE than at other times of the fiscal year, controlling for calendar time. In aggregate, this 10% is $47 billion, valued at cost of goods. One possible explanation for the FYE effect is sales timing, in which executives’ private benefits lead them to pull some post-FYE sales into the FYE. But the literature suggests three alternative hypotheses that can also explain the FYE effect. To test for sales timing, we employ a novel natural experiment based on Germany’s tax code change in 2000, when some firms change their FYEs in a way that is plausibly exogenous to inventory patterns. We find that these German firms also have lower inventory at FYE. This result is robust to corrections for possible treatment selection using the Heckit procedure and propensity scoring. We also examine mediator and moderator effects. For example, the link from FYE to lower inventory is mediated by lower margins and higher sales. Firms whose executive compensations have a higher bonus component have a stronger FYE effect. Taken together, the evidence is consistent with sales timing, but is not explained by the alternative hypotheses. Finally, we estimate that 1 percentage point lower inventory at FYE is associated with an economically significant 1.7% lower Tobin’s \(q\), which is due to lower gross profits and higher costs like inventory holdings expenses and capacity investments. We conclude with a discussion of limitations, next steps, and some intriguing implications for research and practice.

\(^{1}\) Harvard Business School, Morgan Hall, Boston, MA 02163. Tel: (617) 495 6228. Email: rlai@hbs.edu.  
This is a revised version of a paper entitled “Inventory’s Fiscal Year End Effect,” which received Honorable Mention for the 2007 INFORMS M&SOM Student Paper Competition. I am most grateful for encouragement and feedback from my advisors: Ananth Raman (chair), Vishal Gaur, Dale Jorgenson, and Josh Lerner. I also benefited tremendously with feedback from faculty and students at HBS, as well as from seminars as GeorgiaTech, HBS, Katz, Stanford, USC, and Wharton. All errors are mine.
Is Inventory’s Fiscal Year End Effect Caused by Sales Timing?  
A Test Using a Natural Experiment from Germany

1. Introduction

Figure 1 depicts RadioShack’s finished goods inventory levels over time. Before 1992, inventory dips every second calendar quarter. The dips are sizable, about $1 billion in inventory, valued at cost of goods. One explanation for this (e.g., Stevenson (1999), pg. 485) is that inventory varies with demand seasonality in the calendar year. Perhaps at every second quarter, RadioShack’s demand peaks, depleting inventory to its trough.

In 1992, RadioShack changed its fiscal year to end not in the second calendar quarter, but the fourth. Figure 1 shows that after that, inventory is lowest in the new fiscal year ends. We shall call this phenomenon, in which inventory is lower at fiscal year end than at other times of the fiscal year, the “fiscal year end (FYE) effect.”

Inventory variation—both between firm and within firm but over time—is a central topic in empirical operations management—e.g., Rajagopalan and Malhotra (2001), Chen, et al. (2005), Gaur, et al. (2005), Netessine and Roumiantsev (2007). A key objective is to better understand how firms really manage inventory. In most studies, such as all four papers just cited, within-year inventory cycles are assumed to be tied to the calendar year, reflecting fundamental demand seasonality. The FYE effect is intriguing because inventory cycles seem to be also tied to the fiscal year, an accounting artifact constructed by firms.

It has also not been rigorously studied, so this study fills a gap in current research (figure 2). There is much research into calendar year effects (i.e., seasonality) of sales (e.g., Fisher and Raman (1996), Bitran and Mondschein (1997), Taylor (2006)) and inventory (e.g., Nerlove, et al. (1993)). There is some research on the FYE effect of sales. The pioneering work is by Oyer (1998); see also Nevo and Wolfram (2002), Steenburgh (2004), and Larkin (2006). This last research stream sheds only limited light on inventory’s FYE effect, because
inventory is related not just to sales, but also to production, purchasing, and write-offs. Thus, inventory’s FYE effect is related to but distinct from previously studied effects.

In section 2, we report that, in a panel of all U.S. manufacturers, wholesalers, and retailers, inventory is 10% lower in the fourth fiscal quarter than other quarters, controlling for calendar quarters (the concept of FYE is econometrically implemented as the fourth fiscal quarter). In 2006, this FYE dip is about $47 billion in valued at cost of goods sold.

Such a pervasive, economically significant FYE effect might still be just a curiosity, but there is a priori reason to think it is also detrimental to firm’s long-term financial valuation. The media and many executives suggest that the effect might be due to sales timing, in which executives’ private benefits lead them to pull some next-quarter sales into the FYE. In an egregious example, the Securities and Exchange Commission (SEC) alleges that Bristol-Myers Squibb cut prices to induce its distributors to take on $1.3 billion of inventory at 2001 FYE. This inflated FYE revenues by 7% so that its executives could make sales targets. This practice, called channel stuffing, is so pervasive that it has gotten many labels: “loading” in SEC (2006) vs. Virbac, “gallon pushing” in SEC (2005) vs. Coca-Cola, “floor sweeping” in SEC (2003) vs. ClearOne, and “pull forwards” in SEC (2004b) vs. K-mart.

Although sales timing seems like an obvious explanation for inventory’s FYE effect, the literature suggests otherwise. There are at least three alternative hypotheses that could also explain lower inventory at FYE. Also, the SEC cases suggest, but do not conclusively show, that sales timing hurts firm valuation. So if inventory’s FYE effect is indeed due to sales timing, we suspect it has negative financial implications, but we do not really know.

In this paper, we ask if inventory’s FYE effect is due to sales timing and if it lowers firm valuation. To clarify, we seek to find if sales timing is a cause, not if the alternatives are not. We focus on timing because among the hypotheses, it alone seems to have a deleterious effect on firm valuation. We explain this in section 3, in which we first review the literature on sales timing and three alternative hypotheses. There, we also show that sales timing generates different predictions than the three, allowing us to empirically test for timing.
In section 4, we provide evidence for sales timing that is not explained by alternative hypotheses. We exploit a novel natural experiment based on the Germany’s tax code change in 2000, when some firms change their FYE in a way that is plausibly exogenous to inventory patterns. Using a panel dataset of German firms hand-coded from primary sources and from CapitalIQ, we find that firms that changed their FYE have lower inventory in both their old and new fourth fiscal quarters. This result is robust to corrections for possible treatment selection using the Heckit procedure and propensity scoring in a differences-in-difference framework. We will explain why this is clean evidence for sales timing that is not explained by the three alternatives.

In section 5, we provide further evidence for sales timing by directly examining mediators and moderators. For example, we find that the link from FYE to lower inventory is mediated by lower margins and higher sales, and not by reduced production. The FYE effect is stronger for firms that pay higher bonuses and sell durable goods, and weakened when firms are under scrutiny as when they face federal class action suits.

In section 6, we consider financial implications. We find that 1 percentage point lower inventory at FYE is associated with 1.7% lower valuation in industry-adjusted Tobin’s $q$. This elasticity arises from two sources: lower gross profits and higher costs. We find that firms sell more but at lower margins at FYE, so gross profits are a net 5% lower than if there were no FYE effect. Costs could also be higher: inventory fluctuations result in higher holding costs and sales fluctuations in higher capacities.

In section 6, we discuss the limitations of our study and suggest next steps. For example, we have focused only on the importance of the sales timing and not that of the alternative hypotheses; we merely provide evidence for sales timing not explained by others, not evidence for the absence of others. We conclude with implications for research and practice.

We earlier explain that inventory’s FYE effect fills a gap in existing research on calendar and fiscal year effects. More generally, this study builds on the literature that seeks to explain inventory variation, not just in terms of calendar or fiscal year effects—e.g., Lee, et

To summarize, what is new in this study is that it:

1. Identifies inventory’s FYE effect as related to but distinct from those in figure 2;
2. Provides the first empirical evidence that the effect is pervasive and substantive, a finding that is not mechanically implied by previous research on sales’ FYE effect;
3. Provides the first empirical evidence that sales timing is a cause of the effect, even though it is not an obvious cause, with at least three alternative explanations;
4. Introduces a natural experiment that is novel in the empirical investigation of FYE effects, whether for sales or for inventory;
5. Presents the first empirical estimation of the valuation impact of an FYE effect, whether for sales or for inventory.

The rest of the paper provides details of the above.

2. **Motivation: the Pervasive and Substantive FYE Effect**

We motivate this paper by showing that the FYE effect is pervasive and economically significant. Here, we seek only to *identify* the size of the FYE effect. We are not concerned with causality, which we address in the rest of this paper.
We use a panel of all 2,512 U.S. manufacturers, wholesalers, and retailers (NAICS codes 31 through 48) in COMPUSTAT, from 1984 through 2006. We omit observations that are economically insignificant, with missing or negative sales or cost of goods sold (COGS). Table 1, panel (a), shows the variation of the firm-quarter observations by FYE; this variation allows identification of the FYE effect. Panel (b) shows the summary statistics.

Our empirical model is a straightforward reduced form specification:

\[
\nu_{it} = \sum_{f=1}^{4} \nu_{it}^{f} \phi^{f} + \sum_{c=1}^{4} \nu_{it}^{c} \chi^{c} + \gamma_{y} + \kappa_{i} + \eta_{it},
\]

where \(\nu_{it}\) is inventory, measured as finished goods inventory divided by quarterly COGS, of firm \(i\) in quarter \(t\), each of which is indexed with a fiscal quarter label \(f\) and a calendar quarter label \(c\). Different types of inventory—raw materials, work-in-progress—have different dynamics; we focus on finished goods inventory because it has the highest economic value. We also measure inventory without scaling by COGS and scaling by total assets, and obtain the similar results (see appendix for issues regarding measuring inventory). \(\phi^{f}\) is the effect on inventory of being in fiscal quarter \(f\), and \(\chi^{c}\), of being in calendar quarter \(c\); \(\nu\)'s are indicator variables. \(\gamma_{y}\) and \(\kappa_{i}\) are calendar year (indexed by \(y\)) and firm fixed effects, and \(\eta_{it}\) is assumed to be white noise. \(\kappa_{i}\) accounts for time-invariant characteristics, such as industry, inventory accounting method, and LIFO/FIFO reserves. The de-meaned model is as follows, with \(\Delta\) the de-mean operator:

\[
\Delta \nu_{it} = \sum_{f=1}^{4} \Delta \nu_{it}^{f} \phi^{f} + \sum_{c=1}^{4} \Delta \nu_{it}^{c} \chi^{c} + \Delta \gamma_{y} + \Delta \eta_{it}.
\]

In Table 2, we first report in model (1) estimates of calendar quarter effects without fiscal quarters. As with all estimations in this paper, these are obtained with Huber-White robust standard errors in case inventory varies differently by fiscal quarter, and clustered around firms to account for potential within-firm correlation. It is also in log form, so we interpret the estimate as 11.3% lower inventory in the fourth calendar quarter, \(\chi^{4}\), and a little higher in the quarters before and after—\(i.e.,\) we use \(\chi^{2}\) as the base. In model (2), we
see a large FYE effect: inventory is 10.3% lower in the fourth fiscal quarter, $\phi^4$. As expected, $\phi^3$ is statistically weak. Interestingly, $\phi^1$ is positive, a point we address in the next section. We also note that calendar quarter effects are much diminished, with just 5.1% for $\chi^4$, suggesting that the FYE effect might be even more substantive than the calendar effect.

For robustness, we consider the possibility that demand seasonality might vary by industry. In models (3) and (4), we show calendar quarters interacting with two extremes of NAICS classification, at 2 digits and all 6 digits. The main result holds: inventory is lower by about 10% in the fourth fiscal quarter.

Thus, RadioShack is not an isolated case. Inventory is about 10% lower in the fourth fiscal quarter, about $47$ billion based on average quarterly values in 2006 in our dataset of U.S. firms. This raises the question: why?

3. Literature Review and Hypotheses Development

We organize previous research by the hypotheses that might explain the FYE effect. In figure 3, we summarize these hypotheses and their predictions, which are of three types:

1. **Effects** of how fiscal quarters affect inventory levels—e.g., low in some fiscal quarter;
2. **Mediators** through which fiscal quarters affect inventory—e.g., fourth fiscal quarters have lower inventory, via higher sales. Mediators are sometimes called channels of influence, mechanisms, pathways, or intervening variables;
3. **Moderators** that affect the strength of the link from fiscal quarter to inventory—e.g., fourth fiscal quarters have especially lower inventory for the sub-sample of firms with stronger bonus incentives to reach sales targets. Moderators are sometimes called cross-sectional predictions or interaction effects.

3.1 Baseline Hypothesis: Sales Timing

With sales timing, executives’ private benefits lead them to pull some next-quarter sales into fourth fiscal quarters, depleting inventory in a way that is not compensated by
increased production or purchasing. There is past research on why and how this happens:

**Why?** There are at least five motivations for sales timing. One has to do with sales bonuses. Joseph and Kalwani (1998) find that 95% of 215 senior and sales executives surveyed are rewarded on bonuses, largely structured as non-linear functions of sales levels determined at FYE. Such bonus structures could lead executives to time sales so as to make targets. This phenomenon has been studied as the hockey stick effect in Chen (2005) and Sohoni, et al. (2005) and as push contracts in Lariviere and Porteus (2001) and Taylor (2006).

Another motivation is the time value of bonuses (Jensen and Murphy (1990)). Bonuses associated with a sale right after the current fiscal year might be paid at the end of the next fiscal year, so there is incentive to book that sale before the current fiscal year ends.

Third, the equity market is more sensitive to sales figures in the fourth fiscal quarter than other quarters (Collins, et al. (1984), Mendenhall, et al. (1988)). If executives avoid drops in their firms’ equity prices—perhaps because of their equity interests (Jensen and Murphy (1990))—then they also have incentives to time sales into the fourth fiscal quarter.

Fourth, executives are more likely to resign just after getting their year-end bonuses (Blakemore, et al. (1987)), so there is incentive to time sales into the current fiscal year if the bonus associated with a next-year sale after the resignation is discounted or forfeited.

Finally, executives might be motivated to reach sales goals even if these are not associated with explicit bonuses. There might be career concerns (*e.g.*, Holmstrom (1999)) or simply psychological motivation (*e.g.*, Holmstrom and Milgrom (1987)).

**How?** Just how does sales timing cause inventory to be lower at FYE? This requires: (1) sales to be higher and (2) production not replenish inventory at the higher rate. On the former, Oyer (1998) shows that, at the industry level, sales are 2.6% higher in FYE. But how do firms enhance sales? One possibility, as Oyer (1998) also shows at the industry level, is that firms cut prices by 1.6% at FYE. His finding is supported by Nevo and Wolfram (2002), Larkin (2006), Roychowdhury (2006), and Chapman and Steenburgh (2007), as well as the large trade promotion literature, such as Krishnan, et al. (2004). Oyer (1998) and Chapman
and Steenburgh (2007) also show that timing is especially prevalent in durable goods industries, since it is harder to get customers to take perishable inventory.

It has also been suggested that production might not replenish inventory to compensate for the higher sales. One reason is that the sales department is not necessarily aligned with production. For example, Porteus and Whang (1991) point out that a sales department might want higher sales, but a production department is incentivized to keep inventory low to reduce holding costs. Second, even where incentives are aligned, production often relies on demand forecasts provided by the sales department. If these forecasts are used to set sales targets, and given the sales department’s incentive to meet sales targets, its forecasts—especially those at FYE—are often sandbagged (see Davis and Mentzer (2007) for a review). Working with these low-balled forecasts, production might not be able to produce or buy enough to maintain inventory levels. Third, the kind of bonus incentives described as motivations for sales timing is much less prevalent among production functions, in the U.S., the U.K., or Australia (Heywood, et al. (1997)). Using Australian establishment data, Drago and Heywood (1995) report that bonuses apply to just 1.1% of the workforce.

**Predictions.** Taken together, the above imply specific predictions (recall figure 3):

- **Effects. P1a:** *Inventory is lower at FYE.* This follows from our discussion of how sales timing leads to lower inventory at FYE.

- **Mediators. P1b, P1c:** *FYE leads to lower inventory via lower margin and higher sales.* This also follows directly from how sales timing works.

- **Moderators.** The following are from our discussion of why firms want to, can, and are sometimes prevented from sales timing.
  - **P1d:** *FYE effects (P1a) are stronger when executive pay has a higher bonus portion.*
  - **P1e:** *FYE effects are stronger for firms in durable good industries.*
  - **P1f:** *FYE effects are stronger for firms under less scrutiny.* Such scrutiny might be by auditors, analysts, board members, or regulators, to name some examples.

**Post-FYE Effect.** We now turn to the observation from the U.S. dataset in the previous
section, in which inventory is abnormally high in first fiscal quarters. In sales timing, customers might stockpile that goods pushed to them at FYE that demand is dampened in the next quarter. A firm also has less visibility on inventory “stuffed” down the supply chain, since there is likely to be stuffing by competitors, too (Armony and Plambeck (2005)). Finally, returns might also be more likely. In the more egregious cases, firms might even provide favorable terms to customers for returns, a practice called round-tripping—e.g., the SEC (2004a) alleges that Bristol-Myers Squibb provides such guarantees, so that $35 million in inventory was returned to the company right after FYE in 2001.

These suggest the following additional predictions (figure 3 again):

- **Effects.** P1g: Inventory is higher post-FYE.
- **Mediators.** P1h: Post-FYE leads to higher inventory via lower sales.
- **Moderators.** The moderators for FYE effect also apply to the post-FYE effect.

For clarity, we use “FYE effect” to mean lower inventory at FYE and “post-FYE effect” to mean higher inventory in the quarter right after FYE.

### 3.2 Alternative Hypotheses

Our point in this section is that the alternative hypotheses have predictions that are sufficiently different than those of sales timing that we can empirically identify sales timing.

**Sales effort hypothesis.** Even with the motivations for sales timing just described, executives could simply exert more effort to generate sales at FYE, without having to pull in sales that might naturally occur in the next quarter (Basu, et al. (1985), Kocabiyikoglu and Popescu (2007)). There is also empirical literature supporting a sales effort story. However, much of these investigate sales effort averaged over time, and not specifically at FYE—e.g., Bratkovich and Steele (1989) and Lazear (2000). An important exception is Steenburgh (2004), who finds that—for an office products manufacturer—“lump-sum bonuses primarily motivate salespeople to work harder” (pg. 1).

The sales effort story predicts that (figure 3 again):
• **Effects.** **P2a:** Inventory is lower at FYE. As an alternative explanation, sales effort has this same prediction as sales timing;

• **Mediators.** **P2b:** FYE leads to lower inventory via higher sales. However, unlike sales timing, there is no prediction of lowering margins to enhance sales;

• **Moderators.** **P2c:** FYE effects (P2a) are stronger when executive pay has a higher bonus portion. However, while sales timing is easier for firms pushing durable goods down the supply chain, there is no such prediction here in the sales effort story. Nor is there any prediction about firms under greater scrutiny.

**Stock taking hypothesis.** In this story, FYE is associated with activities that mechanically lead to lower inventory. For one, FYE audits are often when firms write off inventory, so that could explain lower reported inventory at FYE. Importantly, FYE audits are mandated by accounting guidelines while interim quarter audits are sometimes not: “even when regulations require a review of interim earnings by auditors, the review can be done at the time of the annual audit” Basu, et al. (2001), pg. 4). The write-offs at FYE could be larger those at interim quarters because FYE audits are more conservative (Basu, et al. (2001)) and the write-offs in the interim quarters tend to be unreported and get booked only at FYE, in a process often called “settling up” (AICPA (1973)).

One other way in which inventory might be lower at FYE is that firms often produce less, to simplify the “sight audit” of inventory.

Stock taking then produces the following predictions:

• **Effects.** **P3a:** Inventory is lower at FYE. Again, this is why stock taking is a possible explanation for the FYE effect.

• **Mediators.** **P3b, P3c:** FYE leads to lower inventory via higher write-offs and lower production. This is a prediction not from sales timing or sales effort.

• **Moderators.** We find no theoretical consensus on how the FYE effect under stock taking would be moderated, so we make no explicit prediction. However, intuition suggests that the predictions are the opposite for those for sales timing.
For example, there is no prediction for the bonus moderator. And the FYE effect could be *weaker* for firms in durable goods industries if write offs are greater for perishable goods. The effect is also *weaker* if firms face more scrutiny; they more diligently write off in the interim quarters rather than “settle up” only at FYE.

**FYE setting hypothesis.** This is a story of endogeneity, in which firms set their FYE to when inventory is the lowest. One motivation for this is that equity markets assign higher valuations to firms with low or decreasing inventory—*e.g.*, Thomas and Zhang (2002), Lai (2006). Also, the equity market is more sensitive to figures disclosed at FYE than at other times (Collins, et al. (1984), Mendenhall, et al. (1988)), and if executives have interests in their firms’ equity prices, then firms have incentives to set the FYE to when inventory is lowest or have decreased the most. The same mechanics applies not only to low inventory, but to high sales, so if high sales are linked to low inventory as described above for sales timing, we again have FYE setting.

Firms might also “use a fiscal period to attempt to measure performance at a time when they have concluded most operating activities” (Stickney and Weil (2000), pg. 102), which could be the time when inventory is most depleted. It is also easier to conduct inventory audits and more resources are freed up to close the year’s accounting books.

The FYE setting hypothesis has one only prediction:

- **Effects. P4a:** *Inventory is not lower at FYE, after accounting for endogeneity.* This is different than the earlier predictions.

We conclude this section by pointing out that of all the hypotheses, sales timing seems to have the most deleterious impact on firm valuation. Sales timing involves margin discounting, more returns, and shifting of sales that could damage brand equity. Sales effort, on the other hand, might even be a positive, if the extra effort produces incremental revenues that outweigh the bonus payment. Stock taking seems neutral, because lower inventory just more accurately reflects the true inventory level. FYE setting also seems neutral, since it involves just an accounting choice in setting the FYE. To reiterate the point
made in the introduction, it is for these reasons that we focus on sales timing.

We now turn to tests of the predictions. In figure 3, we summarize all key findings in the rightmost column. The evidence is consistent only with sales timing, and is not consistently explained by the alternative hypotheses.

4. Test of “Effects” Predictions Using a German Natural Experiment

We first test for sales timing by investigating the “effects” predictions.

4.1 A Natural Experiment

The biggest empirical challenge is to rule out endogeneity in FYE setting. To be sure, our U.S. results in table 2 already address endogeneity to a degree, and it is one of the arguments used in Oyer (1998). The idea is that if endogeneity is present, then finer controls of calendar quarters, using interactions with finer industry classifications, should reduce the significance of the FYE effect. This is not the case in models (3) and (4) in table 2. Still, this evidence is indirect. Another approach is to see if the FYE effect is still significant after firms change their FYEs, as in RadioShack. The identifying assumption is that demand patterns do not change as quickly as the FYE changed, over one year. In regressions using specification (A) but only the 32 firms in the U.S. dataset that changed their FYE, we find that inventory is 15.4% ($p=0.000$) lower at the old and new FYEs. Unfortunately, from Factiva news reports on the circumstances surrounding FYE changes, we find that the changes are often made to synchronize FYEs after mergers and acquisitions of companies, so there might be confounding changes in product lines that affect inventory patterns.

We find a unique natural experiment that can more cleanly rule out FYE setting. In 2000, Germany reduced its corporate tax rate from 40% to 25%. Consider a hypothetical firm with FYE in the middle of the calendar year, as in figure 4. The law stipulates that firms pay the tax for the full fiscal year at the rate at the start of that year. This is unlike many tax reforms, such as those in the U.S. The left panel shows the firm if it does not change its FYE.
It pays 40% for two fiscal years (indicated by the horizontal full black lines), and 25% thereafter. The right panel shows the firm if it changes its FYE in 2000 to end in calendar 2000. It pays the lower 25% for the full 2001 year, capturing the tax savings shown. Thus, some German firms change their FYE to end with the calendar year. This change is not exogenous to factors such as taxable income or the cost of FYE change. It is, however, plausibly exogenous to inventory patterns, our dependent variable.

4.2 Data

We start with a panel dataset of all 661 German manufacturers, wholesalers, and retailers in CapitalIQ. The data is spotty on quarterly inventory data, which we hand-code from primary sources—annual and quarterly reports, direct communications with the firms. Among the 76 firms whose FYE are not already at the end of the calendar year, 19 change their FYE. The rest have not changed presumably because the tax benefit is smaller than the cost of the change. We summarize the data in table 3.

One complication is that we obtain only total inventory, not finished goods inventory. In the U.S. data, the correlation coefficient between finished goods and total inventory is 0.78 when measured in absolute levels, and 0.94 when measured as quarters of COGS. So this issue may not be material, but our results are subject to this qualification.

4.3 Empirical Approach

Our empirical model is a modification of specification (A), with an interaction indicator that is coded 1 for observations after the 2000 tax change:

\[ v_{it} = \sum_{f=1}^{4} t_i^f \phi_i^f + \sum_{f=1}^{4} \left[ t_i^f \phi_i^f \times I(\text{after 2000}) \right] + I(\text{after 2000}) \sum_{c=1}^{4} t_{ic}^c X_{it} + \gamma_y + \kappa + \eta_{it}. \]

We are interested in \( \phi_i^f \) and \( \phi_i^l \) (the latter for the post-FYE effect). Further, we expect that the interaction terms are zero—i.e., \( \phi_i^f \) and \( \phi_i^l \) effects are maintained over the FYE change. For firms that change their FYE (and for firms that already have their FYE at calendar year end), the interaction terms pick up the post-change updating of both \( \phi \)'s and
\( \chi \)'s. But in practice, these interactions are likely to be just \( \phi \) updating. First, we have no \textit{a priori} reason to believe that the \( \chi \) update is anything but zero. Second, in a direct estimation of (C) without the \( \phi \)'s and with the \( \chi \)'s interacted with I(\textit{after2000}), we find that the \( \chi \) update is indeed statistically indistinguishable from zero \((p=0.56)\). Third, if we find the interaction terms to be statistically indistinguishable from zero (as is the case, see below), it is unlikely that the \( \phi \) update is not zero and is exactly cancelled by an opposite \( \chi \) update.

To be conservative, in this as well as subsequent estimations reported here, we measure \( \phi^4 \) and \( \phi^1 \) using \( \phi^2 = \phi^3 \) as a baseline, so the resulting \( \phi^4 \) and \( \phi^1 \) estimates can be interpreted as deviations from both bases, rather than just from \( \phi^2 \), as in the previous U.S. estimation. We also undertake estimations with just \( \phi^2 \) as a base and the results are similar. Further, the \( \phi^3 \) estimate is always economically or statistically insignificant, or both.

Given the small sample size, we report results using first differences (FD) rather than fixed effects (FE) estimation because the bias of FD estimators is independent of sample size, while that for FE vanishes at the reciprocal of sample size. Also, FE estimators are more sensitive to non-normality of the disturbance term (Wooldridge (2002)). For robustness, we also execute FE estimation, which produces similar results and are unreported.

The small sample size means that our test has low power, but that suits our empirical objective since it only biases us against finding sales timing as a cause even when it is one. Another issue is potential sample selection bias. We attend to that after the results below.

4.4 Results

In Table 4, panel (a), we report our FD estimates under specification (A). The FYE effect of 21.5\% is maintained over the FYE change.\(^2\) This is not explained by the FYE setting

\(^2\) That the German effect is larger than the U.S. one is not germane to our empirics, but is consistent with the large “law and finance” literature (e.g., La Porta, et al. (1998)) suggesting that the German commercial code provides weaker corporate governance than common law in the U.S.
hypothesis. We also note that, as in the U.S., there is a post-FYE effect that is not explained by the sales effort and stock taking hypotheses (P2a and P3a). Taken together, these results are consistent only with sales timing (P1a and P1g). We also note that the interaction terms are not economically and statistically significant, as predicted. Given the small sample, the $p$ value for the entire specification is unsurprisingly high.

4.5 Treatment Selection Bias

The small sample size may involve treatment (i.e., whether a firm changes its FYE) selection bias. Heckman (1979) notes that there are two types of potential bias: selection on unobservables and on observables. Specification (C) is a treatment equation, and conceptually, there is also a selection equation that models how firms select to change their FYE. Selection on unobservables arises from correlation between the disturbances in the two equations, biasing our estimation. Selection on observables arises from omitting selection variables in the treatment equation.

Following Heckman (1979), we address selection on unobservables with a Heckit procedure. Since we have panel data, we modify his first stage estimation to use a population-averaged probit model for the selection of observations—see Kyriazidou (1997):

\[ \zeta^*_i = \pi_i + \omega_i + \mu_i, \]

where $\zeta^*_i$ is a latent continuous dependent variable representing the selection, $\pi_i$ represents covariates that could explain the selection, $\omega_i$ accounts for unobserved firm fixed effects, and $\mu_i$ is white noise. We then define a dichotomous dependent variable:

\[ \zeta_i = \begin{cases} 1 & \text{if } \zeta^*_i \geq 0 \\ 0 & \text{otherwise.} \end{cases} \]

We include several covariates in $\pi_i$. The first is taxes saved, calculated as in the blue area in figure 4, based on firms’ reported taxable profits. The second is the counter-balancing cost of changing FYE, using log COGS as a proxy. The third is an indicator for whether a firm already has its FYE at calendar year end, in which case there is no need to
change FYE. For the first stage, we run specification (D) with \( \zeta_{it} \) on all observations.

The second-stage regression is standard, and includes the inverse Mill’s ratio as an additional covariate in the treatment equation. We obtain a Chi-square of 79 \((p=0.000)\) in a test for exclusion restriction.

The result is in table 4, panel (a), model (2). It is almost identical to our FD estimates. The inverse Mill’s ratio is insignificant, suggesting that there is not a large bias in the first place. We are concerned that firms might have anticipated the tax change, and conversely, inventory pattern changes might have a lag after the change. But our estimates are robust to various pre-change and post-change windows, with the former at various end years (1998 through 2000), and the latter, various start years (2000 through 2002).

To address treatment on observables, we use the propensity scoring method proposed by Rosenbaum and Rubin (1983). The method is designed for only one treatment at a time, so for us, it is more natural now to test for the \( \phi^4 \) and \( \phi^1 \) effects separately. We do that in a differences-in-differences framework. Consider \( \phi^4 \). We ask whether after 2000, the fourth calendar quarter \( (i.e., \) also the fourth fiscal quarter then) has lower-than-annual-average inventory than the fourth calendar quarter before 2000—the “first difference”:

\[
F_{it} = \begin{cases} 
F_{it}(0) & \text{if } C_{it} = 0 \text{ (no change),} \\
F_{it}(1) & \text{if } C_{it} = 1 \text{ (change).}
\end{cases}
\]

\( F_{it} \) equals how far inventory at the fourth calendar quarter is below the annual mean, for firm \( i \) in calendar quarter \( t \). \( C_{it} \) is whether the firm changes its FYE—\( i.e., \) the treatment. We want the average treatment effect (ATE):

\[
\tau_{it} = E[F_{it}(1) - F_{it}(0)].
\]

To address selection bias, we compare the first difference for treated firms against that for untreated firms—hence, the “second difference.” The difficulty, of course, is that we observe \( F_{it}(1) \) only if firm \( i \) in the treated group and \( F_{it}(0) \) only if it is in the untreated group.

The propensity scoring method develops a score as the probability of selecting into the
treated group, conditional on some observable matching covariates $M_{it}$. We use as covariates the three used in the Heckit procedure, plus firm fixed effects. In robustness tests, we build the propensity score using just year 2000 data (no firm fixed effects, but focused at the year of change) and we obtain similar results.

Heckman, et al. (1998) note that the method performs well under three conditions:

1. **Ignorability.** Selection is ignorable conditional on the matching covariates:

   \[ C_{it} \perp (F_{it}(1), F_{it}(0)) | M_{it}; \]

2. **Common support.** The intuition is that we require the probability distribution of the matching covariates to be bounded away from zero for the treated observations, on the range of values taken by the untreated observations:

   \[ \exists \epsilon > 0 : \epsilon < \Pr(C_{it} = 1 | M_{it} = m_{it}) < 1 - \epsilon, \text{ for all } m_{it} \text{ in the support of } M_{it}; \]

3. **Heterogeneous distributions.** The propensity score distribution of the treated is skewed toward higher values than that of the untreated.

We have addressed ignorability with our Heckit procedure (and find no significant bias from that). To address the last two conditions, we implement $k$th-neighbor matching with caliper restrictions, so as to contain matching within specific ranges.

In table 4, panel (b), we see that inventory is 19.8% lower at FYE and 16.9% higher post-FYE. The estimates are robust to calipers from 0.001 through 0.5, and $k$th-neighbors from 1 through 20. We also report (not in the table) that for $\phi^t$, the treatment effect on just the treated (called ATT) and untreated (ATU) are similar, at -0.193 and -0.206 respectively. For $\phi^t$, these are both 0.169. These similarities suggest that the treated are representative and endogeneity is probably not a big issue, as is the case with the U.S. dataset.

Taken together, the above provide a clean test that sales timing is a cause of the FYE effect, and at the magnitudes estimated, a sizeable cause at that.
5. Tests of Mediators and Moderators

We now turn to tests of mediators and moderators (figure 3 again). We report results using the U.S. dataset because it has higher power, it appears that endogeneity is not significant, and the German dataset has limited variables. Wherever variables are available in the German data (COGS, margin, and production as mediators), we run the estimations on it and find qualitatively the same results.

5.1 Mediators

We describe our empirical approach, data, and results.

**Empirical approach.** We follow Baron and Kenny (1986) and test each mediator by running two estimations: regressing the mediator on specification (A) covariates and estimating specification (A) with the mediator as an additional covariate. From these estimations, we construct a $Z$ statistic to summarize the presence of mediation. We construct three versions of $Z$ proposed—Sobel, Goodman, and Aroian—and obtain the same significance on each. Here, we report the Aroian $Z$ because MacKinnon and Warsi (1995) show that it does not assume that the multiple of the standard errors from the two estimations are vanishingly small and it performs best in Monte Carlo studies.

**Data.** We use COGS as a measure of sales, to separate sales from gross margin, our other mediator. Gross margin is defined as one minus COGS divided by revenues. Production is defined as change in inventory plus COGS. We do not have inventory write-off data, but only general write-offs, so that write-off results must be viewed with caution.

**Results.** In Table 5, panel (a), we report the results of our tests. The evidence is that the FYE effect on lower inventory is mediated by higher sales, as predicted by sales timing and sales effort, but not by the other two hypotheses. The FYE effect is also mediated by lower gross margin and the post-FYE effect, by lower sales. These are both predicted by sales timing ($P_{1c}$ and $P_{1h}$) and not by any of the alternative hypotheses, not even by sales effort. Finally, we find weak or no evidence of mediation via write-offs or lower production ($P_{4b}$,
This is not only supportive of sales timing, but it is also not supportive of stock taking.

5.2 Moderators

In figure 3, we have predictions for three moderators: bonuses, durability, and scrutiny.

**Empirical approach.** The bonus test checks if the FYE effect is stronger for firms whose executives have higher bonus components. There are two standard ways to check this: interact the fiscal quarter effects in specification (A) with a bonus measure, or estimate specification (A) using sub-samples with high and low bonuses and see if the FYE effect is stronger in the high-bonus sub-sample. We do both but report results from the latter, which does not assume that high- and low-bonus firms have the same covariate estimates or the same distribution in their disturbances (Brame, et al. (1998)). It is less information-efficient but with our large U.S. dataset, the estimation will not be too noisy.

The durability test checks if the FYE effect is stronger for firms in durable goods industries. The mechanics is the same as for the bonus test.

For even sharper results, we report here a test that interacts the bonus and durability sub-samples, instead of tests with them as univariate partitions of sub-samples (which we do, with stronger results than reported here).

Finally, we test the “scrutiny” moderator with an event study and estimate if a firm’s FYE effect is weakened after it faces a federal class action suit. We employ this different empirical approach because we are concerned about reverse causality: a strong FYE effect might lead to greater scrutiny. Reverse causality is less likely in the durability test, and in the bonus test, if it happens, it works in our favor since a strong FYE effect might lead to compensating lower bonus as a component of total compensation.

**Data.** We construct a concordance of our U.S. dataset with the three moderators. We obtain firm-year bonus data from ExecuComp. ExecuComp provides bonus data at the executive-year level, so we construct our firm-year bonus measure as the median executive bonus (as a percent of the executive’s total compensation) each year. We include only
executives with an explicit sales or marketing function. Given that ExecuComp captures bonuses for only the top few executives (up to 15, median is 5 executives), this is only a proxy for sales bonuses company-wide, so our result is subject to this caveat. We tag each observation in our U.S. dataset with an indicator of whether the firm is in a durable good industry, as defined by the U.S. Census. The Census defines a good as durable if its life expectancy is three years or more. Finally, we obtain federal class action suits from the Stanford Securities Class Action Clearinghouse. We include all 39 suits in all years available (1996-06) that have at least one of these words: sale*, revenue*, inventor*, bonus*.

**Result.** In Table 5, panel (b), we report the results of FYE and post-FYE effects in four sub-samples constructed by dividing firm-quarter observations using the median bonus and durability. The top-left sub-sample, with higher bonus and in durable goods, has the strongest effects, and bottom-right has the weakest. We compare the effects across every pair of sub-samples using standard errors as in Brame, et al. (1998) and Paternoster, et al. (1998), and the differences are as predicted. For example, the top-left of the FYE matrix is significantly different ($p=0.08$) from the bottom-right. It is also different ($p=0.09$) than the bottom-left, which is less significantly different ($p=0.12$) than the top-right. The difference along the bonus dimension for the FYE effect is consistent with sales timing (P1d in figure 3) and sales effort (P2c), but is not explained by alternative hypotheses. The difference along the durability dimension and the results for post-FYE effects fit only the sales timing predictions (P1d and P1e), and is not explained by others, not even by sales effort.

For robustness, we divided the bonus dimension into not just two fractiles, but three or four. The results are qualitatively the same and are not reported here.

In Table 5, panel (c), we show the results of interactions of the fiscal year effects with an indicator for whether observations are before or after the suit. The results are robust to various definitions of before- and after-windows, as well as to whether we include the year of the suit in either window, or neither. In models (1) and (2), we show two examples where the before window is years (-19,-2)—where 0 represents the year of the suit—and the
after windows are (0,1) and (0,5).

In model (1), we estimate these firms have a staggering 33.1% lower inventory at FYE before the suit. After the suit, this inventory dip is reduced by 17.6 percentage points. It is this that is consistent with the scrutiny prediction in sales timing ($P1f$ in figure 3), and is not explained by other hypotheses. Also, one interpretation is that the 17.6% represents the explanatory power of the alternative hypotheses, and the difference ($33.1 - 17.6 = 15.5$) represents that of sales timing. That would suggest that sales timing is at least as important as other explanations. The estimates in model (2), with a longer after-window, are similar. We do not find significant post-FYE effects. This might be due to noise with the small number of observations.

6. **Financial Implication of the FYE Effect**

We first show evidence that FYE and post-FYE effects are associated with lower firm value. These effects might be picking up broader characteristics like governance or could be proxies for other characteristics like operational competence, so we run firm fixed effects regressions that partial out time-invariant characteristics. Since these characteristics might also change over time, we next dig deeper to directly check that lower valuation is due to lower gross profits and higher costs. Lower gross profits could arise if price discounting is not compensated by higher sales. Higher costs might arise if inventory fluctuations lead to higher inventory holding costs, and sales fluctuations to higher capacity investments.

6.1 **Fiscal Quarter Effects and Valuation**

For the reasons before, we report results using the U.S. dataset. Our estimations using the German data produce qualitatively the same results and are not reported here.

In first-stage regressions, we estimate firm-specific FYE and post-FYE effects using specification (A) without the firm effects term. In the second stage, we see if firms with stronger effects have lower valuations. We prefer that the latter include firm fixed effects to
account for unobserved firm heterogeneity, so we run our first-stage regressions on periods of the dataset to obtain a time series of effects.

In the second stage, we follow standard $q$ regressions by Shin and Stulz (2000). We use Tobin’s $q$, a standard measure of valuation, and regress it on the first stage FYE and post-FYE effects, with firm fixed effects and the log of total assets. The industry-adjusted $q$ is the sum of total assets and market capitalization, less common equity and deferred taxes, as a deviation from the industry median $q$. We use the NAICS 3-digit industry classification.

In Table 6, panel (a), we see that 1 percentage point in FYE effect (lower inventory) is associated with 1.74% lower valuation. The post-FYE elasticity of valuation is -3.56, signed as expected. Here, the first-stage regressions use the dataset divided into 4 periods, but the result is robust to dividing into 5 through 10 periods. It is also robust to Fama and MacBeth (1973) estimation in the second stage, which accounts for serial correlation.

The higher elasticity associated with post-FYE effects is intriguing. We conjecture, but leave for future research, that this is consistent with two forces: (1) a sign effect, in which the equity market is more concerned about inventory peaks at post-FYE than troughs at FYE, because the former are due to lower sales and the latter to higher sales (achieved with lower margins); and (2) an uncertainty effect, in which inventory peaks are unambiguously bad (due to sales timing) but troughs might be good or bad (part of it due to sales timing, part to any or all of the alternative hypotheses).

Although the elasticities are large, we make no claim about whether shareholders suffer. For example, shareholders may not have suffered if they have priced the lower valuation at the time of contracting—i.e., when they buy into the firms’ shares (Christie and Zimmerman (1994)). Nevertheless, the valuation elasticities suggest opportunities for improving firm valuation, which we discuss in the concluding section.

6.2 Plausible Explanations of $q$ Regressions: Gross Profits and Costs

Why would the FYE and post-FYE effects lower valuations? We suggest two reasons:
• **Lower gross profits.** Table 6, panel (b) shows how much sales (COGS) and gross margin—mediators from the previous section—vary at fiscal quarters. At FYE, sales are 4.9% higher and margins 3% lower; gross profits are 0.5% higher, although that is not statistically distinguishable from zero. But post-FYE, sales are 4.8% lower while margins are not any higher; gross profits are 5.0% lower. The net 5% loss in gross profits paints a picture of sales timing being injurious to the firm’s long-term valuation, and is consistent with the literature on managerial short-termism—e.g., Lai (2006).

In unreported estimations, we explicitly test gross profit as a mediator in the previous q regression and find that it does play that role—e.g., regressing log gross profit on fiscal quarter effects produce the predicted signs (both negative) and including log gross profit into the q regression eliminates the significance of fiscal quarter effects in panel (a).

• **Higher costs.** This could arise from two other factors:
  - *Inventory fluctuations increase Inventory holding costs.* If cost is a convex function of inventory, then varying inventory implies higher cost. Our findings suggest that firms are not factoring sales timing into inventory management in post-FYE, resulting in higher inventories;
  - *Sales fluctuations increase capacity investments.* Since we find sales timing as a cause, sales vary, too. That in turn means capacity investments might be more than if there were no FYE effect.

We are not ready to estimate the financial impact of the above, but it seems reasonable that, with the FYE effect on the order of 10%, the financial impact could be significant.

7. **Discussion and Conclusion**

This study has several limitations and paths for future research. First, we have not considered the relative importance of sales timing versus the alternative hypotheses. The evidence is that sales timing is an important cause, but we have not really rigorously examined the relative importance among hypotheses. Second, it would be interesting to
consider interactions between hypotheses. For example, are sales timing and sales effort complements or substitutes? Third, we have not rigorously considered inventory carrying costs and capacity investments. A more rigorous treatment of these and welfare effects in general is needed. Finally, it would be intriguing to test the FYE effect on other operational measures, such as capacity utilization in service industries, or R&D spending.

We conclude with some example implications:

• **Research.** One implication for empirical work is that structural estimations of inventory levels with covariates that correlate with fiscal quarters should include the latter, or suffer omitted variable bias. For example, if one wants to use seasonality factors to forecast demand, the estimation of these factors should partial out fiscal effects. Separately, it is also desirable to enhance models of inventory management to account for sales timing in particular and agency issues in general, not just between firms and their suppliers, but also between firms and their shareholders.

• **Practice.** These implications for practice are more for shareholders than executives, since we find that the FYE effect is caused by sales timing, an agency behavior by executives themselves. For shareholders, we may divide the implications into two types:
  - **Screening firms for investing.** With the large elasticity of valuation to FYE effects, private equity investors might find it attractive to acquire firms with high FYE effects and reduce these after acquisition. Conversely, public markets investors might consider low FYE-effect firms if there is a market flight to quality. While agency behavior in general is difficult to observe for outsiders, inventory and sales levels are publicly released data, and with appropriate empirics, FYE effects can be estimated. Furthermore, one can rely on mediator and moderator effects (margins, product durability) to ascertain the magnitude of inventory’s FYE effect.
  - **Improving valuation after investing.** One of our more interesting findings is that firms do not appear to account for sales timing in their inventory management post-FYE. Correcting this seems to be a candidate for improvement that could lead to lower
inventory holding costs. One could also run through the why’s and how’s for sales timing in section 3.1 to design a program to improve valuation. For example, one of the why’s is that both firm and equity market overweight financial disclosures at FYE; this could be mitigated with more emphasis on interim reporting. Another example might be to increase scrutiny, perhaps with independent board members and rotating auditors. Interestingly, our \( q \) regressions—which account for industry effects—suggest that it is possible to have lower FYE effects and higher valuation even if competitors do not.

This study extends previous research—on sales seasonality, inventory seasonality, and sales FYE effects—to inventory’s FYE effect. We find that sales timing is an important cause and measure the financial implications. We hope others will build upon these findings to develop a more complete picture of how firms actually manage inventory.

8. References


SEC. Sec Vs. Clearone Communications, Frances M. Flood, and Susie Strohm, Civil No: 2 103 Cv 055 Dak. Salt Lake City, Utah, 2003
---. Sec V. Bristol-Myers Squibb Company. Newark, New Jersey: SEC, 2004a
---. Sec Vs. Various K-Mart Executives. Detroit, MI: SEC, 2004b
---. Sec Administrative Proceeding File No. 3-11902 Vs. Coca Cola. New York: SEC, 2005
---. Sec V. Virbac Corporation, Civil Action No. 4-06cv-453-A. Fort Worth, TX: SEC, 2006
Figure 1—RadioShack Inventory Over Time

The vertical axis is RadioShack level in finished goods inventory, measured in quarters of cost-of-goods sold (COGS). The squares show inventory at the last quarter of fiscal years and the circles, at other quarters. Until 1991, the fiscal year ends in the second calendar quarter. After that, it changes to end in the fourth calendar quarter.

Figure 2—Inventory’s Fiscal Year End (FYE) Effect in the Context of Current Research

Citations in cells are example studies, but are not meant to suggest number of studies. There are considerably more studies on calendar year than fiscal year effects.

<table>
<thead>
<tr>
<th>Calendar year effects (“seasonality”)</th>
<th>Fiscal year effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td></td>
</tr>
<tr>
<td>Inventory</td>
<td></td>
</tr>
<tr>
<td>Nerlove, et al. (1993)</td>
<td></td>
</tr>
<tr>
<td>Rajagopalan and Malhotra (2001)</td>
<td>Current study</td>
</tr>
<tr>
<td>Gaur, et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>Netessine and Roumiantsev (2007)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3—Hypotheses, their Predictions, and Findings

“FYE” is fiscal year end, econometrically implemented as the last quarter of the fiscal year; “post-FYE” means the quarter after that. A dash means no prediction.

<table>
<thead>
<tr>
<th>Effects</th>
<th>Baseline hypothesis</th>
<th>Alternative hypotheses</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>FYE inventory is...</td>
<td>Sales timing</td>
<td>Sales effort</td>
<td>Stock taking</td>
</tr>
<tr>
<td>Post-FYE inventory is...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FYE leads to lower inventory via...</td>
<td>P1b. Higher sales</td>
<td>P2b. Higher sales</td>
<td>P3b. Higher write-offs</td>
</tr>
<tr>
<td></td>
<td>P1c. Lower margins</td>
<td></td>
<td>P3c. Less production</td>
</tr>
<tr>
<td>Post-FYE leads to higher inventory via...</td>
<td>P1h. Lower sales</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FYE and post-FYE effect stronger for firms with...</td>
<td>P1d. Higher bonus %</td>
<td>P2c. Higher bonus %</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>P1e. Durable goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P1f. Less scrutiny</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The idea here is that observed lower inventory at FYE is a result of poor econometrics. After properly controlling for endogeneity, inventory is not lower at FYE.

Figure 4—German Tax Change

In 2000, Germany reduced its corporate tax rate from 40% to 25%. Consider a hypothetical firm that starts its fiscal year in the middle of the calendar year. The laws stipulate that firms pay the tax rate for the full fiscal year, depending on rate at the start of that year. The left panel shows the firm if it does not change its fiscal year end (FYE), so it pays 40% for two fiscal years (indicated by the horizontal full black lines), and then pays 25% thereafter. The right panel shows the firm if it changes its FYE in 2000 to end in calendar 2000. Therefore, it pays the lower 25% for the next full 2001 year, capturing the tax savings in blue.
Table 1—Summary Statistics (U.S. Dataset)
The data is an amalgam from the COMPUSTAT quarterly financials tape, COMPUSTAT annual financials tape, Standard & Poor's ExecuComp database, US Census data on whether an industry sells durable goods, and the Stanford Class Action Suit Clearinghouse. Each observation is a firm-quarter, for all 2,512 U.S. manufacturers, wholesalers, and retailers (NAICS codes 31 through 48) in COMPUSTAT. The data is for 1984 through 2006. All monetary amounts are in US$ millions, unless otherwise indicated.

(a) Variation of fiscal year ends among firm-quarters

<table>
<thead>
<tr>
<th>FYE*</th>
<th>No. of firm-quarters</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8,544</td>
<td>29.8</td>
</tr>
<tr>
<td>2</td>
<td>2,367</td>
<td>8.3</td>
</tr>
<tr>
<td>3</td>
<td>3,434</td>
<td>12.0</td>
</tr>
<tr>
<td>4</td>
<td>14,318</td>
<td>50.0</td>
</tr>
<tr>
<td>Total</td>
<td>28,663</td>
<td>100.0</td>
</tr>
</tbody>
</table>

* The FYE is in one of four calendar quarters.

(b) Firm-quarter observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calendar year</td>
<td>28,663</td>
<td>2005</td>
<td>5.60</td>
<td>1984</td>
<td>2006</td>
</tr>
<tr>
<td>Inventory (qtrs of COGS)</td>
<td>28,663</td>
<td>0.92</td>
<td>12.16</td>
<td>0.0004</td>
<td>955</td>
</tr>
<tr>
<td>-raw mat</td>
<td>13,642</td>
<td>0.30</td>
<td>4.03</td>
<td>0.00</td>
<td>410</td>
</tr>
<tr>
<td>-WIP</td>
<td>13,673</td>
<td>0.11</td>
<td>2.45</td>
<td>0.00</td>
<td>235</td>
</tr>
<tr>
<td>-fin gds</td>
<td>28,663</td>
<td>0.58</td>
<td>10.75</td>
<td>0.0002</td>
<td>929</td>
</tr>
<tr>
<td>COGS (US$ mil)</td>
<td>28,663</td>
<td>71.62</td>
<td>2728.77</td>
<td>0.001</td>
<td>109657</td>
</tr>
<tr>
<td>Sales net (US$ mil)</td>
<td>28,663</td>
<td>115.07</td>
<td>3453.23</td>
<td>0.001</td>
<td>126477</td>
</tr>
<tr>
<td>Gross margin</td>
<td>28,663</td>
<td>0.33</td>
<td>29.94</td>
<td>0.00</td>
<td>1.0</td>
</tr>
<tr>
<td>Production (US$ mil)</td>
<td>27,414</td>
<td>73.77</td>
<td>2441.02</td>
<td>-601.13</td>
<td>68838</td>
</tr>
<tr>
<td>Write-offs (US$ mil)</td>
<td>28,663</td>
<td>0.00</td>
<td>36.48</td>
<td>-3800.00</td>
<td>105</td>
</tr>
<tr>
<td>Bonus (median executive, % of total comp.)</td>
<td>9,112</td>
<td>0.30</td>
<td>0.22</td>
<td>0.00</td>
<td>0.90</td>
</tr>
<tr>
<td>In durable goods industry (indicator)</td>
<td>28,585</td>
<td>0.63*</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Face class action suit (indicator)</td>
<td>28,663</td>
<td>0.09*</td>
<td>0.28</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

* These are means, not medians, which are more meaningful for indicator variables.
Table 2—Identifying the FYE Effect (U.S. Dataset)

We estimate the reduced form:

\[ v_{it} = \sum_{f=1}^{4} \phi_{i}^{f} + \sum_{c=1}^{4} \chi_{i}^{c} + \gamma_{i} + \kappa_{i} + \eta_{it} \]

where \( v_{it} \) is the inventory level, measured as inventory divided by quarterly COGS, of firm \( i \) in fiscal quarter \( f \), calendar quarter \( c \), and calendar year \( t \). \( \phi_{i}^{f} \) is the effect on inventory of being in fiscal quarter \( f \), and \( \chi_{i}^{c} \)'s are indicator variables. \( \chi_{i}^{c} \) is the effect on inventory of being in calendar quarter \( c \). \( \gamma_{i} \) and \( \kappa_{i} \) are calendar year (indexed by \( y \)) and firm fixed effects, and \( \eta_{it} \) is assumed to be white noise. The base fiscal quarter is \( \phi_{2} \) and the base calendar quarter is \( \chi_{2} \). Estimation is done in log form and clustered around firms. Figures in brackets are Huber-White sandwich robust standard errors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiscal quarters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \phi_{i}^{f} )</td>
<td>-1.03***</td>
<td>-0.94***</td>
<td>-0.98***</td>
<td></td>
</tr>
<tr>
<td>Before ( ( \phi_{i}^{f} ) )</td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>After ( ( \phi_{i}^{f} ) )</td>
<td>(0.13)</td>
<td>(0.14)</td>
<td>(0.13)</td>
<td></td>
</tr>
<tr>
<td>Calendar quarters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \chi_{i}^{c} )</td>
<td>-1.13***</td>
<td>-0.51***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before ( ( \chi_{i}^{c} ) )</td>
<td>(0.09)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>After ( ( \chi_{i}^{c} ) )</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Calendar yr effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>N</td>
<td>28,663</td>
<td>28,663</td>
<td>28,663</td>
<td>28,663</td>
</tr>
<tr>
<td>F</td>
<td>6.5</td>
<td>11.9</td>
<td>10.2</td>
<td>11.4</td>
</tr>
<tr>
<td>( p )</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

*** = significant at the 1% level, ** at the 5%, * at the 10% level.

Table 3—Summary Statistics (German Dataset)

This panel dataset of 661 German firms is hand-coded from primary sources—annual and quarterly interim reports, direct communications with the firms—and from CapitalIQ. Tax savings is calculated as in the blue area in figure 4, based on firms’ reported taxable profits and an assumed corporate tax of 40%.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calendar yr</td>
<td>8,944</td>
<td>2,003</td>
<td>1.99</td>
<td>1,997</td>
<td>2,005</td>
</tr>
<tr>
<td>I(not December)</td>
<td>8,944</td>
<td>0.16*</td>
<td>0.36</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>I(change FYE)</td>
<td>8,944</td>
<td>0.05*</td>
<td>0.21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Inventory (qtrs of COGS)</td>
<td>8,944</td>
<td>0.75</td>
<td>10.51</td>
<td>0</td>
<td>385</td>
</tr>
<tr>
<td>Tax savings (E mil)</td>
<td>8,944</td>
<td>0.00</td>
<td>64</td>
<td>-38</td>
<td>865</td>
</tr>
</tbody>
</table>

* These are means, not medians, which are more meaningful for indicator variables.
Table 4—Test of Effects Predictions: Natural Experiment from Germany

The dataset consists of German firms summarized in the previous table.

(a) First Difference and Heckit Estimations
The dependent variable is log inventory. The model is:

\[ v_i = \sum_{j=1}^{4} \phi_j \phi^j + \sum_{j=1}^{4} \left[ \phi_j \phi^j \times I(\text{after} \ 2000) \right] + I(\text{after} \ 2000) + \sum_{c=1}^{4} \gamma_c + \kappa_i + \eta_i \]

where the notation is as in the U.S. dataset (see table 2). \( I(\text{after} \ 2000) \) is an indicator for whether a firm-quarter observation is after year 2000, the year of the German tax reform. Estimation is done in log form and clustered around firms. Figures in brackets are Huber-White sandwich robust standard errors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) First Difference</th>
<th>(2) Heckit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi^4 )</td>
<td>-.215* (.113)</td>
<td>-.215* (.114)</td>
</tr>
<tr>
<td>( \phi^1 )</td>
<td>.147** (.065)</td>
<td>.147** (.065)</td>
</tr>
<tr>
<td>I(After)</td>
<td>.001 (.001)</td>
<td>.001 (.001)</td>
</tr>
<tr>
<td>( \phi^4 \times I(\text{After}) )</td>
<td>.000 (.001)</td>
<td>.000 (.001)</td>
</tr>
<tr>
<td>( \phi^1 \times I(\text{After}) )</td>
<td>.000 (.001)</td>
<td>.000 (.001)</td>
</tr>
<tr>
<td>Inverse Mill’s ratio</td>
<td>.017 (.029)</td>
<td></td>
</tr>
<tr>
<td>Calendar qtr effects</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Calendar yr effects</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>First stage probit unrestricted log likelihood</td>
<td>-166.8</td>
<td></td>
</tr>
<tr>
<td>( N )</td>
<td>262</td>
<td>262</td>
</tr>
<tr>
<td>( F )</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>( p )</td>
<td>.291</td>
<td>.364</td>
</tr>
</tbody>
</table>

(b) Average Treatment Effect (ATE) Estimates using Propensity Scores
The dependent variable is log inventory. We test for the \( \phi^4 \) and \( \phi^1 \) effects separately, using a differences-in-differences framework. For the \( \phi^4 \) effect, we calculate for each firm-year the deviation of fourth calendar quarter inventory from annual average. Then we consider how different post-change deviations are from pre-change ones, comparing this difference for treated versus that for untreated firms.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi^4 )</td>
<td>-.198*</td>
<td></td>
</tr>
<tr>
<td>( \phi^1 )</td>
<td>.169*</td>
<td></td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Caliper</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>( k )-th-neighbors</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>( N ) (on and off support)</td>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>-Treated on common support</td>
<td>52</td>
<td>55</td>
</tr>
<tr>
<td>-Untreated on common support</td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>Selection model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Log likelihood</td>
<td>-50.2</td>
<td>-43.6</td>
</tr>
</tbody>
</table>

*** = significant at the 1% level, ** at the 5%, * at the 10% level.
Table 5—Tests of Mediators and Moderators Predictions (U.S. Dataset)

(a) Mediators
We use the U.S. dataset for these estimations. The predictions refer to those in figure 3. The Aroian Z is:

$$Aroian = \frac{a \times b}{\sqrt{b^2 s^2_a + a^2 s^2_b}}$$

where $a$ is the estimate of the fiscal quarter effect in a regression of the mediator on specification (A) covariates and $b$ is the estimate of the mediator when it is included in specification (A) as an additional covariate. The $s$'s are standard errors of these estimates.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Relationship</th>
<th>Mediator</th>
<th>Aroian $Z$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1b, P2b</td>
<td>$\phi^I$ to lower inventory</td>
<td>Higher sales (COGS)</td>
<td>7.92</td>
<td>.000</td>
</tr>
<tr>
<td>P1b</td>
<td>$\phi^I$ to lower inventory</td>
<td>Lower gross margin</td>
<td>5.79</td>
<td>.000</td>
</tr>
<tr>
<td>P1g</td>
<td>$\phi^I$ to higher inventory</td>
<td>Lower sales (COGS)</td>
<td>2.21</td>
<td>.000</td>
</tr>
<tr>
<td>P4b</td>
<td>$\phi^I$ to lower inventory</td>
<td>Higher write-offs</td>
<td>1.91</td>
<td>.057</td>
</tr>
<tr>
<td></td>
<td>$\phi^I$ to lower inventory</td>
<td>Less production</td>
<td>1.22</td>
<td>.222</td>
</tr>
</tbody>
</table>

(b) Moderators: Bonus × Durability

The dependent variable is log inventory; we run specification (A) on four sub-samples of the U.S. dataset, partitioned by firms’ bonus component paid to their executives (as a percent of total compensation) and durability of goods sold by the firms. The bonus data is from ExecuComp and durability data from the U.S. Census. Estimation is done in log form and clustered around firms. Figures in brackets are Huber-White sandwich robust standard errors. Total $N$ for all four cells is 19,556.

<table>
<thead>
<tr>
<th>FYE effect ($\phi^I$)</th>
<th>Post-FYE effect ($\phi^I$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above median bonus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.45***</td>
</tr>
<tr>
<td></td>
<td>(.027)</td>
</tr>
<tr>
<td></td>
<td>-1.35***</td>
</tr>
<tr>
<td></td>
<td>(.022)</td>
</tr>
<tr>
<td>Below</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.112***</td>
</tr>
<tr>
<td></td>
<td>(.24)</td>
</tr>
<tr>
<td></td>
<td>-.108***</td>
</tr>
<tr>
<td></td>
<td>(.27)</td>
</tr>
</tbody>
</table>

(c) Moderator: Scrutiny

The dependent variable is log inventory; we run specification (A) on the U.S. dataset, combined with federal class action suits obtained from the Stanford Clearing House. $I(\text{afterSuit})$ is an indicator for whether an observation is after the suit has been filed, so the comparison is for fiscal quarter effects in the before- versus after-window. Estimation is done in log form and clustered around firms. Figures in brackets are Huber-White sandwich robust standard errors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Windows (0-year of suit)</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi^I$</td>
<td>-3.31 (0.083)***</td>
<td>-3.31 (0.086)***</td>
<td></td>
</tr>
<tr>
<td>$\phi^I$</td>
<td>-.056 (.067)</td>
<td>-.05 (.066)</td>
<td></td>
</tr>
<tr>
<td>I(suit)</td>
<td>.033 (.185)</td>
<td>.011 (.158)</td>
<td></td>
</tr>
<tr>
<td>$\phi^I \times I(\text{afterSuit})$</td>
<td>.176 (.082)***</td>
<td>.166 (.087)***</td>
<td></td>
</tr>
<tr>
<td>$\phi^I \times I(\text{afterSuit})$</td>
<td>.01 (.062)</td>
<td>.049 (.055)</td>
<td></td>
</tr>
<tr>
<td>Calendar qtr effects × NAICS 3 digits</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Calendar yr effects</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>1,082</td>
<td>1,826</td>
<td></td>
</tr>
<tr>
<td>$P$</td>
<td>.000</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

*** = significant at the 1% level, ** at the 5%, * at the 10% level.
Table 6—Financial Implication of the FYE Effect
Estimation is done in log form and clustered around firms. Figures in brackets are Huber-White sandwich robust standard errors.

(a)—Association with Tobin’s \( q \)
The dependant variable is industry-adjusted \( q \), as a deviation from the industry median \( q \). We use the NAICS 3-digit industry classification. \( \phi^4 \) and \( \phi^1 \) are firm-specific effects in time series obtained from OLS first-stage regressions, using 4-year periods. In this version reported here, only effects significant at the 10% level or better are treated as non-zero. The result is robust to treating all estimates as non-zero, or at the 5% level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Industry-adjusted ( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi^4 ) effects</td>
<td>1.74 (.996)*</td>
</tr>
<tr>
<td>( \phi^1 ) effects</td>
<td>-3.56 (1.59)**</td>
</tr>
<tr>
<td>Log total assets</td>
<td>-.274 (.175)</td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>Y</td>
</tr>
<tr>
<td>N</td>
<td>302</td>
</tr>
<tr>
<td>( F )</td>
<td>2.32</td>
</tr>
<tr>
<td>( p )</td>
<td>.077</td>
</tr>
</tbody>
</table>

(b)—Association with Sales, Gross Margin, Gross Profit
The dependent variables are in the column headings.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi^4 )</td>
<td>Sales (COGS)</td>
<td>Gross margin</td>
<td>Gross profit</td>
</tr>
<tr>
<td>( \phi^1 )</td>
<td>.049 (.011)**</td>
<td>-.030 (.013)**</td>
<td>.005 (.833)</td>
</tr>
<tr>
<td>Calendar qtr effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Calendar yr effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Firm fixed effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>( N )</td>
<td>28,403</td>
<td>27,497</td>
<td>27,497</td>
</tr>
<tr>
<td>( F )</td>
<td>36.5</td>
<td>1.5</td>
<td>30.7</td>
</tr>
<tr>
<td>( p )</td>
<td>.000</td>
<td>.041</td>
<td>.000</td>
</tr>
</tbody>
</table>

*** = significant at the 1% level, ** at the 5%, * at the 10% level
Appendix—Measuring Inventory

There are many views on how “inventory” should be measured. Which ones to use? That depends on the question, which (for us) is: “what does it mean to hold ‘lower’ inventory?”:

1. *End-of-quarter inventory.* This measure has the merit of being parsimonious.

2. *End-of-quarter inventory scaled by quarterly COGS* (cost of goods sold). This is the view in operations management—e.g., Gaur, et al. (2005). Inventory supports demand, so if one observation has a lower unscaled inventory level than another but is associated with a much lower sales level, we would not conclude that this former observation holds less inventory. In this view, proper comparison of inventory requires absolute levels to be scaled by COGS.

3. *End-of-quarter inventory scaled by end-of-quarter total assets.* This view is often associated with the accounting literature—e.g., Roychowdhury (2006). It is analogous to the previous one, but the idea here is that inventory is working capital, so it is comparable only as part of total assets.

Similar Results. It turns out that empirically, the different measures produce qualitatively similar results. For example, Table 2, model (2) shows that inventory is 10.3% lower at FYE, when we measure inventory as absolute inventory scaled by COGS. The corresponding estimates are 5.9% for inventory measured as a absolute level and 3.4% for inventory measured as absolute level scaled by total assets. The former is about US$21.8 billion in inventory dollars and the latter, US$18.2 billion in asset dollars, using our U.S. dataset.

Implications of Scaling by Exogenous Variable. It might appear that scaling by COGS creates a measure that becomes mechanically tied to COGS. This turns out to be immaterial:

---

3 When we test COGS as a mediator, one of the two estimations (see the section on mediator tests) includes COGS as an additional covariate in specification (A), with inventory as dependent variable.
• **Conceptually,** all three measures are functions of several exogenous variables. To see this, recall that:

\[
\text{Absolute inventory}_t = \text{Absolute inventory}_{t-1} + \text{COGS}_t - \text{Production}_t - \text{Writeoffs}_t,
\]

Dividing the above by COGS does not make it any more or less tied to COGS.

• **Empirically,** the scaled measures are not highly correlated with COGS. The correlations with COGS (the signs are irrelevant) are:
  - Absolute inventory: 0.838
  - Scaling by COGS: -0.014
  - Scaling by assets: -0.006

  If anything, the absolute inventory measure has a higher correlation with COGS.

• **Econometrically,** that the measures are correlated with COGS is a requirement for identification. What is needed is that the dependent variable is neither orthogonal to nor collinear with the covariate of interest (see Wooldridge (2002)), which is true for all measures.

Finally, we reiterate that all the estimations with inventory do not say much about why inventory is higher or lower—*i.e.,* which exogenous variables (sales, production, or writeoffs) are driving inventory down. This is why we undertake explicit tests of various mediators to see the pathways with which FYE leads to low inventory.