# Conditional Methods in Event Studies and an Equilibrium Justification for Standard Event-Study Procedures

**N. R. Prabhala**
Yale University

***The literature on* conditional *event-study methods criticizes standard event-study procedures as being misspecified if events are voluntary and investors are rational. We argue, however, that standard procedures (1) lead to statistically valid inferences, under conditions described in this article; and (2) are often a superior means of inference, even when event-study data are generated* exactly *as per a class of rational expectations specifications introduced by the conditional methods literature. Our results provide an equilibrium justification for traditional event-study methods, and we suggest how these simple procedures may be combined with conditional methods to improve statistical power in event studies.***

I am grateful to Stephen Brown, my dissertation chairman, for stimulating my interest in the topic, his guidance, and numerous valuable suggestions. Thanks are also due to Franklin Allen (the executive editor), Yakov Amihud, Mitchell Berlin, Silverio Foresi, Robert Hansen, Kose John, L. Misra, Robert Whitelaw, and especially Bent Christensen, William Greene, Chester Spatt (the editor), and an anonymous referee whose extensive feedback has greatly improved the article. I have also benefitted from discussions with my colleagues at NYU, and from comments of seminar participants at various universities: British Columbia, Columbia, Cornell, Georgia Institute of Technology, Michigan, New York University, Purdue, Rutgers, Strathclyde, University of California, Los Angeles, Virginia Polytechnic Institute, and Yale. Any errors that remain are solely mine. Address correspondence to N. R. Prabhala, School of Management, Yale University, 135 Prospect Street, New Haven, CT 06520.

*The Review of Financial Studies* Spring 1997 Vol. 10, No. 1, pp. 1–38

Event studies are widely used to study the information content of corporate events. Such studies typically have two purposes: (1) to test for the existence of an "information effect" (i.e., the impact of an event on the announcing firm's value) and to estimate its magnitude, and (2) to identify factors that explain changes in firm value on the event date.

To test for the existence of an information effect, empirical finance has primarily employed the technique developed in Fama, Fisher, Jensen, and Roll (1969) (referred to as FFJR hereafter). FFJR suggest that if an event has an information effect, there should be a nonzero stock-price reaction on the event date. Thus, inference is based on the statistical significance of the average announcement effect[1] for a sample of firms announcing the event in question. The FFJR test is usually followed by a linear regression of announcement effects on a set of firm-specific factors to identify those factors that explain the cross-section of announcement effects. Most event studies in the applied literature have been based on the above methods.[2]

However, recent literature on *conditional* event-study methods [Acharya (1988, 1993), Eckbo, Maksimovic, and Williams (1990)] argues that the traditional methods are misspecified in a rational expectations context. Briefly, the argument is that corporate events are voluntary choices of firms and are typically initiated when firms come to possess information not fully known to markets. The *unexpected* portion of such information should determine the stock-price reaction to the event.

When events are modeled in this manner within simple equilibrium settings, the resulting specifications are typically nonlinear cross-sectional regressions[3] that bear little resemblance to the simple models conventionally used in event studies. Hence, it has been suggested that the conventional methods are misspecified and lead to unreliable inferences, implying that such methods should not be used in practice. More generally, this debate does raise the important issue that though the standard event-study procedures have been widely used in empirical work, little is understood about their consistency and power

[1] Announcement effect (or abnormal return) denotes the excess of the actual event-date stock return over the unconditional expected return for the stock. The latter is usually estimated via the market model, calibrated on pre-event data [see Brown and Warner (1985) for a more complete discussion].

[2] A partial list of applications includes studies of (1) equity and debt issues [Asquith and Mullins (1986), Eckbo (1986)], (2) timing of equity issues [Korajczyk, Lucas, and McDonald (1991)], (3) takeovers [Asquith, Bruner, and Mullins (1983)], (4) dividends [Bajaj and Vijh (1990)], and (5) stock repurchases [Vermaelen (1984)].

[3] The nonlinearity stems from the endogeneity of events. Endogeneity truncates the statistical distribution of announcement effects.

in rational expectations settings, such as those underlying conditional methods.

This article has three purposes. First, we present a simple exposition of conditional methods that focuses on their economic content. We show that all conditional models have essentially the same economic intuition, and derive all received models within a common framework that reflects this perspective. This synthesis reconciles different specifications proposed in the literature, clarifies their shared intuition, and suggests how one might choose between or extend such models in practice.

Our second point is that while traditional event-study techniques are indeed misspecified in the conditional methods context, they still lead to valid inferences, under certain statistical conditions described in this article. Specifically, even when event-study data are generated *exactly* as per conditional models of the sort introduced by Acharya (1988), (1) the FFJR procedure remains a well-specified test for detecting the *existence* of information effects; and (2) the conventional cross-sectional procedure yields parameter estimates *proportional* to the true conditional model parameters, under the conditions mentioned before. The proportionality factor has a simple interpretation in terms of the informational parameters of the event. These results provide, for the first time, an equilibrium justification for the procedures conventionally used to conduct event studies.

Finally, if both traditional and conditional methods lead to equivalent inferences, how does one choose between the two in practice? Working in the context of the conditional model proposed by Acharya (1988), we develop simulation evidence on this issue. Our evidence suggests that one's choice would depend mainly on whether one has, besides the usual event-study data, an additional sample of "nonevent" firms, that is, firms that were partially anticipated to announce but did not announce the event in question. If such nonevent data are available, conditional methods are powerful means of conducting event studies and may be implemented effectively using a simple "two-step" estimator. Absent nonevent data, conditional methods appear to offer little value relative to traditional procedures.

This article is organized along the above lines. Section 1 presents conditional methods for event studies. Section 2 presents and discusses the main analytic results, regarding the equivalence of inferences via conditional and traditional event-study methods. Section 3 motivates the question of choosing between the two approaches, and Section 4 outlines the structure of the simulations conducted to address this question. Simulation results are presented in Section 5, and Section 6 offers concluding comments.

## 1. On Conditional Methods

Section 1 develops conditional models for event studies. The main point made here is that all conditional models have essentially the same economic intuition: they relate announcement effects to the unexpected information revealed in events. While the notion of relating announcement effects to unexpected information is not new, we show here that it is the common theme that underlies all conditional models. We demonstrate that all received models may be derived in terms of this framework, and that the models differ only because they make different assumptions about the information structure underlying events.

The exposition proceeds as follows. Section 1.1 opens with a discussion of the intuition underlying conditional methods. Section 1.2 discusses alternative ways of modeling the information structure in events, and Sections 1.3 through 1.5 develop econometric models for announcement effects for each of these information structures.

### 1.1 Intuition underlying conditional methods

To begin, note the potential dichotomy between the *fact* of an event and the *information* it reveals. For instance, the event "takeover" is plausibly less surprising for a bidder with announced acquisition programs than for a bidder with no history of acquisitions. Similarly, the event "dividend increase" is less surprising for a firm with an unusually good spell of earnings than for a firm with flat or declining earnings. Thus, a given event may convey less information for some firms and more for others. Further, it should be the *unexpected information* revealed in events that causes the stock-price changes around event dates.[4]

This discussion suggests the following empirical procedure for carrying out event studies: (1) estimate for each firm the unexpected information that the event reveals; (2) compute the cross-sectional correlation between information and abnormal return and test for its significance. A nonzero correlation would indicate that abnormal return is systematically related to information revealed in the event (i.e., there exists an information effect). Conversely, zero correlation implies lack of an information effect. This intuition underlies every conditional specification analyzed in this article.

Central to the conditional paradigm is the notion of "information revealed in events." Next, we discuss how this might be modeled in the event-study context.

[4] Malatesta and Thompson (1985), Thompson (1985), and Chaplinsky and Hansen (1993) also recognize the role of partial anticipation of events and examine its implications for event studies based on FFJR-style procedures.

## 1.2 Specifying the information structure

As argued before, events reveal the information that their announcement is conditioned on. Suppose that this information consists of a variable $\tau_i$, which arrives at firm $i$ on an *information arrival* date. Information $\tau_i$ is subsequently revealed to markets, via the event, on an *event date*.

What do markets learn from the revelation of $\tau_i$? Clearly, this depends on what markets knew, prior to the actual event date, about the arrival of information $\tau_i$ at firm $i$. Here, we allow for three possibilities:

***Assumption 1.*** *Markets know, prior to the event, that the event-related information $\tau_i$ has* arrived *at firm i (but not its exact content).*

***Assumption 2.*** *Markets do not know, prior to the event, that information $\tau_i$ has arrived at firm i.*

***Assumption 3.*** *Markets assess a probability $p \in (0, 1)$ that information $\tau_i$ has arrived at firm i.*

Under Assumption 1, information arrival is common knowledge prior to the event; under Assumption 2, markets do not know about information arrival prior to the event-date. Finally, Assumption 3 is the encompassing case that permits markets to make probabilistic assessments about information arrival.[5] Each assumption leads to a different econometric specification for announcement effects, as we show below.

For expositional ease and because previous work has been based on Assumptions 1 and 2, we first develop the methodology under Assumptions 1 and 2, in Sections 1.3 and 1.4. We then present an encompassing specification, based on Assumption 3, in Section 1.5.

## 1.3 Model I: information arrival known prior to event

We begin by making Assumption 1: that markets know, prior to the event, that the event-related information $\tau_i$ has arrived at firm $i$. In general, this leads markets to form expectations about $\tau_i$. Suppose these expectations are given by

$$E_{-1}(\tau_i) = \underline{\theta}' \underline{x}_i = \sum_{j=1}^{n} \theta_j x_{ij} \tag{1}$$

[5] The following example illustrates the distinction between the three assumptions. Consider the event "takeover" and suppose takeovers occur if and only if the acquirer-bidder synergy ($\tau$) is positive. Assumption 1 implies that markets always know, prior to each takeover announcement, that the bidder had identified the target in question. The only uncertainty is whether $\tau$ is positive or not. In contrast, Assumption 2 implies that markets do not know, prior to each announcement, that the target had been identified. Under Assumption 3, markets assign probability $p \varepsilon (0, 1)$ that the target had been identified.

where $\underline{x}_i$ denotes a vector of firm-specific variables in the pre-event market information set, and $\underline{\theta}$ is a vector of parameters. Given Equation (1), firm $i$'s private information $\psi_i$ is given by

$$\psi_i = \tau_i - E_{-1}(\tau_i), \tag{2}$$

where $E_{-1}(\psi_i) = 0$, with no loss of generality.

In what follows next, we model an *event* as an announcement that each firm chooses to make (or not to make), depending on the nature of its information $\tau_i$. Our goal is to develop econometric models for the resultant announcement effect.

To fix matters, consider a situation in which each firm $i$ must choose between two mutually exclusive and collectively exhaustive alternatives on the event date: either the firm must announce the *event* ($E$) or the *nonevent* ($NE$). Suppose that the firm's decision depends on its information $\tau_i$, as follows:

$$E \Leftrightarrow \tau_i \geq 0 \Leftrightarrow \psi_i + \underline{\theta}'\underline{x}_i \geq 0 \tag{3}$$

$$NE \Leftrightarrow \tau_i < 0 \Leftrightarrow \psi_i + \underline{\theta}'\underline{x}_i < 0 \tag{4}$$

The choice model, Equations (3) and (4), reflects that the decision to announce an event is an endogenous choice of firms: here, event $E$ is announced if and only if conditioning information $\tau$ is "large enough." Otherwise, the "nonevent" $NE$ is announced.[6]

What do markets learn from firm $i$'s announcement? Given Equations (3) and (4), firm $i$'s choice between $E$ and $NE$ partially reveals its private information $\psi_i$, and thereby leads markets to form revised expectations about the value of $\psi_i$. The revised expectation, $E(\psi_i \mid C)$, $C \in \{E, NE\}$, constitutes the *unexpected information* on the event date.

If there is an information effect (i.e., the information revealed has a stock-price effect), we should find abnormal returns (say $\epsilon$) to be related to unexpected information. This relationship is linear under the following (jointly sufficient) assumptions.

***Assumption 4.*** *Risk Neutrality: Investors are risk-neutral towards the event risk.*

***Assumption 5.*** *Linearity: Conditioning information is a linear signal of expected stock return. That is, $E(r_i \mid \psi_i) = \pi\psi_i$, where $r_i$ stands for stock return, and $\psi_i$ for conditioning information.*

[6] Conditioning on $\tau$ being "large enough," — equivalently, a *sample selection bias* — characterizes all voluntary corporate events. For instance, takeovers plausibly occur if and only if the personal or corporate gains ($\tau$) from acquiring are positive; dividend increases are announced only when future earnings ($\tau$) are "large enough" to sustain higher dividends, and so on. The fact that $\tau$ is a function of elements $x$ in the pre-event information set captures the effect that some firms are more likely to announce the event than others.

Thus, if the event has an information effect, $\pi$ should be significant in the nonlinear cross-sectional specifications:

$$E(\epsilon_i \mid E) = \pi E(\psi_i \mid E) = \pi E(\psi_i \mid \underline{\theta}' \underline{x}_i + \psi_i \geq 0), \tag{5}$$

and

$$E(\epsilon \mid NE) = \pi E(\psi_i \mid NE) = \pi E(\psi_i \mid \underline{\theta}' \underline{x}_i + \psi_i < 0), \tag{6}$$

where $\epsilon_i$ is the *event-date* abnormal return for firm $i$. Intuitively, when firm $i$ makes an announcement $C$, it signals that the expected return, given its information, is $E(r_i \mid C) = \pi E(\psi_i \mid C)$ (via Assumption 5). Under risk neutrality, $E(r_i \mid C)$ also equals the expected *event-date* abnormal return $E(\epsilon_i \mid C)$.[7]

If private information $\psi_i$ is distributed normally, $N(0, \sigma^2)$, the above models may be rewritten as

$$E(\epsilon_i \mid E) = \pi\sigma \frac{n(\underline{\theta}' x_i)/\sigma}{N(\underline{\theta}' x_i/\sigma)} = \pi\sigma\lambda_E(\underline{\theta}' \underline{x}_i/\sigma), \tag{7}$$

and

$$E(\epsilon_i \mid NE) = \pi\sigma \frac{-n(\underline{\theta}' x_i/\sigma)}{1 - N(\underline{\theta}' x_i/\sigma)} = \pi\sigma\lambda_{NE}(\underline{\theta}' \underline{x}_i/\sigma), \tag{8}$$

where $n(\cdot)$ and $N(\cdot)$ denote the normal density and distribution, respectively, and $\lambda_C(\cdot)$ denotes the updated expectation of private information $\psi$, given the firm's choice $C\epsilon\{E, NE\}$.

Equation (7), our first "conditional" specification for announcement effects, was introduced by Acharya (1988). The model admits to two sets of hypothesis tests:

1. *Test for existence of information effect*: A test for significance of $\pi$ indicates whether announcement effects ($\epsilon$) are related to the information revealed in the event [$\lambda_C(\cdot)$], that is, whether there exists an information effect.

2. *Factors explaining announcement effects*: A test for significance of coefficients $\theta_j$ ($j = 1, 2, \ldots, k$) identifies from the set $x_j$ ($j = 1, 2, \ldots, k$) those factors that explain the cross-section of announcement effects.

[7] With risk aversion, we have two cases of potential interest. For firm-specific events of the sort analyzed here, announcement effects will be shifted upwards since (priced) uncertainty is resolved on the event date. In other words, $E_{-1}(\epsilon_i) > 0$. An interesting second case relates to events aggregate in character (such as federal interventions in fixed-income markets) and in which event risk is priced. Here $\lambda_C(\cdot)$, the "private information" is aggregate, and may be interpreted as a zero mean innovation in a priced APT factor. If the event risk is priced under a linear pricing operator, the risk-premium for the event could be estimated using cross-sectional and time-series data, much as in standard empirical APT studies [e.g., McElroy and Burmeister (1988)].

That the model is consistent with equilibrium follows from (1) risk neutrality towards event risk, and (2) the fact that the ex ante expected abnormal return is zero:

$$\begin{aligned}\Sigma_{k=E,NE}E(\epsilon_i \mid k) * Prob(k) \\ &= \pi\sigma\{\lambda_E(\cdot) * N(\underline{\theta}'\underline{x}_i/\sigma) + \lambda_{NE}(\cdot) * [1 - N(\underline{\theta}'\underline{x}_i/\sigma)]\} \\ &= 0. \end{aligned} \tag{9}$$

With this discussion in hand, it is fairly straightforward to develop binary event models under the alternate information structures, Assumptions 2 and 3. We present these models next and close Section 1 with a discussion on how one might choose between the three specifications in practice.

### 1.4 Model II: information arrival not known prior to event

Equation (7) was based on Assumption 1, under which markets knew ex ante about the arrival of information $\tau$. Suppose instead that the framework is Assumption 2: markets do not know ex ante about information arrival.[8] We now consider the conditional model for this situation.

Given Assumption 2, pre-event expectations about $\tau_i$ were not formed. Hence, $\tau_i$ itself (as opposed to $\psi_i$ in the previous section) is firm $i$'s private information. As before, the conditional expectation of private information (here $\tau_i$), given event $E$, constitutes the information revealed by $E$. This variable must be related to announcement effects, linearly so under Assumptions 1 and 2, if the event has an information effect. That is, $\pi$ should be significant in the model

$$\begin{aligned} E(\epsilon_i \mid E) &= \pi E(\tau_i \mid E) = \pi E(\tau_i \mid \tau_i \geq 0) \\ &= \pi\left[\underline{\theta}'\underline{x}_i + \lambda_E(\underline{\theta}'\underline{x}_i/\sigma)\right], \end{aligned} \tag{10}$$

where the last equality is obtained by using $\tau_i = \underline{\theta}'\underline{x}_i + \psi_i$. Equation (10) — hereafter, the EMW model — was, in essence, introduced by Eckbo, Maksimovic, and Williams (1990).

For some intuition, compare the EMW model, Equation (10), with the Acharya model, Equation (7). The EMW model has the extra term $\underline{\theta}'\underline{x}_i$ — the unconditional expectation of $\tau_i$. In the Acharya model, pre-event expectations of $\tau$ led to its unconditional expectation, $(\underline{\theta}'\underline{x}_i)$, being incorporated into the stock price *prior to the event*. Here, pre-event expectations were not formed (under Assumption 2); hence,

[8] This is the case, for instance, in takeover announcements involving bidders with no history of acquisitions or targets not previously in play. Here, markets plausibly do not know, prior to the actual takeover announcement, that the acquirer had identified the relevant target and that some announcement related to the acquisition was forthcoming.

the term $\underline{\theta}'\underline{x}_i$ appears in the expression for the abnormal return on the event date. Thus, contrary to a claim in Acharya (1993), we note that the EMW model is *not* nested within the Acharya model. The two models differ in their assumptions about the underlying information structure.

Both models are, in fact, limiting cases of a binary event model based upon Assumption 3. We derive this encompassing specification next and clarify the sense in which it nests the Acharya and EMW models.

### 1.5 Model III: information arrival partially known

Suppose now that the information structure is described by Assumption 3: markets assess a probability $p$ that information $\tau_i$ has arrived at firm $i$. Given Equation (1), the stock-price reaction in light of the assessed probability $p$ is given by

$$E_{-1}(\epsilon_i) = p\pi\underline{\theta}'\underline{x}_i.$$

Now, if event E does occur, it conveys two pieces of information. First, it confirms that information $\tau$ has arrived at firm $i$, that is, the probability of information arrival is raised from $p$ to 1. Second, it conveys via choice model Equations (3) and (4) that $\underline{\theta}'\underline{x}_i + \psi > 0$. Together, the two pieces of information lead to an announcement effect given by

$$E(\epsilon_i \mid E) = \pi\left[(1-p)\underline{\theta}'\underline{x}_i + \sigma\lambda_E\left(\frac{\underline{\theta}'\underline{x}_i}{\sigma}\right)\right]. \tag{11}$$

It is easily seen that Equation (11) nests the EMW and Acharya models as the special cases $p = 0$ and $p = 1$, respectively. The traditional event-study methods never obtain as the appropriate specifications, for any value of $p$.

How does one choose between these conditional specifications in practice? The preceding discussion demonstrates that this choice is essentially a matter of picking the informational assumption appropriate to one's context. Specifically, the EMW model is probably a good approximation of Equation (11) for nonrepetitive announcements whose timing is not well-identified ex ante. On the other hand, when markets are reasonably certain that some event-related announcement is forthcoming, the Acharya model is appropriate. For intermediate situations, Equation (11) is appropriate. Its practical value is not known and awaits empirical applications, as all received work is based on the EMW and Acharya models.

For the discussion that follows, we focus on the Acharya model, Equation (7), (i.e., the case when $p \approx 1$) since the EMW model, Equa-

tion (10), a "truncated regression" specification [see Greene (1993) or Maddala (1983)] has been treated fairly extensively in the econometric literature. By contrast, the properties of Equation (7) are not as well-understood: they are related to, but differ in interesting ways from, those of standard "selectivity" models. Hence, we focus on the Acharya model, Equation (7), next and through the rest of this article.

## 2. On Inferences Via Traditional Methods

The conditional specifications developed in Section 1 are quite different from traditional event-study procedures. How might one interpret inferences via traditional methods in light of this difference?

Working in the context of the Acharya model, Equation (7), we make two points. Specifically, we argue that even when event-study data are generated exactly as per Equation (7), (1) the FFJR procedure is well-specified as a test for *existence* of information effects (i.e., the hypothesis $\pi = 0$), whether or not any of the factors $\underline{x}$ explain announcement effects; and (2) the traditional cross-sectional procedure yields regression coefficients *proportional* to the true cross-sectional parameters $\underline{\theta}$, under conditions to be described shortly. Thus, while traditional techniques are indeed misspecified in the sense discussed before, the implications of such misspecification are probably not as serious as the previous literature [Acharya (1988, 1993), Eckbo, Maksimovic, and Williams (1990)] suggests. Conventional methods do allow one to conduct significance tests for both $\pi$ and cross-sectional parameters $\theta$, despite these parameters being potentially estimated inconsistently.

It is relatively straightforward to establish that the FFJR procedure may be viewed as a test of the hypothesis $\pi = 0$.[9] The cross-sectional results need some argument though, and we present these in what follows next.

### 2.1 The conventional cross-sectional procedure

The conventional cross-sectional procedure may be written as a test of significance of regression coefficients $(\beta_1, \ldots, \beta_k)$ in the linear

[9] Take expectations over conditioning factors $x$ in Equation (7). The unconditional (over $x$) announcement effect, given event $E$, is given by $E_x(\epsilon_i \mid E) = \pi\sigma E_x[\lambda_E(\cdot)]$, which is nonzero if and only if $\pi$ is nonzero [since $\lambda_E(\cdot) > 0$]. Hence, detecting a nonzero unconditional announcement effect, as in the FFJR procedure, is equivalent to an observation that $\pi$ is nonzero. Variants of the FFJR procedure, such as those introduced in Schipper and Thompson (1983), possess a similar interpretation.

regression

$$E(\epsilon \mid E) = \beta_0 + \underline{\beta}'\underline{x} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{12}$$

estimated for a sample of firms announcing event $E$ (firms announcing $NE$ are not considered by the conventional procedure), where we have dropped firm-specific subscript $i$ for notational ease. The linear model, Equation (12), is clearly misspecified, given the conditional model, Equation (7). What sort of inferences might it yield, if used anyway? That is, are regression coefficients $\beta$ related in some way to the true cross-sectional parameters $\underline{\theta}$ of Equation (7)?

Such a relationship does exist and, under fairly weak conditions, it takes a simple form: each linear regression coefficient $\beta_j$ is proportional to true coefficient $\theta_j$. Additionally, every $\beta_j$ is *biased towards zero*, relative to $\theta_j$.

The underlying intuition is illustrated by the following observation: the true slope $s_j$ of Equation (7) is *attenuated* relative to $\theta_j$.[10] Formally,

$$s_j = \frac{\partial E(\epsilon \mid E)}{\partial x_j} = -\theta_j \pi \delta(y), \tag{13}$$

where $y = \underline{\theta}'\underline{x}/\sigma$ and $\delta(y) = \lambda_E(y)[\lambda_E(y) + y]$. Since (1) $\mid \pi \mid < 1$ (it is a correlation), and (2) $0 < \delta(y) < 1$,[11] it follows immediately from Equation (13) that $\mid s_j \mid < \mid \theta_j \mid$. One might conjecture on this basis that each regression coefficient $\beta_j$ is biased towards zero, relative to $\theta_j$, with the opposite sign if $\pi > 0$.[12] Further, Equation (13) also suggests that downward bias should be greater when

1. $\mid \pi \mid$ *is small.* Here, announcement effects are less sensitive to conditioning information. Hence, regression coefficients $\beta_j$ should be smaller.

2. $\delta(y)$ *is small.* This happens when $y = \underline{\theta}'\underline{x}/\sigma$ is large [Goldberger (1983)], that is, for highly anticipated events. Here, little information is contained in firms' announcements of E or resultant abnormal returns. Once again, estimated regression coefficients $\beta_j$ should be smaller.

Precisely these results obtain when regressors $\underline{x}$ are multivariate normally distributed. Proposition 1 contains the formal statement.

---

[10] In a different setting, Lanen and Thompson (1988) also suggest that slope $s_j$ may be attenuated due to partial anticipation of the event.

[11] Interpreting $\pi$ as a correlation involves the normalization $var(\epsilon) = \sigma$. The bounds on $\delta(y)$ follow from two properties of the standard normal variable $z$ — (1) $E(z \mid z > -y) = \lambda_E(y)$ is decreasing in y, that is, $\lambda'_E(y) = -\delta(y) < 0$; and (2) $var(z \mid z > -y) = 1 - \delta(y) > 0$ [Greene (1993)].

[12] The linear regression itself does *not* necessarily estimate the slope of the nonlinear function [see, e.g., Stoker (1986), White (1980)].

***Proposition 1.*** *Suppose (1) event E occurs if and only if* $\theta_0 + \sum_{j=1}^{k} \theta_j x_j + \psi > 0$*; and (2) information* $\psi$ *and abnormal return* $\epsilon$ *are bivariate normal with correlation* $\pi$ *and marginal distributions N(0,1); and regressors* $(x_1, \ldots, x_k)$ *are multivariate normal, independent of* $\psi$*.*[13] *Then, coefficients* $(\beta_1, \ldots, \beta_k)$ *in the linear model, Equation (12), estimated for a sample of firms announcing E are given by*

$$\beta_j = -\theta_j \pi \frac{(1 - R^2)(1 - t)}{t + (1 - R^2)(1 - t)} = -\theta_j \pi \mu, \tag{14}$$

*where*

*1.* $t = var(\tau \mid E)/var(\tau)$, $\tau = \underline{\theta}' \underline{x} + \psi$.

*2.* $R^2$ *= coefficient of determination ("explained variance") in the population regression of* $\tau$ *on* $(1, x_1, \ldots, x_k)$.

*3.* $\mu = \frac{(1-R^2)(1-t)}{t+(1-R^2)(1-t)}$

See the Appendix for the proof.

To interpret the proportionality factor $\mu$, observe that (1) the term $(1-R^2)$ represents the variance of $\tau$ not explained by public information $\underline{x}$, that is, the *unexpected* component of information $\tau$; and (2) term $(1 - t)$ proxies the information revealed by event E. Therefore, the product of the two — and hence the term $\mu$ — represents *the unexpected component of information* $\tau$ *revealed by event E.* Another way of viewing this is to consider the fraction of information $\tau$ that is lost by restricting oneself to event E. Part of information $\tau$ is lost to (1) pre-event expectations and (2) the nonevent NE. The constant $\mu$ represents the fraction of information $\tau$ that remains in event E. Thus, $\mu$ is small when the event reveals little information; conversely, $\mu$ is large for highly surprising events. This intuition is formalized in Lemma 1.

***Lemma 1.*** *Let* $\mu$ *be as defined above. Then (1)* $0 < \mu < 1$*, and (2)* $\mu$ *is small when event E is, on average, highly anticipated.*

See the Appendix for the proof.

[13] In the choice model underlying Proposition 1, firms choose between *E* and *NE* based on latent information $\tau$. Condition (2) specifies how the latent information maps into stock-return information since it is the latter that causes observed announcement effects. Multivariate normality of $(\underline{x}, \tau)$ is stronger than what is needed for Proposition 1 to obtain. All we need is that the conditional expectation $E(\underline{x} \mid \tau)$ be linear in $\tau$. Multivariate normality is sufficient, though not necessary for this condition to hold. Finally, note that while firms have two choices (*E* or *NE*) in the event modeled here, Proposition 1 also applies to events in which each announcing firm has more than two choices — such as dividend announcements, wherein firms have three choices (increase, keep unchanged, or decrease dividends).

With these results in hand, one can readily establish useful comparative statics about regression coefficients $\beta_j$:

• *Downward bias in $\beta_j$'s.* This is an immediate consequence of $0 < \mu < 1$ (Lemma 1), $| \pi | < 1$, and Equation (14); together, these imply that $| \beta_j | \leq | \theta_j |$.

• *Opposite Sign.* Each $\beta_j$ is signed opposite to $\theta_j$, provided $\pi > 0$, as seen from Equation (14). To understand this result, note that $\theta_j$ reflects the marginal impact of an increase in regressor $x_j$ on the *probability* of event $E$, while $\beta_j$ reflects the marginal impact on the *announcement effect* associated with $E$. Since an increase in the probability of event $E$ decreases the expected announcement effect upon announcing $E$ if $\pi > 0$, $\theta_j$ and $\beta_j$ have the opposite sign when $\pi > 0$.

• *More attenuation when $| \pi |$ is small.* This follows directly from Equation (14).

• *More attenuation when events are highly anticipated.* For highly anticipated events, $\mu$ is small, from part 2 of Lemma 1. From Proposition 1, this implies that $| \beta_j |$ is small.

Summarizing, Proposition 1 has the interesting implication that the traditional cross-sectional procedure may be used for cross-sectional inferences in event studies. Specifically, a statistical test for significance of regression coefficient $\beta_j$, $(j = 1, \ldots, k)$, is equivalent to a test for significance of the corresponding cross-sectional parameter $\theta_j$ of the conditional model.

However, for practical purposes, two questions remain. One, while Proposition 1 provides an interpretation of the linear regression coefficients, are the usual OLS standard errors appropriate for use in significance tests? Second, how robust is Proposition 1 to the assumption that regressors $\underline{x}$ are multivariate normal? Simulation evidence needs to be developed on these issues.

## 3. Issues in Choosing Event-Study Methodology

Section 2 suggests that under certain conditions, both conditional and traditional methods are valid means of inference. How might one choose between the two approaches in practice? We address this issue in the context of cross-sectional inferences, as conditional methods are likely to be useful only when cross-sectional hypotheses are being tested.[14]

One's choice between the two approaches would depend primarily on the performance of each method (i.e., the likelihood of making

[14] Simulation evidence on the FFJR procedure (reported in earlier versions of this article) attest to this point. These results are available upon request.

correct inferences about the sign and significance of cross-sectional parameters $\underline{\theta}$), given typical event-study samples. We argue that the relative performance of the two approaches must be considered in two distinct cases:

1. *All assumptions satisfied*: To begin, suppose that event-study data are generated exactly as per Equation (7) and that the assumptions underlying Proposition 1 are satisfied, so that both conditional and traditional methods are equally valid means of inference. Even so, is there any reason why one method might be preferred over the other? Section 3.1 considers this question. We argue that even in this "ideal" case, one's choice should depend on what data one has — specifically, whether one has a sample of *nonevent* firms or not.

2. *Some assumptions not satisfied*: To motivate the second case, observe that the conditional model, Equation (7), places a fairly tight statistical structure on announcement effects. Not all of its econometric assumptions may be satisfied in practice. Hence, we also consider separately the question of methodological choice when some of the assumptions underlying the conditional model are not satisfied. This issue is addressed in Section 3.2.

One's choice would also depend, to some extent, on the computational burden involved in estimation, which is likely to be greater for the conditional model. Section 3.3 discusses these computational issues and presents a brief summary that motivates the empirical work to follow.

### 3.1 Choice when all assumptions are satisfied

With the identifying normalization $w = \pi\sigma$ and $\underline{\theta} = \frac{\theta}{\sigma}$ [equivalently, $var(\psi) = \sigma^2 = 1$], the conditional model, Equation (7), consists of a "probit" model governing firms' choices between event $E$ and nonevent $NE$:

$$C = \begin{cases} E & if \quad \underline{\theta}'\underline{x}_i + \psi > 0 \\ NE & if \quad \underline{\theta}'\underline{x}_i + \psi < 0 \end{cases} \tag{15}$$

coupled with a heteroskedastic cross-sectional regression for announcement effects,

$$\epsilon_i = w\lambda_C(\underline{\theta}'\underline{x}_i) + e_i, \tag{16}$$

and

$$var(e_i \mid C) = \sigma_\epsilon^2 - w^2\lambda_C(\underline{\theta}'\underline{x}_i)[\lambda_C(\underline{\theta}'\underline{x}_i) + \underline{\theta}'\underline{x}_i],$$

where $C \in$ {E, NE } is firm $i$'s choice, $\sigma_\epsilon^2 = var(\epsilon_i)$, and $e_i$ is an error term.

In this section, we argue that even if event-study data are generated exactly as above, and the results of Section 2 hold, conditional meth-

ods would be, a priori, one's preferred means of inference, under some conditions. Specifically, if one has, besides a sample of firms announcing event $E$, additional data on nonevent firms (those announcing $NE$), conditional methods are likely to be preferred over traditional methods.

To see why, observe that under the traditional approach, event studies are conducted using only data on firms that announced event $E$. However, Equations (15) and (16) point to the existence of a second category of firms: *nonevent* firms, that is, firms that were partially anticipated to announce but chose not to announce the event in question. Such firms are not used under the traditional procedure. On the other hand, nonevent information may be exploited in the conditional framework by estimating the conditional model with both event and nonevent data. Thus, when nonevent data are available, conditional methods should be more powerful relative to traditional methods.

In most practical situations though, nonevent data are not available. Nonevent data include (1) a set of firms that were anticipated to announce but chose not to announce the event in question; (2) the time when markets learn of the non-announcement; and (3) cross-sectional and announcement effect data on this date. Usually, such information cannot be obtained [Lanen and Thompson (1988) make a similar point], and one must work with only the firms that have announced event $E$. Here, both conditional and traditional procedures use the same data in estimation, and there is little to choose between the two procedures from an informational viewpoint, if Proposition 1 holds. However, Proposition 1 does not generalize for arbitrary distributions of regressors $\underline{x}$. Thus, if $\underline{x}$ does not satisfy the distributional assumptions of Proposition 1, conditional methods may be preferred since traditional procedures might give misleading inferences in this instance. We develop some Monte-Carlo simulation evidence on the seriousness of this issue.

Thus, in the benchmark case, one's choice of methodology depends on what data one has. If nonevent data are available, conditional methods are likely to be preferred; absent nonevent data, one's choice is less clear, and the issue warrants empirical investigation.

### 3.2 Choice when some assumptions are not satisfied

In the benchmark case, we assumed that event-study data are generated exactly as per the conditional model. However, some of the model's statistical assumptions may not be satisfied in practice. In this section, we discuss why such deviations might arise and examine the implications for one's choice of event-study methodology.

The first issue, raised in the econometric literature on "selectivity" models, relates to the sensitivity of Equations (15) and (16) to

nonnormality of private information $\psi$. Recollect that in developing Equations (16), $\psi$ was assumed to be normally distributed. If $\psi$ is in fact nonnormal, estimates based on the normality assumption are inconsistent.

There is little consensus on the seriousness of the nonnormality issue or on the value of alternate models robust to this distributional assumption. Received evidence has been essentially mixed on both scores [Arabmazar and Schmidt (1982); Goldberger (1983); Lee (1982, 1983); Newey, Powell, and Walker (1990)]. However, all reported evidence pertains to selectivity models that are quite different in focus[15] from that of event-study specification, Equation (16). Its relevance to the event-study situation is not clear, motivating the need to develop evidence more specific to our context.

A second concern relates to data problems that are endemic to the event-study situation [Brown and Warner (1985)]. Two issues are of particular relevance here:

1. *Noise in announcement effects*: Announcement effects are inevitably measured with noise, due to the need to estimate the calibrating market model. Additionally, noise may be induced by uncertainty about the true event date, non synchronous trading, or bid-ask spreads in prices, as well as any changes in sources of valuation not observable to the econometrician. Thus, it is difficult to isolate the portion of stock returns exclusively attributable to announcement of the event.

2. *Cross-sectional correlation*: Models such as Equation (16) are almost always estimated assuming that latent information $\psi$ is cross-sectionally independent. However, in the event-study context, $\psi$ is often cross-sectionally correlated, due to common macroeconomic or industry influences on firms' decisions to announce events, or due to clustering of event dates in calendar time.

While problems such as these do exist in practice, little is known about how they impact one's inferences via either event-study approach. In particular, are inferences via the conditional model affected more than those via the simpler traditional procedure? We develop some evidence along these lines in the work to follow.

[15] The prototype selectivity model is the regression $E(\epsilon_i \mid E) = \underline{\beta}'z + \pi\lambda_E(\underline{\theta}'x)$. This class of models focuses almost exclusively on consistent estimation of parameters $\underline{\beta}$ of the unconditional mean function $\underline{\beta}'z$. In contrast, our interest primarily centers on estimates of *selectivity* parameters $\underline{\theta}$, which is a subject of lesser interest in the selectivity literature (note that in the EMW model, Equation (10), $\beta = \theta$ since $p$ is essentially zero). However, parameter $\underline{\beta}$ may also be of interest sometimes in the event-study context. For example, in the encompassing model, Equation (11), where $\underline{\beta} = \underline{\theta}(1-p)$, estimates of $\underline{\beta}$ might be of interest since they allow one to estimate parameter $p$ and conduct related hypothesis tests. No such tests have been reported yet in the literature.

### 3.3 Computational considerations

The practical value of conditional specifications such as Equations (15) and (16) depends, to some extent, on the computational burden involved in estimation. Estimation turns out to be quite straightforward, provided one has both event and nonevent data. However, estimation is more demanding when one has only event data, as we discuss below.

To begin, consider the issue of estimating the conditional model when one has both event and nonevent data. In this setting, two natural estimation techniques are (1) maximum likelihood (ML) and (2) nonlinear least squares (NLS). However, consistent estimation may also be achieved through a simple two-step procedure [Heckman (1979)]. To motivate the two-step procedure, observe that in Equation (16), if one had consistent estimates of parameters $\underline{\theta}$ [and hence of $\lambda_C(\cdot)$], regression of $\epsilon$ on $\lambda_C(\cdot)$ would lead to consistent estimates of $w$. This suggests a computationally simple procedure for estimating the conditional model:

1. Estimate the "probit" choice model, Equation (15), to obtain consistent estimates of $\underline{\theta}$, say $\underline{\hat{\theta}}$.

2. Use $\underline{\hat{\theta}}$ to compute $\lambda_C(\cdot)$ for each observation and obtain $w$ by OLS estimation of Equation (16), adjusting standard errors appropriately [see Greene (1981) and Heckman (1979)].

The two-step procedure offers two other advantages, relative to ML and NLS estimators, in the particular context of event studies. First, cross-sectional inferences under the two-step procedure do not require announcement-effect data. Hence, the abnormal return measurement problems discussed in Section 3.2 are no longer an issue in cross-sectional inferences. A more interesting consequence is that cross-sectional estimates via the two-step method are consistent not only for the Acharya model, Equation (16), but also for the EMW model, Equation (10), and the encompassing specification, Equation (11) as well, unlike ML and NLS estimates, which are model-specific.

Robustness, however, comes at a cost: the two-step procedure is not efficient precisely for the same reason it is robust, that is, the first step does not use announcement effect data. Announcement effects represent markets' assessments about the information conditioning the event and should add efficiency in cross-sectional parameter estimation. Nevertheless, our evidence indicates that such efficiency gains are typically small.[16] Hence, in the empirical work that follows, we

[16] The evidence was found in simulations with ML and NLS estimators. With large samples, or when $w$ is known, additional information in announcement effects should clearly make cross-sectional parameter estimates more precise relative to those obtained via the two-step method. In finite

use the two-step estimator whenever both event and nonevent data are available.

Suppose, on the other hand, that one has only event data. In this setting, the two-step procedure is not available, and the conditional model must be estimated via ML or NLS. Estimation now involves optimization of a nonstandard maximand, and, as we describe later, successful convergence requires some experimentation with numerical procedures and parameters governing the optimization process. From a computational perspective, the conditional model is less attractive since estimation now involves greater effort. Whether such effort leads to more powerful inferences, relative to those via the much simpler traditional procedure, remains to be seen, and we explore this issue in the empirical work to follow.

To summarize, Section 3 has raised issues concerning one's choice of event-study methodology, and we address these issues via simulations. Our discussion also suggests how such simulation experiments should be organized. Specifically, simulations ought to be carried out for the case when all assumptions underlying the conditional model are met and, separately, when they are not met. In each instance, two sets of data should be used: one comprising event data only, and another comprising both event and nonevent data.

Section 4 details the broad structure of our simulation experiments, and Section 5 presents the simulation results.

## 4. Experiment Design

In broad terms, our experiment consists of three steps: (1) simulation of event-study data as per the conditional model; (2) introduction, where relevant, of problems such as nonnormality into the data; and (3) estimation of model parameters by alternative event-study techniques. Some remarks on our design choices are in order before moving to the details.

Two of our choices are different from ones made in related work by Acharya (1993). Acharya simulates regressors $\underline{x}$ from the normal distribution. We sample $\underline{x}$ from the uniform distribution. As analytic results (Section 2.1) are available for normally distributed regressors, a nonnormal alternative was desired, leading to our choice. Second, our sample sizes (100 or 250) are much smaller than the size (800) used in Acharya (1993). Our choice is governed by two considerations: (1) these represent sample sizes typically used in event studies

samples and when $w$ must be estimated, the gains should be smaller. Indeed, under ML and NLS, the primary improvement relative to the two-step method was found to lie in the precision of estimates of $w$.

in corporate finance; (2) given the results of Section 2, both conditional and traditional procedures yield equivalent inferences in large samples. Thus, large sample comparisons of the two methods, the focus of previous work, are not relevant to the question of choosing between them in practice.

Our regressor and parameter choices give an adjusted $R^2$ of about 10% if the linear probability model is used to estimate the underlying choice model. The experiment design is similar to ones used previously in econometric literature [Nelson (1984), Paarsch (1984), Wales and Woodland (1980)].

We now describe in some detail the methodology used to (1) simulate the event and abnormal return data, and (2) construct the test statistics.

### 4.1 Broad design

1. *Event*: For $i = 1, 2, \ldots, n$ (the sample size), the event was simulated to occur as

$$\text{E} \quad \Leftrightarrow \quad \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \psi_i \geq 0,$$

and

$$\text{NE} \quad \Leftrightarrow \quad \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \psi_i < 0,$$

where

- $\theta_0$ = 1, $\theta_1 = -1$, $\theta_2 = 0.01$
- Regressor $x_{2i}$, $i = 1, 2, \ldots, n$, was drawn independently and identically distributed (i.i.d.) from a uniform distribution U(0,100) once for each set of 400 simulations.
- Regressor $x_{1i}$, $i = 1, 2, \ldots, n$, was also drawn once for every set of 400 simulations, i.i.d. from a uniform distribution with support of unit length. The support location was varied, so as to get average event probabilities of 25% or 50%.[17]
- $\psi_i$ was drawn i.i.d. from the normal distribution N(0,1), except in experiments addressing nonnormality issues. Here, $\psi_i$ was drawn from one of four nonnormal distributions: Laplace, Logistic, Chi-square or Student's $t$. We normalized the draw to unit variance/zero mean by (1) subtracting the mean, and (2) dividing by the standard deviation of the relevant distribution.

2. *Abnormal return*: Data for abnormal return were simulated as $\epsilon_i = w\psi_i + v_i$, where $v_i$ is i.i.d. noise with (a) $E(v_i) = 0$, and (b) $var(v_i) = 1 - w^2$ (this normalization sets the unconditional variance

[17] When the average event probability was 50%, $x_1$ had support (1, 2), except when $\psi$ was $\chi^2$ distributed, when it had support (1.2, 2.2). For 25% average event probability, the support varied from (1.5, 2.5) to (1.75, 2.75), depending on the distribution.

of $\epsilon_i$ to unity) and the same distribution as $\psi_i$.[18] Each set of 400 simulations was repeated for three values of $w$ — 0.30, 0.50, and 0.75 — going from less-informative (in terms of information contained in abnormal return) to more-informative events.

3. *Sample Size*: Each set of 400 simulations was repeated for two sample sizes — 100 and 250.

4. *Number of replications*: Statistics are based on 400 replications of each simulation.

### 4.2 Simulation methodology

A typical set of 400 simulations proceeds as follows:

1. Fix the desired average event probability, sample size, $w$, and distribution for $\psi_i$.

2. Draw a sample of regressors $x_{2i}$ and $x_{2i}$ to conform to the target average event probability fixed in step (1) above. This sample of $\underline{x}$'s is fixed for all 400 repetitions. We then repeat steps (3) through (5) 400 times.

3. Simulate $\psi_i$ from the target distribution. Normalize $\psi_i$ to zero mean/unit variance.

4. Simulate abnormal return data for each firm $i$, as in Section 4.1.

5. Estimate parameters $\underline{\theta}$, $w$, and $\underline{\beta}$ as appropriate and compute associated $t$-statistics.

Each set of 400 simulations is repeated for (1) three values of $w$ — 0.30, 0.50, and 0.75; (2) two average event probabilities — 25% and 50%; and (3) two sample sizes — 100 and 250.

### 4.3 Reported statistics

Every set of 400 simulations yields 400 point estimates of $w$ and $\underline{\theta}$, together with associated $t$-statistics. From these, we compute the following statistics.

- *Mean*: This denotes the average of the 400 point estimates of the relevant parameter.
- *Std. error*: This is the sample standard deviation of the 400 parameter estimates, from their simulated distribution. This should be equal to the standard error implied in the asymptotic $t$-statistics in individual simulations, provided the relevant estimator attains its asymptotic distribution.
- *Mean t-stat*: Every simulation produces a $t$-statistic for each model parameter. Mean $t$-stat denotes the average of the 400 $t$-statistics.

[18] There are no natural form representations for the implied bivariate distributions of $(\psi, \epsilon)$. Known bivariate forms corresponding to the univariate distributions used here (e.g., bivariate logistic, bivariate Student's $t$) have nonlinear conditional first moments, which is inconsistent with our Assumption 5.

**Table 1**
**Performance of conditional model: base case**

| Model parameter | Truth | Sample size = 100 Mean | Std. error | Mean $t$-stat | Sample size = 250 Mean | Std. error | Mean $t$-stat |
|---|---|---|---|---|---|---|---|
| Average event probability = 25% | | | | | | | |
| $w$ | 0.30 | 0.31 | 0.13 | 2.30 | 0.30 | 0.09 | 4.27 |
| $w$ | 0.50 | 0.50 | 0.13 | 3.81 | 0.50 | 0.08 | 6.01 |
| $w$ | 0.75 | 0.74 | 0.11 | 6.38 | 0.75 | 0.08 | 10.13 |
| $\theta_0$ | 1.00 | 1.04 | 0.51 | 2.05 | 0.99 | 0.31 | 3.27 |
| $\theta_1$ | −1.00 | −1.04 | 0.52 | −1.97 | −0.98 | 0.32 | −3.12 |
| $100\theta_2$ | 1.00 | 1.02 | 0.52 | 1.99 | 0.96 | 0.31 | 3.08 |
| Average event probability = 50% | | | | | | | |
| $w$ | 0.30 | 0.30 | 0.12 | 2.37 | 0.32 | 0.08 | 4.45 |
| $w$ | 0.50 | 0.50 | 0.12 | 4.12 | 0.50 | 0.08 | 6.64 |
| $w$ | 0.75 | 0.76 | 0.11 | 6.83 | 0.74 | 0.07 | 10.74 |
| $\theta_0$ | 1.00 | 0.99 | 0.75 | 1.36 | 0.97 | 0.46 | 2.08 |
| $\theta_1$ | −1.00 | −1.00 | 0.48 | −2.15 | −0.98 | 0.30 | −3.36 |
| $100\theta_2$ | 1.00 | 1.00 | 0.47 | 2.20 | 1.02 | 0.29 | 3.46 |

Table 1 presents statistics relating to the conditional model's performance when all underlying assumptions are satisfied.

We simulate event E to occur if and only if $\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \psi_i > 0$, where $\theta_0 = 1$, $\theta_1 = -1$, and $\theta_2 = 0.01$. Information $\psi_i$, $i = 1, 2, \ldots, n$ ($n$ is the sample size), is i.i.d. normal, N(0,1). Expected event-date abnormal return, conditional on $\psi_i$, is simulated as $E(\epsilon_i \mid \psi_i) = w\psi_i$.

In each simulation, we sample $\psi_i$ and $\epsilon_i$ as above. We then apply the two-step method to estimate parameters $w$ and $\theta$. This process is repeated 400 times; each set of 400 simulations is carried out for (1) 3 values of $w$ — (0.30, 0.50, and 0.75), (2) two sample sizes — (100 and 250), and (3) two average event probabilities (25% and 50%). From each set of 400 repetitions, we compute and report the following statistics: (1) *Mean*: mean parameter point estimate, averaged over the 400 repetitions; (2) *Std. error*: standard error of parameter estimate, computed as the sample standard deviation of the 400 point estimates; (3) *Mean t-stat*: Each repetition generates a $t$-statistic for the relevant parameter estimate. Mean $t$-stat refers to the mean of the 400 $t$-statistics. As estimates of $\theta$ are obtained independent of $w$, we report only one set of statistics for $\theta$ that applies to all three values of $w$.

## 5. Simulation Results

We present the simulation results in three parts. Section 5.1 discusses the conditional model's performance when both event and nonevent data are available. Here, we examine the model's sensitivity to nonnormality and various event-study data problems mentioned before. Section 5.2 evaluates the traditional cross-sectional procedure, and finally, Section 5.3 discusses the conditional model's performance when only event data are available.

### 5.1 Conditional model: event and nonevent data

**5.1.1 Base case.** Our first set of results concerns the conditional model's performance when all underlying assumptions are satisfied. These results are presented in Table 1. As first-step estimates of $\underline{\theta}$ are

independent of the true $w$, we report only one set of statistics for $\underline{\theta}$, which applies to all four values of $w$.

**Discussion of results.** It is useful to analyze our results in two parts, one pertaining to the Probit estimates of $\underline{\theta}$, and the other pertaining to second-step estimates of $w$.

Probit estimates of $\theta_1$ and $\theta_2$ (our main objects of interest) are close to truth on average, as expected. The average $t$-statistic for $\theta_1$ (or $\theta_2$) is about 2.0 for a sample size of 100 and about 3.0 for a sample size of 250. This corresponds to power of rejecting the hypothesis $\theta_1 = 0$ of about 50% and 86%, respectively (at 5% size).[19]

Next, consider the point estimates of $w$. Average estimates of $w$ are close to truth. As true $w$ increases, associated $t$-statistics get larger. For samples of size 100, the average $t$-statistic increases from about 1.4 for $w = 0.20$ to about 3.8 for $w = 0.50$, corresponding to power of 29% and 96%, respectively. Larger $w$ are virtually certain to be picked up.

Finally, observe that standard errors used within simulations to construct the $t$-statistics are close to their true values. For instance, for 25% event probability and sample size of 100, the standard error implied in the probit $t$-statistic for $\theta_1$ is 0.528 ($\frac{1.04}{1.97} = \frac{\theta_1}{t_{\theta_1}}$). This is almost exactly equal to the true standard error, 0.52 (see column Std. error), obtained from the simulated distribution of parameter estimates $\theta_1$. Similar results are seen to hold for every other parameter, highlighting that the estimates do conform to their asymptotic distribution.

In the remainder of Section 5.1, we artificially introduce statistical and data problems into the data and report the conditional model's performance under these conditions.

**5.1.2 Nonnormality.** What sort of inferences does the procedure based on normality yield, when $\psi_i$ is actually nonnormal? To address this issue, we simulate data as in Section 5.2, from the assumed nonnormal distribution, and estimate $\underline{\theta}$ and $w$ using the two-step procedure based on normality. Four nonnormal distributions were considered, based on previous literature [Arabmazar and Schmidt (1982), Goldberger (1983), and Lee (1982, 1983)]: (1) logistic, (2) Student's $t$ (5 degrees of freedom), (3) Chi-square (5 degrees of freedom), and (4) Laplace. The conditional moments and density functions of the four distributions are detailed in Goldberger (1983) and Lee (1982). The degrees of freedom in distributions (2) and (3) were kept small to maximize their departure from normality.

[19] Here, we define power as the probability of not committing a Type II error, that is, of rejecting the null hypothesis $H_0 : \theta_k = 0$. In practice, the test's power is obtained as the fraction of the 400 $t$-statistics for $\theta_1$ exceeding $t_{critical} = 1.96$, the cutoff for a 5% significance level.

**Table 2**
**Normality-based estimator when information is laplace distributed**

| Model parameter | Truth | Sample size = 100 | | | Sample size = 250 | | |
|---|---|---|---|---|---|---|---|
| | | Mean | Std. error | Mean *t*-stat | Mean | Std. error | Mean *t*-stat |
| Average event probability = 25% | | | | | | | |
| $w$ | 0.30 | 0.28 | 0.14 | 2.13 | 0.28 | 0.09 | 3.32 |
| $w$ | 0.50 | 0.47 | 0.14 | 3.40 | 0.47 | 0.07 | 5.55 |
| $w$ | 0.75 | 0.70 | 0.12 | 6.06 | 0.70 | 0.08 | 8.86 |
| $\theta_0$ | 1.00 | 1.26 | 0.59 | 2.22 | 1.24 | 0.36 | 3.46 |
| $\theta_1$ | −1.00 | −1.25 | 0.53 | −2.41 | −1.23 | 0.31 | −3.85 |
| $100\theta_2$ | 1.00 | 1.25 | 0.53 | 2.43 | 1.21 | 0.31 | 3.85 |
| Average event probability = 50% | | | | | | | |
| $w$ | 0.30 | 0.27 | 0.13 | 2.10 | 0.28 | 0.08 | 3.50 |
| $w$ | 0.50 | 0.46 | 0.12 | 3.73 | 0.45 | 0.07 | 5.77 |
| $w$ | 0.75 | 0.69 | 0.12 | 6.40 | 0.69 | 0.07 | 9.76 |
| $\theta_0$ | 1.00 | 1.41 | 0.80 | 1.80 | 1.36 | 0.46 | 2.90 |
| $\theta_1$ | −1.00 | −1.41 | 0.50 | −2.80 | −1.36 | 0.30 | 4.54 |
| $100\theta_2$ | 1.00 | 1.40 | 0.49 | 2.82 | 1.35 | 0.31 | 4.44 |

Table 2 presents statistics relating to the conditional model's performance, when conditioning information $\psi_i$ is incorrectly assumed to be standard normal.

We simulate event E to occur if and only if $\theta_0+\theta_1 x_{1i}+\theta_2 x_{2i}+\psi_i > 0$, where $\theta_0 = 1, \theta_1 = -1$, and $\theta_2 = 0.01$. Information $\psi_i$, $i = 1, 2, \ldots, n$ ($n$ is the sample size), is i.i.d., sampled from the Laplace distribution (normalized to unit variance). Expected event-date abnormal return, conditional on $\psi_i$, is simulated as $E(\epsilon_i/\psi_i) = w\psi_i$.

In each simulation, we sample $\psi_i$ and $\epsilon_i$ as above. We then apply the two-step method based on normality to estimate parameters $w$ and $\underline{\theta}$. This process is repeated 400 times; each set of 400 repetitions is carried out for (1) 3 values of $w$ (0.30, 0.50, and 0.75), (2) two sample sizes (100 and 250), and (3) two average event probabilities (25%, 50%). From each set of 400 simulations, we compute and report the following statistics: (1) *Mean*: mean parameter point estimate, averaged over the 400 repetitions; (2) *Std. error*: standard error of parameter estimate, computed as the sample standard deviation of the 400 point estimates; and (3) *Mean t-stat*: Each repetition generates a *t*-statistic for the relevant parameter estimate. Mean *t*-stat refers to the mean of this *t*-statistic, averaged over the 400 repetitions. As estimates of $\underline{\theta}$ are obtained independent of $w$, we report only one set of statistics for $\underline{\theta}$ for each average event probability/sample size. This applies to all three values of $w$.

**Discussion of results.** Table 2 presents results for the Laplace distribution. Qualitatively similar results obtain for other distributions and are not reported here.

Not surprisingly, point estimates of $\theta_1$ and $\theta_2$ in Table 2 are quite different from their true values. The difference reflects that the normality-based estimator is inconsistent when the true distribution of $\psi$ is nonnormal. However, note that the *sign* and order of these estimates are similar to those in Table 1, where data actually come from the normal distribution; the associated *t*-statistics are also of a similar magnitude. Thus, for the nonnormal distributions considered here, there seems to be little impact on one's inferences, reflecting the apparent robustness of inferences via the probit step.

Next, consider estimates of $w$. Point estimates of $w$ are close to truth everywhere but are slightly attenuated. For example, when true

$w$ = 0.75, average estimates range from 0.69 to 0.70. Nonnormality introduces a new source of noise into the second-step regressor $\lambda_k(\cdot)$, that of approximating it by its counterpart based on normality. This introduces "measurement error" into $\lambda_k(\cdot)$, which attenuates estimated $w$. In all instances though, the amount of attenuation is small, and the $t$-statistics for $w$ are close to their corresponding values in Table 1.

As before, standard errors implied in $t$-statistics are roughly equal to their values from the empirical distribution of simulated parameter estimates, despite the misspecification engendered by nonnormality. Thus, nonnormality does lead to inconsistent parameter estimates but does not appear to impact one's inferences about the significance of model parameters.

**5.1.3 Noise in announcement effects.** As discussed in Section 3.2, announcement effects are always measured with noise in practice. In this section, we examine how such noise affects the conditional model's performance.

The experiment here involves simulating abnormal return data as in Section 4.2, and adding noise $n_i$, drawn i.i.d. N$(0, n^2)$, to the true abnormal returns. The two-step procedure is then applied to the data as usual, to estimate $\underline{\theta}$ and $w$. Simulations were carried out for four values of $n^2$ — 0.20, 0.40, 0.65, and 1.00 — corresponding to noise levels of 20%, 40%, 65% and 100%, respectively (since var$[\epsilon] = 1$).

**Qualitative discussion of results.** For brevity, we do not report the complete simulation results but only provide a qualitative discussion instead.

Noise in announcement effects does not affect estimates of cross-sectional parameters $\underline{\theta}$ under the two-step procedure since the procedure does not use announcement-effect data in estimation of $\underline{\theta}$.[20] However, noise in estimated announcement effects does lead to larger error terms in the second-step regression. Hence, second-step estimates of $w$, while consistent, are now less precise and the average $t$-statistics for $w$ are smaller as a result.

For moderate amounts of noise, there is little impact on one's inferences. For instance, when $w$ = 0.30, we found that an increase in

[20] We also examined the effect of noise on other estimators of the conditional model that use announcement effects in estimating cross-sectional parameters. Simulation evidence (based on the maximum likelihood procedure for the case where there are both event and nonevent data) shows that noise in announcement effects makes estimates of $w$ less precise. However, there was little effect on estimates of $\underline{\theta}$, whose distribution remained virtually unchanged for all levels of noise from 20% to 100%. Thus, in the context of Equation (7), most information relevant to cross-sectional estimation seems to be contained in whether firms announced an event or not, rather than the associated announcement effects.

noise from 0% to 40% reduces power (i.e., fraction of $t$-statistics exceeding 1.96) only from 64% to 52%. Thus, the two-step procedure does stand up to moderate amounts of noise at levels typical of event studies that use windows of a few days to measure announcement effects.

**5.1.4 Cross-sectional correlation in information.** Thus far, we have assumed information $\psi$ to be i.i.d. across firms. What effect does cross-sectional correlation have on the conditional model? Cross-sectional correlation does not affect consistency of probit estimates of $\underline{\theta}$, and by extension, second-step estimates of $w$. However, the $t$-statistics for tests of significance could be inappropriate. We develop some evidence on the direction of the potential bias.

Simulation methodology follows that of Section 4.2, with one change — information $\psi_i$ is sampled differently to artificially introduce cross-sectional correlation into the data. This is accomplished by simulating $\psi_i$, $i = 1, 2, \ldots, n$ ($n$ is the sample size), as

$$\psi_i = \alpha c + u_i\sqrt{1-\alpha^2},$$

where (1) $c$ is sampled from the normal distribution N(0,1) once for each repetition; (2) $u_i$, $i = 1, 2, \ldots, n$, is sampled i.i.d. N(0,1). This procedure effectively produces a sample with (a) $E(\psi_i)=0$, and (b) $\text{var}(\psi_i)=1$ and correlation $E(\psi_i\psi_j) = \alpha^2$, $\forall i \neq j$. We carried out simulations for two values of $\alpha^2$ — 0.25 and 0.50, corresponding to cross-sectional correlation of 25% and 50%, respectively. Apart from this change, simulations exactly follow the procedure outlined in Section 4.2.

**Qualitative discussion of results.** As in the previous section, we only provide a qualitative discussion of simulation results.

The simulation results had two features of interest. First, point estimates and standard errors of both $\underline{\theta}$ and $w$ were quite close to the values reported in Table 1. Second, standard errors of estimates of $w$ were 20% to 40% *smaller* than those reported in Table 1; that is, estimates of $w$ appeared to be more precise in the presence of cross-sectional correlation. The standard errors were roughly equal to those computed via the empirical distribution of estimates of $w$. Hence, the lower standard error did reflect more-precise estimates of $w$.

To summarize the simulation results thus far, nonnormality and cross-sectional correlation in private information appear to matter less than imprecisely measured announcement effects. Noise in announcement effects somewhat degrades second-step estimates of $w$, though not significantly so for moderate amounts of noise.

### 5.2 The conventional cross-sectional procedure

How does the traditional cross-sectional procedure perform when data are generated as per the conditional model? We develop some evidence on this question here.

Simulation methodology is similar to that of Section 5.1, with one exception: here, we simulate only firms announcing event E. Abnormal returns are then regressed on a constant term and regressors $x_1$ and $x_2$ to obtain regression coefficients $\beta_0 - \beta_2$ of the linear model, Equation (12), and the associated $t$-statistics. Results are presented in Table 3, for average event probabilities of 25% and 50%.

**5.2.1 Discussion of results.** Table 3 is best interpreted by comparing statistics for OLS regression coefficients $\underline{\beta}$ with corresponding statistics for probit estimates of $\underline{\theta}$ in Table 1.

To begin, observe that while regressors $\underline{x}$ do not satisfy the distributional assumptions of Proposition 1, our simulation results are consistent with its implied comparative statics:

- *Opposite sign*: Average point estimates of $\beta_1$ and $\beta_2$ are signed opposite to $\theta_1$ and $\theta_2$ everywhere.
- *Attenuation*: $\beta_1$ and $\beta_2$ are biased towards zero, relative to $\theta_1$ and $\theta_2$. For instance, $\theta_1$ = -1 everywhere, but average point estimates of $\beta_1$ range from 0.12 to 0.57.
- *More attenuation for smaller $w$*: Consider results of panel A, for instance. As $w$ falls off from 0.75 to 0.30 (going upwards in the table), mean $\beta_1$ drops from 0.57 to 0.22.
- *More attenuation for highly anticipated events*: For instance, keeping $w$ fixed at 0.30, estimates of $\beta_1$ drop from 0.21 to 0.12 as we go from panel A (25% event probability) to panel B (50% event probability).

Second, compare the empirically estimated standard errors (see column Std. error) with OLS standard errors implied in reported $t$-statistics. For instance, consider in panel A the evidence for $w$ = 0.30 and sample size = 100. The $t$-statistic for $\beta_1$ is 0.62; the point estimate of $\beta_1$ is 0.21. Thus, the implied standard error produced by OLS is $\frac{0.21}{0.62} = 0.33$. This is equal to the empirical standard error, which is based on the actual distribution of the simulated parameter estimates. A similar correspondence is seen to hold for every regression coefficient in Tables 3. Consequently, the usual OLS standard errors seem to be appropriate for carrying out significance tests for cross-sectional parameters $\beta_k$.

Finally, how does the linear regression measure up in terms of power, compared to the conditional model? To judge the statistical power of the two procedures, compare the standard errors and $t$-

**Table 3**
**Performance of conventional cross-sectional procedure**

Panel A: Normally distributed information and 25% average event probability

| | | Sample size = 100 | | | Sample size = 250 | | |
|---|---|---|---|---|---|---|---|
| True parameter | Estimated parameter | Mean | Std. error | Mean $t$-stat | Mean | Std. error | Mean $t$-stat |
| Correlation ($w$) = 0.30 | | | | | | | |
| $\theta_0 = 1.00$ | $\beta_0$ | 0.05 | 0.74 | 0.07 | 0.02 | 0.47 | 0.04 |
| $\theta_1 = -1.00$ | $\beta_1$ | 0.21 | 0.33 | 0.62 | 0.22 | 0.20 | 1.05 |
| $100\theta_2 = 1.00$ | $100\beta_2$ | −0.22 | 0.33 | −0.64 | −0.23 | 0.20 | −1.11 |
| Correlation ($w$) = 0.50 | | | | | | | |
| $\theta_0 = 1.00$ | $\beta_0$ | 0.02 | 0.71 | 0.02 | 0.05 | 0.48 | 0.10 |
| $\theta_1 = -1.00$ | $\beta_1$ | 0.39 | 0.32 | 1.26 | 0.39 | 0.22 | 1.86 |
| $100\theta_2 = 1.00$ | $100\beta_2$ | −0.38 | 0.29 | −1.29 | −0.38 | 0.19 | −1.95 |
| Correlation ($w$) = 0.75 | | | | | | | |
| $\theta_0 = 1.00$ | $\beta_0$ | 0.02 | 0.58 | 0.03 | 0.02 | 0.42 | 0.02 |
| $\theta_1 = -1.00$ | $\beta_1$ | 0.57 | 0.25 | 2.21 | 0.57 | 0.18 | 3.25 |
| $100\theta_2 = 1.00$ | $100\beta_2$ | −0.57 | 0.26 | −2.18 | −0.56 | 0.18 | −3.31 |

Panel B: Normally distributed information and 50% average event probability

| | | Sample size = 100 | | | Sample size = 250 | | |
|---|---|---|---|---|---|---|---|
| True parameter | Estimated parameter | Mean | Std. error | Mean $t$-stat | Mean | Std. error | Mean $t$-stat |
| Correlation ($w$) = 0.30 | | | | | | | |
| $\theta_0 = 1.00$ | $\beta_0$ | 0.08 | 0.58 | 0.14 | 0.05 | 0.35 | 0.16 |
| $\theta_1 = -1.00$ | $\beta_1$ | 0.17 | 0.38 | 0.47 | 0.19 | 0.20 | 0.89 |
| $100\theta_2 = 1.00$ | $100\beta_2$ | −0.19 | 0.32 | −0.60 | −0.18 | 0.23 | −0.89 |
| Correlation ($w$) = 0.50 | | | | | | | |
| $\theta_0 = 1.00$ | $\beta_0$ | 0.12 | 0.55 | 0.22 | 0.08 | 0.32 | 0.26 |
| $\theta_1 = -1.00$ | $\beta_1$ | 0.30 | 0.35 | 0.90 | 0.32 | 0.20 | 1.50 |
| $100\theta_2 = 1.00$ | $100\beta_2$ | −0.30 | 0.33 | −0.92 | −0.31 | 0.22 | −1.48 |
| Correlation ($w$) = 0.75 | | | | | | | |
| $\theta_0 = 1.00$ | $\beta_0$ | 0.15 | 0.50 | 0.30 | 0.15 | 0.27 | 0.66 |
| $\theta_1 = -1.00$ | $\beta_1$ | 0.47 | 0.29 | 1.57 | 0.47 | 0.16 | 2.70 |
| $100\theta_2 = 1.00$ | $100\beta_2$ | −0.47 | 0.28 | −1.68 | −0.48 | 0.17 | −2.89 |

Table 3 presents statistics describing performance of the linear model when the true event/abnormal return data are generated by the conditional model, Equation (7), in the text.

Event E is simulated to occur if and only if $\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \psi_i > 0$, where $\theta_0 = 1$, $\theta_1 = -1$, and $\theta_2 = 0.01$. Information $\psi_i$, $i = 1, 2, \ldots, n$ ($n$ is the sample size), is sampled i.i.d. normal, N(0,1). Expected abnormal return, conditional on information $\psi_i$, is simulated as $E(\epsilon_i \mid \psi_i) = w\psi_i$.

In each simulation, we generate a sample of $n$ firms announcing event E, and abnormal returns thereto. We then estimate $\beta_0$, $\beta_1$, and $\beta_2$ in the linear regression $E(\epsilon_i \mid E) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ by OLS. This process is repeated 400 times, and each set of 400 repetitions is done for (1) two sample sizes (100 and 250) and (2) three values of $w$ (0.30, 0.50, and 0.75). From each set of 400 simulations, we compute and report the following statistics: (1) *Mean*: the average parameter point estimate; (2) *Std. error*: the standard error of parameter estimate, computed as the sample standard deviation of the 400 point estimates; and (3) *Mean t-stat*: each simulation yields a $t$-statistic for $\beta_j$, $j = 1, \ldots, 3$. Mean $t$-stat refers to the average of the 400 $t$-statistics.

Based on Section 2.1, we expect regression coefficients $\beta_i$, $i = 1, 2$, to be (1) signed opposite to $\theta_i$ and (2) attenuated, relative to $\theta_i$ (i.e., $|\beta_i| < |\theta_i|$).

statistics for the $\beta_j$, $j = 1, 2$, with those for corresponding probit coefficients $\theta_j$ in Table 1.

On the one hand, the standard errors for the linear regression coefficients $\beta_1$ and $\beta_2$ are 30% to 40% smaller than those for the conditional model estimates of $\theta_1$ and $\theta_2$. Even so, the $t$-statistics for the linear regression coefficients are generally smaller. Thus, even though the linear model produces smaller standard errors, it is *less* powerful than the conditional model in picking up cross-sectional effects. In practice, the gap between the two procedures' performance is likely to be even wider than indicated above, since announcement effects are likely to be measured with error. With measurement error in announcement effects, $t$-statistics for cross-sectional parameters $\beta_j$ of the linear model will be smaller, whereas those for the corresponding $\theta_j$ of the two-step procedure will remain unchanged since the procedure does not use announcement-effect data in cross-sectional parameter estimation.

Attenuation — equivalently, the information in nonevent firms, which is lost here — plays a central role in reducing the linear regression's power. As an illustration of this phenomenon, note that based on Proposition 1, we expect little attenuation and hence more power for the linear regression when $w$ is large and the event is highly informative (i.e., has a low probability). The simulation results are consistent with this intuition: when the average event probability is small (25%) and $w$ is large (0.75), regression coefficients have the largest $t$-statistics relative to all other parameter choices.

The lower power of the linear model suggests the following conjecture: the statistical significance of the linear regression coefficients $\beta_j$ serves as a *lower bound* on significance of the $\theta_j$. In other words, the linear regression is a conservative means of conducting cross-sectional inferences. Hence, if one rejects the hypothesis $\beta_j$ at significance level $\alpha$, one also rejects the hypothesis $\theta_j = 0$, at a significance level of at least $\alpha$. The generality of this conjecture is unknown and warrants investigation in future work.

### 5.3 Conditional model without nonevent information

With both event and nonevent data, the conditional model is superior to the usual OLS procedure and appears to be fairly robust along several dimensions. Absent nonevent data, how does the model perform? We develop some evidence here.[21]

[21] Comments and suggestions of an anonymous referee have motivated and vastly improved much of this section.

We use the methodology described in Section 5.2 to simulate a sample of firms announcing the event and use this data to estimate the conditional model. However, estimation is a more delicate matter in this setting. We discuss some issues that arise, before presenting the simulation results.

With only event data, the conditional model may be estimated by nonlinear least squares (NLS), applied to regression Equation (16), or by maximum likelihood (ML). We began by experimenting with the NLS estimator. This estimator displayed poor statistical properties (it led to upward biased parameter estimates, and standard errors were overstated by a factor of 8 through 10), and was computationally as intensive as the (more-efficient) ML estimator. Accordingly, parameters were estimated via maximum likelihood. This involves maximization of the log-likelihood function

$$L(\underline{\theta}, w, \sigma_\epsilon) = -ln\,\sigma_\epsilon + ln\; n(\epsilon/\sigma_\epsilon) + ln\; N\left(\frac{\underline{\theta}'x + w\epsilon/\sigma_\epsilon}{\sqrt{1-w^2}}\right) - ln\; N(\underline{\theta}'x). \tag{17}$$

The optimization process requires choices along several dimensions. Below, we briefly describe the alternatives experimented with, as well as the decisions made in the final empirical work.

1. *Starting values*: Two sets of starting values were used to initialize the iterations. In the first "benchmark" set, starting values were set equal to the true parameter values. In the second set, all parameters were initialized to zero, except the parameter $\sigma_\epsilon$, which was set to the sample standard deviation of announcement effects. Simulation results from the two sets were not appreciably different.

2. *Optimization algorithm*: We experimented with three algorithms: (1) Gauss-Newton, (2) Davidon-Fletcher-Powell (DFP), and (3) Broyden-Fletcher-Goldfarb-Shanno (BFGS) techniques. The first of these led to repeated problems of nonconvergence or noninvertibility of the Hessian, causing the optimization routine to abort. By contrast, the DFP and BFGS algorithms were better behaved, and the latter was chosen for empirical work.

3. *Gradient and Hessian*: Analytical (not numerical) gradients were employed in the iterations, as these improved computational and convergence properties. $T$-statistics are based on standard errors using the analytical Hessian.

4. *Convergence and exit criteria*: After some experimentation, we deemed the optimization routine to have converged when the relative change in all gradients was less than $10^{-4}$. The maximum number of iterations was set at 175, beyond which there was no appreciable change in convergence behavior.

One final point deserves mention. Observe that the maximand, Equation (17), contains the term $\sqrt{1-w^2}$. For the maximand to be computable, we require that $-1 < w < 1$, a constraint that must be imposed at all stages in the optimization. To get around this difficulty, we reparametrized the problem by setting $w_1 = \frac{w}{\sqrt{1-w^2}}$; equivalently $w = \frac{w_1}{\sqrt{1+w_1^2}}$. The advantage of this formulation is that unlike $w$, parameter $w_1$ is not constrained to lie in (-1,1), freeing the optimization process of this additional constraint. The best set of results obtained with these settings, and estimation was carried out using the OPTMUM routine in GAUSS.

**Discussion of results.** For both sets of starting values, our experience with regard to convergence was satisfactory for all but one set of parameter values.[22] Table 4 reports the results for the subsample of simulations that did converge, and we recognize the potential bias built in favor of conditional methods in interpreting these results.

Two facts emerge from the simulation results. First, the *t*-statistics reported in Table 4 are much smaller than those in Table 1. Thus, the absence of nonevent data has a severe negative impact on the conditional model's performance. Why are nonevent data so crucial to the conditional model's performance? We consider two explanations in this context. First, by using nonevent and event data rather than just event data only, one effectively increases the sample size, and this leads to greater statistical power. A second possibility is that the use of nonevent data expands the type of information being used in estimation. If the type of information represented by nonevent firms is useful in estimation, more powerful inferences should result. Indeed, our analysis supports the second conjecture. In all our simulations, the sample size, that is, the total number of firms (event *plus* nonevent firms) is fixed (at 100 or 250). Nevertheless, the conditional model's performance depends on the type of data within the sample: it performs better when there are both event and nonevent data rather than event data only.

The simulation results also indicate that the statistical properties of the ML estimator are somewhat unsatisfactory. Specifically, (1) parameter estimates are upward-biased, and (2) standard errors are slightly understated. The upshot is that the ML *t*-statistics appear to

[22] Aberrant behavior was displayed in only one instance, when $w$ was small (0.30), and the sample was small (100). Here, only 60% of the iterations converged, and the reason for such behavior is intuitively straightforward. When $w = \text{corr}(\epsilon, \psi)$ is small, little useful information is contained in announcement effects. Thus, the likelihood function becomes flat with respect to $\underline{\theta}$ [this may be verified by allowing $w \to 0$ in Equation (17)], and location of extrema becomes difficult, especially in small samples.

**Table 4**
**Performance of conditional model with truncated samples 50% average event probability**

| | | Sample size = 100 | | | Sample size = 250 | | |
|---|---|---|---|---|---|---|---|
| Parameter | Truth | Mean | Std. error | Mean $t$-stat | Mean | Std. error | Mean $t$-stat |
| $w = 0.30$ | | | | | | | |
| $w_1 = \frac{w}{\sqrt{1-w^2}}$ | 0.31 | 0.93 | 1.19 | 1.39 | 0.67 | 0.79 | 1.91 |
| $\theta_0$ | 1.00 | 2.43 | 11.98 | 2.41 | 0.91 | 3.32 | 0.37 |
| $\theta_1$ | −1.00 | −2.24 | 6.09 | −0.73 | −1.19 | 2.39 | −0.79 |
| $100\theta_2$ | 1.00 | 2.46 | 4.96 | 0.69 | 1.90 | 2.99 | 0.99 |
| $w = 0.50$ | | | | | | | |
| $w_1 = \frac{w}{\sqrt{1-w^2}}$ | 0.58 | 1.12 | 1.39 | 1.69 | 0.89 | 0.81 | 2.37 |
| $\theta_0$ | 1.00 | 1.52 | 3.69 | 0.66 | 1.17 | 1.84 | 1.02 |
| $\theta_1$ | −1.00 | −1.56 | 2.15 | −0.96 | −1.37 | 1.34 | −1.51 |
| $100\theta_2$ | 1.00 | 1.60 | 2.06 | 0.99 | 1.17 | 1.08 | 1.40 |
| $w = 0.75$ | | | | | | | |
| $w_1 = \frac{w}{\sqrt{1-w^2}}$ | 1.13 | 1.45 | 1.07 | 1.91 | 1.17 | 0.54 | 2.77 |
| $\theta_0$ | 1.00 | 0.91 | 1.29 | 1.03 | 0.90 | 0.85 | 1.52 |
| $\theta_1$ | −1.00 | −1.20 | 0.82 | −1.54 | −1.16 | 0.63 | −2.28 |
| $100\theta_2$ | 1.00 | 1.53 | 1.00 | 1.45 | 1.17 | 0.54 | 2.21 |

Table 4 presents statistics describing performance of the conditional model, Equation (7), in the text, when data is available only for the firms announcing the event.
Event E is simulated to occur if and only if $\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \psi_i > 0$, where $\theta_0 = 1, \theta_1 = -1$, and $\theta_2 = 0.01$. Information $\psi_i$, $i = 1, 2, \ldots, n$ ($n$ is the sample size), is sampled i.i.d. normal, N(0,1). Expected abnormal return, conditional on information $\psi_i$, is simulated as $E(\epsilon_i \mid \psi_i) = w\psi_i$.
In each simulation, we generate a sample of $n$ firms announcing event E, and abnormal returns thereto. We then estimate parameters $w_1 = \frac{w}{\sqrt{1-w^2}}$, $\theta_0$, and $\theta_1, \theta_2$ by maximum likelihood, the likelihood function being defined in Equation (17) in the text. This process is repeated 400 times, and each set of 400 repetitions is done for (1) two sample sizes (100 and 250) and (2) three values of $w$ (0.30, 0.50, and 0.75). From each set of 400 simulations, we compute and report the following statistics: (1) *Mean*: the average parameter point estimate; (2) *Std. error*: the standard error of parameter estimate, computed as the sample standard deviation of the 400 point estimates; and (3) *Mean t-stat*: each simulation yields a $t$-statistic for parameters $w_1$ and $\theta_j$, $j = 1, \ldots, 3$. Mean $t$-stat refers to the average of these $t$-statistics.

be somewhat larger than their true values. As an illustration of this phenomenon, consider the instance when the sample size is 100 and $w$ = 0.50. Here, the average ML point estimate of $\theta_1$ is -1.56, and the average $t$-statistic for $\theta_1$ is -0.96, implying that the ML standard error is about $\frac{1.56}{0.96} = 1.62$. However, the true standard error (see column Std. error) is larger at 2.15. Thus, the reported ML $t$-statistic appears to be overstated. From Table 4, we see that this is more of an issue for small $w$ and less so for larger $w$ and sample sizes.

Despite this upward bias in the ML $t$-statistics, the $t$-statistics are no better than those produced by OLS (see Table 3, panel B). In this context, one situation — $w$ = 0.75 and sample size = 250 — is somewhat interesting. For this parameter set, every one of the conditional model simulations converged; even so, ML $t$-statistics are 25% smaller

than the corresponding OLS counterparts! Thus, absent nonevent data, there is little evidence that the specification of the conditional model analyzed here has any practical value, relative to the much simpler OLS procedure.

## 6. Conclusions

Conditional methods offer an interesting perspective of event studies. Such methods are derived in the context of a well-defined economic equilibrium and have simple and appealing intuition: they relate announcement effects to the unexpected information revealed in events. Hence, conditional methods are potentially attractive means of conducting event studies.

The conditional model proposed by Acharya (1988) is a natural choice for many corporate events. When is it an empirically valuable tool? We find that it performs well only when one has, in addition to data on firms announcing the event, a set of nonevent firms, that is, firms that were partially anticipated to announce but chose not to announce the event in question. If such data are available, the conditional model is a valuable means of inference since it allows one to exploit nonevent information — which lies unused under traditional methods — in a straightforward manner.

In such settings, a simple two-step procedure appears to be an attractive method of estimating the conditional model. This estimator has three useful properties as a means of conducting cross-sectional tests in event studies. First, cross-sectional parameters are estimated without using abnormal return information. Thus, the usual data problems associated with event studies — event-date uncertainty, clustering of event dates, etc. — which are known to adversely afflict inferences via conventional procedures, do not affect cross-sectional inferences via the two-step procedure. Second, our evidence suggests cross-sectional inferences are not severely affected by incorrect distributional assumptions. Nevertheless, if this is a concern, there does exist a body of literature for distribution-free estimation that may be employed. Finally, cross-sectional inferences via the two-step procedure are also valid for other conditional models [e.g., that of Eckbo, Maksimovic, and Williams (1990)] discussed in this article. Thus, when nonevent data are available, the two-step procedure appears to be a desirable way of estimating the conditional model.

However, in most practical situations, nonevent data — a sample of firms that chose not to announce the event, the timing of this nonevent, and cross-sectional data and announcement effects at the time markets learn of the nonevent — are not available. One must then work only with firms that have announced the event in ques-

tion. Here, we find that the conditional model becomes computationally burdensome and less powerful, even when efficiently estimated.

It is precisely under these circumstances that the results concerning traditional methods assume the greatest force, and we make two points in this context. First, regression coefficients obtained via the traditional linear regression are proportional to the true cross-sectional parameters, under weak conditions. Second, OLS standard errors appear to be appropriate for testing the significance of cross-sectional parameters. Together, the two results imply that the traditional OLS procedure may be used for carrying out cross-sectional inferences, though the coefficients are potentially inconsistently estimated. These results also provide an equilibrium justification for using the standard procedures in practice.

How useful are the results, from a practical perspective? Not surprisingly, OLS is less powerful than the conditional model when both event and nonevent data are available. However, when one has event data only, OLS appears to be a simple and effective substitute for the conditional model, even when the latter is efficiently estimated.

Thus, our results suggest that when the necessary nonevent data are available, inferences should be based on conditional methods. If not, we suggest that the traditional cross-sectional procedure be used and the associated $t$-statistics be interpreted as conservative lower bounds on the true significance level of the parameters.

## Appendix

*Proof of Proposition 1.* Following Goldberger (1981) or Maddala (1983), we first reparametrize the selection equation so that $\underline{x}$ and $\tau$ have mean zero. With this reparametrization, event E occurs if and only if $\tau = \theta'\underline{x} + \psi > c$, where

$$c = -\theta_0 - \sum_{j=1}^{k} \theta_j E(x_j) \tag{18}$$

and $E(x_j)$ denotes the mean of the original regressor $x_j$. The proof simply consists of working through the normal equation defining the OLS estimator of $\beta$, for a sample of firms announcing $E$. We begin by stating some results that aid in this process.

The population regression of $\tau$ on $\underline{x}$ is given by

$$\underline{\theta} = \Sigma_x^{-1} cov(\underline{x}, \tau). \tag{19}$$

With $E(\tau)$ normalized to zero, the population mean of $\underline{x}$, conditional on $\tau$, is given by

$$E(\underline{x} \mid \tau) = \underline{\alpha}\tau, \tag{20}$$

while the variance of $\underline{x}$, conditional on $\tau$, is

$$V = Var(\underline{x} \mid \tau) = \Sigma_x - \underline{\alpha}\,\underline{\alpha}'s^2, \tag{21}$$

where $\alpha$ and $s^2$ are defined as

$$\underline{\alpha} = \frac{cov(\underline{x}, \tau)}{s^2} = \frac{\Sigma_x \underline{\theta}}{s^2}, \tag{22}$$

and

$$s^2 = var(\tau) = \underline{\theta}'\Sigma_x\underline{\theta} + 1. \tag{23}$$

Define $R^2$ and $t$ as

$$R^2 = \frac{var(\underline{\theta}'\underline{x})}{s^2} = \frac{\underline{\theta}'\Sigma_x\underline{\theta}}{\underline{\theta}'\Sigma_x\underline{\theta} + 1}, \tag{24}$$

and

$$t = \frac{var(\tau \mid E)}{var(\tau)} = \frac{var(\tau^*)}{s^2}, \tag{25}$$

where the asterisk denotes the conditioning $\tau > c$. $R^2$ is the "explained variance" in the population regression of $\tau$ on $\underline{x}$.

The principal fact on which the proportionality result rests is as follows: Selection does not alter the conditional distribution of $\underline{x}$, given $\tau$ [Chung and Goldberger (1984)]. Hence, for the sample of firms announcing E, we have

$$E(\underline{x}^* \mid \tau) = \underline{\alpha}\tau, \tag{26}$$

and

$$var(\underline{x}^* \mid \tau) = V, \tag{27}$$

where $V, \underline{\alpha}$ are defined in Equations (21) and (22). With these results in hand, we can solve for the coefficients $\underline{\beta}$ in the regression of event-date abnormal returns on regressors $\underline{x}$. These coefficients are defined by the normal equation

$$\Sigma_x^*\underline{\beta} = cov(\underline{x}^*, \epsilon^*), \tag{28}$$

where $\epsilon^*$ is the event-date abnormal return, conditional on announce-

ment of E (i.e., $\tau > c$). Consider first the left-hand side of Equation (28). We have from the definition of $\Sigma_x^*$,

$$\begin{aligned}
\Sigma_x^* &= var_\tau(E^*(\underline{x} \mid \tau)) + E_\tau(var^*(\underline{x} \mid \tau)), \\
&= var(\underline{\alpha}\tau^*) + V \quad \text{[from Equations (26) and (27)]}, \\
&= \underline{\alpha}\, var(\tau^*)\underline{\alpha}' + \Sigma_x - \underline{\alpha}\,\underline{\alpha}' s^2 \quad \text{[from Equation (21)]}, \\
&= \underline{\alpha}(ts^2)\underline{\alpha}' + \Sigma_x - \underline{\alpha}\,\underline{\alpha}' s^2 \quad \text{[from Equation (25)]}, \\
&= \Sigma_x - \underline{\alpha}\,\underline{\alpha}'(1-t)s^2. \qquad (29)
\end{aligned}$$

To simplify the right-hand side of Equation (28), use the bivariate normality of $\psi$ and $\epsilon$ to get

$$E(\epsilon \mid \psi) = \pi\psi = \pi(\tau - \underline{\theta}'\underline{x}),$$

and

$$E(\epsilon \mid E) = E(\epsilon^*) = \pi(\tau^* - \underline{\theta}'\underline{x}^*).$$

Using this in the right-hand side of Equation (28), we have

$$\begin{aligned}
cov(\underline{x}^*, \epsilon^*) &= cov(\underline{x}^*, \pi(\tau^* - \underline{\theta}'\underline{x}^*)), \\
&= \pi\, cov(\underline{x}^*, \tau^*) - \pi\, cov(\underline{x}^*, \underline{\theta}'\underline{x}^*), \\
&= \pi\underline{\alpha}ts^2 - \pi\Sigma_x^*\underline{\theta} \quad \text{[using Equation (26)]}, \\
&= \pi\underline{\alpha}ts^2 - \pi\left[\Sigma_x - \underline{\alpha}\,\underline{\alpha}'(1-t)s^2\right]\underline{\theta} \text{ [using Equation (29)]}.
\end{aligned}$$

Substituting for $\underline{\alpha}$ using Equation (22), we have

$$\begin{aligned}
cov(\underline{x}^*, \epsilon^*) &= \pi\left(\frac{\Sigma_x\underline{\theta}}{s^2}\right)ts^2 - \pi\Sigma_x\underline{\theta} + \pi s^2\underline{\theta}\left(\frac{\Sigma_x\underline{\theta}}{s^2}\right)\left(\frac{\underline{\theta}'\Sigma_x}{s^2}\right)(1-t), \\
&= \pi\Sigma_x\underline{\theta}\left[t - 1 + \left(\frac{\underline{\theta}'\Sigma_x\underline{\theta}}{s^2}\right)(1-t)\right], \\
&= \pi\Sigma_x\underline{\theta}\left[t - 1 + R^2(1-t)\right] \quad \text{[using Equation (24)]}, \\
&= -\pi(1-R^2)(1-t)\Sigma_x\underline{\theta}. \qquad (30)
\end{aligned}$$

Finally, using Equations (29) and (30) in normal Equation (28), we have

$$\underline{\beta}\left[\Sigma_x - \underline{\alpha}\,\underline{\alpha}'(1-t)s^2\right] = \pi\Sigma_x\underline{\theta}(1-R^2)(1-t)$$

and

$$\underline{\beta}\left[\Sigma_x - \left(\frac{\Sigma_x\underline{\theta}}{s^2}\right)\left(\frac{\underline{\theta}'\Sigma_x}{s^2}\right)(1-t)s^2\right] = -\pi\Sigma_x\underline{\theta}(1-R^2)(1-t), \quad (31)$$

where the last equation obtains by substituting for $\underline{\alpha}$ using Equation (22). Multiplying both sides of Equation (31) by $\underline{\theta}'$, we have, as

required,

$$(\underline{\theta}'\Sigma_x)\left\{\underline{\beta}\left[1-\left(\frac{\underline{\theta}'\Sigma_x\underline{\theta}}{s^2}\right)(1-t)\right]\right\} = \underline{\theta}'\Sigma_x\left\{\left[-\pi\underline{\theta}(1-R^2)(1-t)\right]\right\},$$

$$\Rightarrow \underline{\beta} = -\pi\underline{\theta}\frac{(1-R^2)(1-t)}{t+(1-R^2)(1-t)},$$

$$= -\pi\underline{\theta}\mu. \tag{32}$$

***Remark***. *The intercept term $\beta_0$ can be computed as $E(\epsilon^* - \underline{\beta}'\underline{x}^*)$. Using (1) Proposition 1 for $(\beta_1, \ldots, \beta_k)$ (2) $E(\tau^*) = \lambda_E(-c)$ and $E(x^*) = \underline{\alpha}\tau^*$, we obtain $\beta_0 = \pi\lambda_E(-c)(1-\mu R^2)$, where c is defined in Equation (18). Also, while the proof pertains to conditional model defined by Equation (7), an analogous proportionality result may be obtained in similar fashion for encompassing specification, Equation (11), as well.*

*Proof of Lemma 1: Result (1).* The result $0 < \mu < 1$ is an immediate consequence of $0 < t, R^2 < 1$. The bounds on $R^2$ follow from its definition, Equation (24); those for $T$ need some argument. From Equation (25), the definition of $t$, we have

$$t = \frac{var(\tau \mid E)}{var(\tau)} = \frac{var(\tau \mid \tau > c)}{s^2}, \tag{33}$$

$$= var\left(\frac{\tau}{s} \mid \frac{\tau}{s} > \frac{c}{s}\right) = var\left(z \mid z > \frac{c}{s}\right), \tag{34}$$

where $z = \frac{\tau}{s}$ is standard normal (multivariate normality of $\underline{x}$ is being invoked here). But if z is standard normal, $0 < var(z \mid z > \frac{c}{s}) < 1$, [see Greene (1993), and footnote 11 in the text]. Therefore, $0 < t < 1$.

*Proof of Lemma 1: Result (2).* For highly anticipated events, we show that $t$ is large. As $\frac{\partial\mu}{\partial t} < 0$, it follows that $\mu$ is small for such events. Accordingly, consider the following facts:

Fact 1: Event E occurs if and only if $\theta_0 + \sum_{j=1}^{k}\theta_j x_j + \psi > c$, where c is defined in Equation (18). Hence, the smaller the value of c, the greater the average probability of E.

Fact 2: From Equation (34), if $z$ is standard normal, $t = var(z \mid z > \frac{c}{s}) = var(z \mid z < -\frac{c}{s})$, where the last equality follows from symmetry. From Goldberger (1983), $\frac{\partial var(z|z<-\frac{c}{s})}{\partial c} = \frac{\partial t}{\partial c} < 0$.

From Facts 1 and 2, we have the following. For highly anticipated events, c is small (from Fact 1) and t is large (as $\frac{\partial t}{\partial c} < 0$ from Fact 2). But large t implies small $\mu$ since $\frac{\partial\mu}{\partial t} < 0$, using Equation (32). Thus, for highly anticipated events, $\mu$ is small, as claimed.

**References**

Acharya, S., 1988, "A Generalized Econometric Model and Tests of a Signalling Hypothesis with Two Discrete Signals," *Journal of Finance*, 43, 413–29.

Acharya, S., 1993, "Value of Latent Information: Alternative Event Study Methods", *Journal of Finance*, 48, 363–85.

Arabmazar, A., and P. Schmidt, 1982, "An Investigation into the Robustness of the Tobit Estimator to Nonnormality," *Econometrica*, 50, 1055–63.

Asquith, P., and D. Mullins, 1986, "Equity Issues and Offering Dilution," *Journal of Financial Economics*, 15, 61–89.

Asquith, P., R. Bruner, and D. Mullins, "The Gain to Bidding Firms from Merger," *Journal of Financial Economics*, 11, 121–39.

Bajaj, M., and A. Vijh, "Dividend Clienteles and the Information Content of Dividend Changes," *Journal of Financial Economics*, 26, 193–219.

Brown, S., and J. Warner, 1985, "Using Daily Stock Returns: The Case of Event Studies," *Journal of Financial Economics*, 14, 3–31.

Chaplinsky, S., and R. S. Hansen, 1993, "Partial Anticipation, the Flow of Information and the Economic Impact of Corporate Debt Sales," *Review of Financial Studies*, 6, 709–32.

Chung, C-F., and A. Goldberger, 1984, "Proportional Projections in Limited Dependent Variable Models," *Econometrica*, 52, 531–4.

Eckbo, E., 1986, "Valuation Effects of Corporate Debt Offerings," *Journal of Financial Economics*, 15, 119–51.

Eckbo, E., V. Maksimovic, and J. Williams, 1990, "Consistent Estimation of Cross-Sectional Models in Event-Studies," *Review of Financial Studies*, 3, 343–65.

Fama, E., L. Fisher, M. Jensen, and R. Roll, 1969, "The Adjustment of Stock Prices to New Information," *International Economic Review*, 10, 1–21.

Goldberger, A., 1981, "Linear Regression After Selection," *Journal of Econometrics*, 15, 357–66.

Goldberger, A., 1983, "Abnormal Selection Bias," in S. Karlin, T. Amemiya, and L. Goodman, (eds.), *Studies in Econometrics, Time Series and Multivariate Statistics*, Academic Press, Stamford.

Greene, W., 1981, "Sample Selection Bias as a Specification Error: Comment," *Econometrica*, 49, 795–8.

Greene, W., 1993, *Econometric Analysis*, Macmillan, New York.

Heckman, J., 1979, "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–61.

Korajczyk, R., D. Lucas, and R. McDonald, 1991, "The Effect of Information Releases on the Pricing and Timing of Equity Issues," *Review of Financial Studies*, 4, 685–708.

Lanen, W., and R. Thompson, 1988, "Stock Price Reactions as Surrogates for the Net Cashflow Effects of Corporate Financial Decisions," *Journal of Accounting and Economics*, 10, 311–34.

Lee, L-F., 1982, "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies*, 49, 355–72.

Lee, L-F., 1983, "Generalized Econometric Models with Selectivity," *Econometrica*, 51, 507–12.

Maddala, G. S., 1983, *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, MA.

Malatesta, P., and R. Thompson, 1985, "Partially Anticipated Events: A Model of Stock Price Reactions with an Application to Corporate Acquisitions," *Journal of Financial Economics*, 14, 237–50.

McElroy, M., and E. Burmeister, 1988, "Arbitrage Pricing Theory as a Restricted Nonlinear Regression Model," *Journal of Business Economics and Statistics*, 6, 29–42.

Nelson, F., 1984, "Efficiency of the Two-Step Estimator with Endogenous Sample Selection," *Journal of Econometrics*, 24, 181–96.

Newey, W., J. Powell, and J. Walker, 1990, "Semiparametric Estimation of Selection Models: Some Empirical Results," *American Economic Review*, 80, 324–28.

Paarsch, H., 1984, "A Monte-Carlo Comparison of Estimators for Censored Regression Models," *Journal of Econometrics*, 24, 197–213.

Schipper, K., and R. Thompson, 1983, "Evidence on the Capitalized Value of Merger Activity for Acquiring Firms," *Journal of Financial Economics*, 11, 85–119.

Stoker, T., 1986, "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461–81.

Thompson, R., 1985, "Conditioning the Return Generating Process on Firm-Specific Events: A Discussion of Event-Study Methods," *Journal of Financial and Quantitative Analysis*, 20, 151–68.

Vermaelen, T., 1984, "Repurchase Tender Offers, Signalling, and Managerial Incentives," *Journal of Financial and Quantitative Analysis*, 19, 163–81.

Wales, T., and A. Woodland, 1980, "Sample Selectivity and the Estimation of Labor Supply Functions," *International Economic Review*, 21, 437–68.

White, H., 1980, "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21, 149–70.