



ROBERT H. SMITH
SCHOOL OF BUSINESS

Leaders for the **Digital Economy**

Working Paper No. RHS-06-020

September, 2005

Self-Selection Models in Corporate Finance

Kai Li

Sauder School of Business
University of British Columbia

N.R. Prabhala

Robert H. Smith School of Business
University of Maryland

This Paper can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection
<http://ssrn.com/abstract=843105>

Self-Selection Models in Corporate Finance¹

Kai Li

Sauder School of Business
University of British Columbia

N.R. Prabhala

Robert H. Smith School of Business
University of Maryland

September, 2005

Abstract

Corporate finance decisions are not made at random, but are usually deliberate decisions by firms or their managers to *self-select* into their preferred choices. This chapter reviews econometric models of self-selection. The review is organized into two parts. The first part reviews econometric models of self-selection, focusing on the key assumptions of different models and the types of applications they may be best suited for. Part two reviews empirical applications of selection models in the areas of corporate investment, financing, and financial intermediation. We find that self-selection is a rapidly growing area in corporate finance, partly reflecting its recognition as a pervasive feature of corporate finance decisions, but more importantly, the increasing recognition of selection models as unique tools for understanding, modeling, and testing the role of private information in corporate finance.

¹We thank N. K. Chidambaran, Espen Eckbo, Bill Greene, Gordon Phillips, Vojislav Maksimovic, Manju Puri, Jeffrey Smith, and Xinlei Zhao for helpful discussions without implicating them for any errors or omissions, which remain ours. Li acknowledges the financial support from the Social Sciences and Humanities Research Council of Canada, and the W.M. Young Chair in Finance from the Sauder School of Business at UBC. Li also wishes to thank the support and hospitality of the MIT Sloan School of Management where she completed most of her work on this chapter.

Table of Contents

Introduction

I. Modeling self-selection

1. Self-selection: The statistical issue
2. The baseline Heckman selection model
 - 2.1 The econometric model
 - 2.2 Self-selection and private information
 - 2.2.1 Selection: An omitted variable problem
 - 2.2.2 The omitted variable as private information
 - 2.3 Specification issues
 - 2.3.1 Exclusion restrictions
 - 2.3.2 Bivariate normality
3. Extensions
 - 3.1 Switching regressions
 - 3.2 Simultaneity in self-selection models
 - 3.2.1 The Roy model
 - 3.2.2 Structural self-selection models
4. Matching models and self-selection
 - 4.1 Treatment effects
 - 4.2 Treatment effects from selection models
 - 4.3 Treatment effects from matching models
 - 4.3.1 Dimension-by-dimension matching
 - 4.3.2 Propensity score (PS) matching
 - 4.3.3 Implementation of PS methods
5. Panel data with fixed effects
6. Bayesian self-selection models
 - 6.1 Bayesian methods

6.2 Bayesian methods for selection models

II. Empirical applications

7. Event studies

- 7.1 Conditional announcement effects: Acharya (1988)
- 7.2 Two announcements on the same date: Nayak and Prabhala (2001)
- 7.3 Takeovers: Eckbo, Maksimovic and Williams (1990)
- 7.4 Takeover deterrence: Eckbo (1992)

8. The pricing of public debt offerings

- 8.1 Banks underwritings and the Glass-Steagall Act: Puri (1996)
- 8.2 Underwriting syndicate structure: Song (2004)
- 8.3 Underwriter reputation: Fang (2005)
- 8.4 Debt covenants: Goyal (2005)
- 8.5 Discussion

9. Other investment banking applications

- 9.1 Underwriter compensation in IPOs: Dunbar (1995)
- 9.2 Analyst coverage: Ljungqvist, Marston and Wilhelm (2005)

10. Diversification discount

- 10.1 Unobservables and the diversification discount: Campa and Kedia (2002)
- 10.2 Observables and the discount: Villalonga (2004)
- 10.3 Refocusing and the discount: Çolak and Whited (2005)
- 10.4 Discussion

11. Other applications of selection models

- 11.1 Accounting for R&D: Shehata (1991)
- 11.2 Bankruptcy costs: Bris, Welch and Zhu (2005)
- 11.3 Family ownership and value: Villalonga and Amit (2004)

12. Other applications of matching methods

- 12.1 Bank debt versus bonds: Bharath (2004)
- 12.2 Matching and long-run performance: Cheng (2003), Li and Zhao (2005)

13. Bayesian methods

13.1 Matching: Sørensen (2003)

13.2 Switching regressions: Li and McNally (2004), Scruggs (2005)

14. Conclusions

Introduction

Corporate finance concerns the financing and investment decisions made by firms and a broad swathe of decisions within these broad choices. For instance, firms pick their target capital structure, and to achieve the target, must make several choices that includes the timing of security issues, structural features of the securities issued, the investment bank chosen to underwrite it, and so on. These choices are not usually random, but are deliberate decisions by firms or their managers to *self-select* into their preferred choices. This chapter reviews econometric models of self-selection. We review the approaches used to model self-selection in corporate finance and the substantive findings obtained by implementing selection methods.

Self-selection has a rather mixed history in corporate finance. The fact that there is self-selection is probably not news; indeed, many papers at least implicitly acknowledge its existence. However, the literature differs on whether to account for self-selection using formal econometric methods, and why one should do so. One view of self-selection is that it is an errant nuisance, a “correction” that must be made to prevent other parameter estimates from being biased. Selection is itself of little economic interest under this view. In other applications, self-selection is itself of central economic interest, because the self-selection model represents one way of incorporating and controlling for unobservable private information that influences corporate finance decisions. Both perspectives find expression in the literature, although an increasing emphasis in recent work reflects the positive view in which selection models are used to construct interesting tests for private information.

Our review is organized into two parts. Part I focuses on econometric models of self-selection. We approach selection models from the viewpoint of a corporate finance researcher who is implementing selection models in an empirical application. We formalize the notion of self-selection and overview several approaches towards modeling it, including reduced form models, structural approaches, matching methods, fixed effect estimators, and Bayesian methods. As the discussion clarifies, the notion of selection is not monolithic. No single model universally models or accounts for all forms of selection. Instead, there are a variety of approaches, each of which makes its own economic and statistical assumptions. We focus on the substantive economic assumptions underlying the different approaches to illustrate what each can and cannot do and the type of applications a given approach may be best suited for. We do not say much on estimation, asymptotic inference, or computational issues, but refer the reader to excellent texts and articles on these matters.

Part II of our review examines corporate finance applications of self-selection models. We cover a range of topics such as mergers and acquisitions, stock splits, equity offerings, underwriting, analyst behavior, share repurchases, and venture capital. Our objective is to illustrate the wide range of corporate finance settings in which selection arises and the different econometric approaches employed in modeling it. Here, we focus on applications published in the last decade or so, and on articles in which self-selection is a major component of the overall results.²

I. Modeling self-selection

This portion of our review discusses econometric models of self-selection. Our intention is not to summarize the entire range of available models and their estimation. Rather, we narrow our focus to models that have been applied in the corporate finance literature, and within these models, we focus on the substantive assumptions made by each specification. From the viewpoint of the empirical researcher, this is the first order issue in deciding what approach suits a given application in corporate finance. We do not touch upon asymptotic theory, estimation, and computation. These important issues are well covered in excellent textbooks.³

We proceed as follows. Section 1 describes the statistical issue raised by self-selection. Sections 2 to 6 develop the econometric models that can address selection. Section 2 discusses a baseline model for self-selection, the “Heckman” selection model analyzed in Heckman (1979), a popular modeling choice in corporate finance.⁴ We discuss identification issues related to the model, which are important but not frequently discussed or justified explicitly in corporate finance applications. Because the Heckman setting is so familiar in corporate finance, we use it to develop a key point of this survey, the analogy between econometric models of self-selection and private information models in corporate finance. Section 3 considers switching regressions and structural self-selection

²Our attempt is to capture the overall flavor of self-selection models as they stand in corporate finance as of the writing. We apologize to any authors whose work we have overlooked: no slight is intended.

³The venerable reference, Maddala (1983), continues to be remarkably useful, though its notation is often (and annoyingly, to the empirical researcher) different from the ones used in other articles and software packages. Newer material is covered in Wooldridge (2002) and Greene (2004).

⁴Labeling any one model as “the” Heckman model surely does disservice to the many other contributions of James Heckman. We choose this label following common usage in the literature.

models. While these models generalize the Heckman selection model in some ways, they also bring additional baggage in terms of economic and statistical assumptions that we discuss.

We then turn to other approaches towards modeling selection. Section 4 discusses matching models, which are methods *du jour* in the most recent applications. The popularity of matching models can be attributed to their relative simplicity, easy interpretation of coefficients, and minimal structure with regard to specification. However, these gains come at a price. Matching models make the strong economic assumption that unobservable private information is irrelevant. This assumption may not be realistic in many corporate finance applications. In contrast, selection models explicitly model and incorporate private information. A second point we develop is that while matching methods are often motivated by the fact that they yield easily interpretable *treatment effects*, selection methods also estimate treatment effects with equal ease. Our review of methodology closes by briefly touching upon fixed effect models in Section 5 and Bayesian approaches to selection in Section 6.

1 Self-Selection: The statistical issue

To set up the self-selection issue, assume that we wish to estimate parameters β of the regression

$$Y_i = X_i\beta + \epsilon_i \tag{1}$$

for a population of firms. In Eq. (1), Y_i is the dependent variable, which is typically an *outcome* such as profitability or return. The variables explaining outcomes are X_i , and the error term is ϵ_i . If ϵ_i satisfies usual classical regression conditions, standard OLS/GLS procedures consistently estimate β .

Now consider a sub-sample of firms who self-select choice E . For this sub-sample, Eq. (1) can be written as

$$Y_i | E = X_i\beta + \epsilon_i | E \tag{2}$$

The difference between Eqs. (2) and (1) is at the heart of the self-selection problem. Eq. (1) is a specification written for the population but Eq. (2) is written for a subset of firms, those that self-select choice E . If self-selecting firms are not random subsets of the population, the usual OLS/GLS estimators applied to Eq. (2), are no longer consistent estimators of β .

Accounting for self-selection consists of two steps. Step 1 specifies a model for self-selection, using economic theory to model why some firms select E while others do not. While this specification step is not often discussed extensively in applications, it is critical because the assumptions involved ultimately dictate what econometric model should be used in the empirical application. Step 2 ties the random variable(s) driving self-selection to the outcome variable Y .

2 The baseline Heckman selection model

2.1 The econometric model

Early corporate finance applications of self-selection are based on the model analyzed in Heckman (1979). We spend some time developing this model because most other specifications used in the literature can be viewed as extensions of the Heckman model in various directions.

In the conventional perspective of self-selection, the key issue is that we have a regression such as Eq. (1) that is well specified for a population but it must be estimated for sub-samples of firms that self-select into choice E . To estimate population parameters from self-selected subsamples, we first specify a self-selection mechanism. This usually takes the form of a probit model in which firm i chooses E if the net benefit from doing so, a scalar W_i , is positive. Writing the selection variable W_i as a function of explanatory variables Z_i , which are assumed for now to be exogenous,⁵ we have the system

$$C = E \quad \equiv \quad W_i = Z_i\gamma + \eta_i > 0 \quad (3)$$

$$C = NE \quad \equiv \quad W_i = Z_i\gamma + \eta_i \leq 0 \quad (4)$$

$$Y_i = X_i\beta + \epsilon_i \quad (5)$$

where Z_i denotes publicly known information influencing a firm's choice, γ is a vector of probit coefficients, and η_i is orthogonal to public variables Z_i . In the standard model, Y_i is observed when a firm picks one of E or NE , so Eq. (5) would require the appropriate conditioning. Assuming that η_i and ϵ_i are bivariate normal, the likelihood function and the maximum likelihood estimators for Eqs. (3)-(5) follow, although a simpler two-step procedure (Heckman (1979), and Greene (1981)) is

⁵Thus, we preclude for now the possibility that Z includes the outcome variable Y . This restriction can be relaxed at a cost, as we show in later sections.

commonly used for estimation. Virtually all applied work is based on the bivariate normal structure discussed above.

2.2 Self-selection and private information

In the above setup, self-selection is a nuisance problem. We model it because not doing so leads to inconsistent estimates of parameters β in regression Eq. (1). Self-selection is, by itself, of little interest. However, this situation is frequently reversed in corporate finance, because tests for self-selection can be viewed as tests of private information theories. We develop this point in the context of the Heckman (1979) model outlined above, but we emphasize that this private information interpretation is more general.

We proceed as follows. Following a well-established tradition in econometrics, Section 2.2.1 presents selection as an omitted variable problem. Section 2.2.2 interprets the omitted variable as a proxy for unobserved private information. Thus, including the omitted self-selection variable controls for and tests for the significance of private information in explaining ex-post outcomes of corporate finance choices.

2.2.1 Selection: An omitted variable problem

Suppose that firm i self-selects choice E . For firm i , we can take expectations of Eq. (5) and write

$$Y_i | E = X_i\beta + (\epsilon_i | Z_i\gamma + \eta_i > 0) \quad (6)$$

$$= X_i\beta + \pi(\eta_i | Z_i\gamma + \eta_i > 0) + \nu_i \quad (7)$$

Eq. (7) follows from the standard result that $\epsilon_i | \eta_i = \pi\eta_i + \nu_i$ where π is the coefficient in the regression of ϵ_i on η_i , and ν_i is an orthogonal zero-mean error term.⁶ Given the orthogonality and zero-mean properties of ν_i , we can take expectations of Eq. (7) and obtain the regression model

$$E(Y_i | E) = X_i\beta + \pi E(\eta_i | Z_i\gamma + \eta_i > 0) \quad (8)$$

and a similar model for firms choosing not to announce E ,

$$E(Y_i | NE) = X_i\beta + \pi E(\eta_i | Z_i\gamma + \eta_i \leq 0) \quad (9)$$

⁶Note that $\pi = \rho_{\eta\epsilon}\sigma_\epsilon$ where $\rho_{\eta\epsilon}$ is the correlation between ϵ and η , and σ_ϵ^2 is the variance of ϵ .

Eqs. (8)-(9) can be compactly rewritten as

$$E(Y_i|C) = X_i\beta + \pi\lambda_C(Z_i\gamma) \quad (10)$$

where $C \in \{E, NE\}$ and $\lambda_C(\cdot)$ is the conditional expectation of η_i given C . In particular, if η and ϵ are bivariate normal, as is standard in the bulk of the applied work, $\lambda_E(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$ and $\lambda_{NE}(\cdot) = -\frac{\phi(\cdot)}{1-\Phi(\cdot)}$ (Greene (2003), p. 759).

A comparison of Eq. (1) and Eq. (10) clarifies why self-selection is an omitted variable problem. In the population regression in Eq. (1), regressing outcome Y on X consistently estimates β . However, in self-selected samples, consistent estimation requires that we include an additional variable, the inverse Mills ratio $\lambda_C(\cdot)$. Thus, the process of correction for self-selection can be viewed as including an omitted variable.

2.2.2 The omitted variable as private information

In the probit model Eqs. (3)-(4), η_i is the part of W_i not explained by public variables Z_i . Thus, η_i can be viewed as the private information driving the corporate financing decision being modeled. The ex-ante expectation of η_i should be zero, and it is so, given that it has been defined as an error term in the probit model.

Ex-post after firm i selects $C \in \{E, NE\}$, the expectations of η_i can be updated. The revised expectation, $E(\eta_i | C)$, is thus an updated estimate of the firm's private information. If we wished to test whether the private information in a firm's choice affected post-choice outcomes, we would regress outcome Y on $E(\eta_i | C)$. But $E(\eta_i|C) = \lambda_C(\cdot)$ is the inverse Mills ratio term that we add anyway to adjust for self-selection. Thus, correcting for self-selection is equivalent to testing the private information. The omitted variable used to correct for self-selection, $\lambda_C(\cdot)$, is an estimate of the private information underlying a firm's choice and testing its significance is a test of whether private information possessed by a firm explains ex-post outcomes. In fact, a two-step procedure most commonly used to estimate selection models follows this logic.⁷

⁷Step 1 estimates the probit model Eqs. (3)-(4) to yield estimates of γ , say $\hat{\gamma}$, and hence the private information function $\lambda_C(Z_i\hat{\gamma})$. In step 2, we substitute the estimated private information in lieu of its true value in Eq. (10) and estimate it by OLS. Standard errors must be corrected for the fact that γ is estimated in the second step, along the lines of Heckman (1979), Greene (1981), and Murphy and Topel (1985).

Our main purpose of discussing the Heckman model is to highlight the dual nature of self-selection “corrections.” One can think of them as a way of accounting for a statistical problem. There is nothing wrong with this view. Alternatively, one can interpret self-selection models as a way of testing private information hypotheses, which is perhaps an economically more useful perspective of selection models in corporate finance. Selection models are clearly useful if private information is one’s primary focus, but even if not, the models are useful as means of controlling for potential private information effects.

2.3 Specification issues

Implementing selection models in practice poses two key specification issues: the need for exclusion restrictions and the assumption that error terms are bivariate normal. While seemingly innocuous, these issues, particularly the exclusion question, are often important in empirical applications, and deserve some comment.

2.3.1 Exclusion restrictions

In estimating Eqs. (3)-(5), researchers must specify two sets of variables: those determining selection (Z) and those determining the outcomes (X). An issue that comes up frequently is whether the two sets of variables can be identical. This knotty issue often crops up in practice. For instance, consider the self-selection event E in Eqs. (3)-(4) as the decision to acquire a target and suppose that the outcome variable in Eq. (5) is post-diversification productivity. Variables such as firm size or the relatedness of the acquirer and the target could explain the acquisition decision. The same variables could also plausibly explain the ex-post productivity gains from the acquisition. Thus, these variables could be part of both Z and X in Eqs. (3)-(5). Similar arguments can be made for several other explanatory variables: they drive firms’ decision to self-select into diversification and the productivity gains after diversification. Do we need exclusion restrictions so that there is at least one variable driving selection, an instrument in Z that is not part of X ?

Strictly speaking, exclusion restrictions are not necessary in the Heckman selection model because the model is identified by non-linearity. The selection-adjusted outcome regression Eq. (10) regresses Y on X and $\lambda_C(Z'\gamma)$. If $\lambda_C(\cdot)$ were a linear function of Z , we would clearly need some

variables in Z that are not part of X or the regressors would be collinear.⁸ However, under the assumption of bivariate normal errors, $\lambda_C(\cdot)$ is a non-linear function. As Heckman and Navarro-Lozano (2004) note, collinearity between the outcome regression function (here and usually the linear function $X_i\beta$) and the selection “control” function $\lambda_C(\cdot)$ is not a generic feature, so some degree of non-linearity will probably allow the specification to be estimated even when there are no exclusion restrictions.

In practice, the identification issue is less clear cut. The problem is that while $\lambda_C(\cdot)$ is a non-linear function, it is roughly linear in parts of its domain. Hence, it is entirely possible that $\lambda_C(Z'\gamma)$ has very little variation relative to the remaining variables in Eq. (10), i.e., X . This issue can clearly arise when the selection variables Z and outcome variables X are identical. However, it is important to realize that merely having extra instruments in Z may not solve the problem. The quality of the instruments also matters. Near-multicollinearity could still arise when the extra instruments in Z are weak and have limited explanatory power.

What should one do if there appears to be a multicollinearity issue? It is tempting to recommend that the researcher impose additional exclusion restrictions so that self-selection instruments Z contain unique variables not spanned by outcome variables X . Matters are, of course, a little more delicate. Either the exclusions make sense, in which case these should have been imposed in the first place. Alternatively, the restrictions are not reasonable, in which case it hardly makes sense to force them on a model merely to make it estimable. In any event, as a practical matter, it seems reasonable to always run diagnostics for multicollinearity while estimating selection models whether one imposes exclusion restrictions or not.

The data often offer one degree of freedom that can be used to work around particularly thorny cases of collinearity. Recall that the identification issue arises mainly because of the 1/0 nature of the selection variable W_i , which implies that we do not observe the error term η_i and we must take its expectation, which is the inverse Mills ratio term. However, if we could observe the *magnitude* of the selection variable W_i , we would introduce an independent source of variation in the selection correction term and in effect observe the private information η_i itself and use it in the regression

⁸In this case, having a variable in X that is not part of Z does not help matters. If $\lambda_C(\cdot)$ is indeed linear, it is spanned by X whenever Z is spanned by X . Thus, we require extra variables that explain the decision to self-select but are unrelated to the outcomes following self-selection.

in lieu of the inverse Mills ratio. Exclusion restrictions are no longer needed. This is often more than just a theoretical possibility. For instance, in analyzing a sample of firms that have received a bank loan, we do observe the bank loan amount conditional on a loan being made. Likewise, in analyzing equity offerings, we observe the fact that a firm made an equity offering and also the size of the offer. In hedging, we do observe (an estimate of) the extent of hedging given that a firm has hedged. This introduces an independent source of variation into the private information variable, freeing one from the reliance on non-linearity for identification.⁹

2.3.2 Bivariate normality

A second specification issue is that the baseline Heckman model assumes that errors are bivariate normality. In principle, deviations from normality could introduce biases in selection models, and these could sometimes be serious (for an early illustration, see Goldberger (1983)). If non-normality is an issue, one alternative is to assume some specific non-normal distribution (Lee (1982), and Maddala (1983), Chapter 9.3). The problem is that theory rarely specifies that a particular alternative distribution should be used. Thus, whether one uses a non-normal distribution and the type of the distribution should be used are often driven by empirical features of the data. One approach that works around the need to specify parametric structures is to use semi-parametric methods (e.g., Newey, Powell and Walker (1990)). Here, exclusion restrictions are necessary for identification.

Finance applications of non-normal selection models remain scarce, so it is hard at this point of time to say whether non-normality is a first order issue deserving particular attention in finance. In one application to calls of convertible bonds (Scruggs (2005)), the data were found to be non-normal, but non-normality made little difference to the major conclusions.

3 Extensions

We review two extensions of the baseline Heckman self-selection model, switching regressions and structural selection models. The first allows some generality in specifying regression coefficients

⁹Estimation proceeds as usual with one notable exception. We include the magnitude of the private information η_i rather than its conditional expectation represented by the inverse Mills ratio term $\lambda_C(\cdot)$.

across alternatives, while the second allows bidirectional simultaneity between self-selection and post-selection outcomes.¹⁰ While each of these extensions generalizes the Heckman model by allowing some flexibility in specification. However, it should be emphasized that the additional flexibility that is gained does not come for free. The price is that the alternative approaches place additional demands on the data or require more stringent economic assumptions about the behavior of firms. The plausibility and feasibility of these extra requirements should be carefully considered before selecting any alternative to the Heckman model for a given empirical application.

3.1 Switching regressions

As in Section 2, a probit model based on exogenous variables drives firms' self-selection decisions. The difference is that the outcome is now specified separately for firms selecting E and NE , so the single outcome regression Eq. (5) in system Eqs. (3)-(5) is now replaced by two regressions. The complete model is as follows

$$C = E \equiv Z_i\gamma + \eta_i > 0 \quad (11)$$

$$C = NE \equiv Z_i\gamma + \eta_i \leq 0 \quad (12)$$

$$Y_{E,i} = X_{E,i}\beta_E + \epsilon_{E,i} \quad (13)$$

$$Y_{NE,i} = X_{NE,i}\beta_{NE} + \epsilon_{NE,i} \quad (14)$$

where $C \in \{E, NE\}$. Along with separate outcome regression parameter vectors β_E and β_{NE} , there are also two covariance coefficients for the impact of private information on outcomes, the covariance between private information η and ϵ_E and that between η and ϵ_{NE} . Two-step estimation is again straightforward, and is usually implemented assuming that the errors $\{\eta_i, \epsilon_{E,i}, \epsilon_{NE,i}\}$ are trivariate normal.¹¹

¹⁰For instance, in modeling corporate diversification as a decision involving self-selection, structural models would allow self-selection to determine post-diversification productivity changes, as in the standard setup, but also allow anticipated productivity changes to impact the self-selection decision.

¹¹Write Eqs. (13)-(14) in regression form as

$$Y_{C,i} = X_{C,i}\beta_C + \pi_C\lambda_C(Z_i\gamma) \quad (15)$$

where $C \in \{E, NE\}$. The two-step estimator follows: the probit model Eqs. (11)-(12) gives estimates of γ and hence the inverse Mills ratio $\lambda_C(\cdot)$, which is fed into regression Eq. (15) to give parameters $\{\beta_E, \beta_{NE}, \pi_E, \pi_{NE}\}$.

Given the apparent flexibility in specifying two outcome regressions, Eqs. (13)-(14) compared to the one outcome regression in the standard selection model, it is natural to ask why we do not always use the switching regression specification. There are three issues involved. First, theory should say whether there is a single population regression whose LHS and RHS variables are observed conditional on selection, as in the Heckman model, or whether we have two regimes in the population and the selection mechanism dictates which of the two we observe. In some applications, the switching regression is inappropriate: for instance, it is not consistent with the equilibrium modeled in Acharya (1988). A second issue is that the switching regression model requires us to observe outcomes of firms' choices in both regimes. This may not always be feasible because we only observe outcomes of firms self-selecting E but have little data on firms that choose not to self-select. For instance, if we were analyzing stock market responses to merger announcements as in Eckbo, Maksimovic and Williams (1990), implementing switching models literally requires us to obtain a sample of would-be acquirers that had never announced to the market and the market reaction on the dates that the markets realize that there is no merger forthcoming. These data may not always be available (Prabhala (1997)).¹² A final consideration is statistical power: imposing restrictions such as equality of coefficients $\{\beta, \pi\}$ for E and NE firms (when valid), lead to greater statistical power.

A key advantage of the switching regression framework is that we obtain more useful estimates of (unobserved) counterfactual outcomes. Specifically, if firm i chooses E , we observe outcome $Y_{E,i}$. However, we can ask what the outcome might have been had firm i chosen NE , the unobserved counterfactual, and what the gain is from firm i 's having made choice E rather than NE . The switching regression framework provides an estimate. The net benefit from choosing E is the outcome of choosing E less the counterfactual had it chosen NE , i.e., $Y_{E,i} - Y_{NE,i} = Y_{E,i} - X_i\beta_{NE} - \pi_{NE}\lambda_{NE}(Z_i\gamma)$. The *expected* gain for firm i is $X_i(\beta_E - \beta_{NE}) + (\pi_E\lambda_E(.) - \pi_{NE}\lambda_{NE}(.))$.¹³ We return

As before, standard errors in the second step regression require adjustment because $\lambda_C(Z\hat{\gamma})$ is a generated regressor (Maddala (1983), pp. 226-227).

¹²Li and McNally (2004) and Scruggs (2004) describe how we can use Bayesian methods to update priors on counterfactuals. More details on their approach are given in Section 6.

¹³This expression stands in contrast to the basic Heckman setup. There, in Eq. (9), $\beta_E = \beta_{NE}$ and $\pi_E = \pi_{NE}$, so the expected difference is $\pi(\lambda_E(.) - \lambda_{NE}(.))$. There, the sign of the expected difference is fixed: it *must* equal to the sign of π because $(\lambda_E(.) - \lambda_{NE}(.)) > 0$. Additionally, the expected difference in the setup of Section 2 does not vary

to the counterfactuals issue when we deal with treatment effects and propensity scores. We make this point at this stage only to emphasize that selection models do estimate treatment effects. This fact is often not apparent when reading empirical applications, especially those employing matching methods.

3.2 Simultaneity in self-selection models

The models considered thus far presume that the variables Z explaining the self-selection decision (Eqs. (3)-(4) or Eqs. (11)-(12)) are exogenous. In particular, the bite of this assumption is to preclude the possibility that the decision to self-select choice C does not directly depend on the anticipated outcome from choosing C . This assumption is sometimes too strong in corporate finance applications. For instance, suppose we are interested in studying the diversification decision and that the outcome variable to be studied is firm productivity. The preceding models would assume that post-merger productivity does *not* influence the decision to diversify. If firms' decisions to diversify depend on their anticipated productivity changes, as theory might suggest (Maksimovic and Phillips (2002)), the assumption that Z is exogenous is incorrect.

The dependence of the decision to self-select on outcomes and the dependence of outcomes on the self-selection decision is essentially a problem of simultaneity. Structural selection models can account for simultaneity. We review two modeling choices. The Roy (1951) model places few demands on the data but it places tighter restrictions on the mechanism by which self-selection occurs. More elaborate models are less stringent on the self-selection mechanism, but they demand more of the data, specifically exogenous instruments, exactly as in conventional simultaneous equations models.

3.2.1 The Roy model

The Roy model hard-wires the dependence of self-selection on post-selection outcomes. Firms self-select E or NE depending on which of the two alternatives yields the higher outcome. Thus, $\{E, Y_E\}$ is observed for firm i if $Y_{E,i} > Y_{NE,i}$. If, on the other hand, $Y_{NE,i} \geq Y_{E,i}$, we observe $\{NE,$

with β or variables X that are not part of Z : here, it does. In short, the counterfactual choices that could be made but were not are less constrained in the switching regression setup.

$Y_{NE,i}$ }. The full model is

$$C = E \equiv Y_{E,i} > Y_{NE,i} \quad (16)$$

$$C = NE \equiv Y_{E,i} \leq Y_{NE,i} \quad (17)$$

$$Y_{E,i} = X_i\beta_E + \epsilon_{E,i} \quad (18)$$

$$Y_{NE,i} = X_i\beta_{NE} + \epsilon_{NE,i} \quad (19)$$

where the ϵ 's are (as usual) assumed to be bivariate normal. The Roy model is no more demanding of the data than standard selection models. Two-step estimation is again fairly straightforward (Maddala (1983), Chapter 9.1).

The Roy selection mechanism is rather tightly specified on two dimensions. One, the model exogenously imposes the restriction that firms selecting E would experience worse outcomes had they chosen NE and vice versa. This is often plausible. However, it is unclear whether this should be a hypothesis that one wants to test or a restriction that one imposes on the data. Two, the outcome differential is the *only* driver of the self-selection decision in the Roy setup. Additional flexibility can be introduced by loosening the model of self-selection. This extra flexibility is allowed in models to be described next, but it comes at the price of requiring additional exclusion restrictions for model identification.

3.2.2 Structural self-selection models

In the standard Heckman and switching regression models, the explanatory variables in the selection equation are exogenous. At the other end of the spectrum is the Roy model of Section 3.2.1, in which self-selection is driven solely by the endogenous variable. The interim case is one where selection is driven by both exogenous and outcome variables. This specification is

$$C = E \equiv Z_i\gamma + \delta(Y_{E,i} - Y_{NE,i}) + \eta_i > 0 \quad (20)$$

$$C = NE \equiv Z_i\gamma + \delta(Y_{E,i} - Y_{NE,i}) + \eta_i \leq 0 \quad (21)$$

$$Y_{E,i} = X_i\beta_E + \epsilon_{E,i} \quad (22)$$

$$Y_{NE,i} = X_i\beta_{NE} + \epsilon_{NE,i} \quad (23)$$

The structural model generalizes the switching regression model of Section 3.1, by incorporating the extra explanatory variable $Y_{E,i} - Y_{NE,i}$, the net outcome gain from choosing E over NE , in

the selection decision, and generalizes the Roy model by permitting exogenous variables Z_i to enter the selection equation. Estimation of the system Eqs. (20)-(23) follows the route one typically treads in simultaneous equations systems estimation—reduced form estimation followed by a step in which we replace the dependent variables appearing in the RHS by their fitted projections. A trivariate normal assumption is standard (Maddala (1983), pp. 223-239). While structural self-selection models have been around for a while in the labor economics literature, particularly those studying unionism and the returns to education (see Maddala (1983), Chapter 8), applications in finance are of very recent origin.

The structural self-selection model clearly generalizes every type of selection model considered before. The question is why one should not always use it. Equivalently, what additional restrictions or demands does it place on the data? Because it is a type of the switching regression model, it comes with all the baggage and informational requirements of the switching regression. As in simultaneous equations systems, instruments must be specified to identify the model. In the diversification example at the beginning of this section, the identification requirement demands that we have at least one instrument that determines whether a firm diversifies but does not determine the ex-post productivity of the diversifying firm. The quality of one’s estimates depends on the strength of the instrument, and all the caveats and discussion of Section 2.3.1 apply here.

4 Matching models and self-selection

This section reviews matching models, primarily those based on propensity scores. Matching models are becoming increasingly common in applied work. They represent an attractive means of inference because they are simple to implement and yield readily interpretable estimates of “treatment effects.” However, matching models are based on fundamentally different set of assumptions relative to selection models. Matching models assume that unobserved private information is irrelevant to outcomes. In contrast, unobserved private information is the essence of self-selection models. We discuss these differences between selection and matching models as well as specific techniques used to implement matching models.

To clarify the issues, consider the switching regression selection model of Section 3.1, but relabel the choices to be consistent with the matching literature. Accordingly, firms are *treated* and belong

to group E or *untreated* and belong to group NE . This assignment occurs according to the probit model

$$pr(E|Z) = pr(Z\gamma + \eta) > 0 \quad (24)$$

where Z denotes explanatory variables, γ is a vector of parameters and we drop firm subscript i for notational convenience. The probability of being untreated is $1 - pr(E|Z)$. We write post-selection outcomes as Y_E for treated firms and Y_{NE} for untreated firms, and for convenience, write

$$Y_E = X_E\beta_E + \epsilon_E \quad (25)$$

$$Y_{NE} = X_{NE}\beta_{NE} + \epsilon_{NE} \quad (26)$$

where (again suppressing subscript i) ϵ_C denotes error terms, X_C denotes explanatory variables, β_C denotes parameter vectors, and $C \in \{E, NE\}$. We emphasize that the basic setup is identical to that of a switching regression of Section 3.1.

4.1 Treatment effects

Matching models focus on estimating *treatment effects*. A treatment effect, loosely speaking, is the value added or the difference in outcome when a firm undergoes treatment E relative to not undergoing treatment, i.e., choosing NE . To put matters in perspective, note that selection models such as the switching regression specification (Eqs. (11)-(14)) focus on estimating a vector of parameters and covariances for specified regressions. These parameters can be used to estimate treatment effects in selection models. Matching models go directly to treatment effect estimation, thereby circumventing the step of estimating particular regression structures specified in selection models.

The key question in the matching literature is whether treatment effects are significant. In specifications Eqs. (24)-(26), this question can be posed statistically in a number of ways.

- At the level of any individual firm i , the effectiveness of a treatment can be judged by asking whether $E(Y_{E,i} - Y_{NE,i}) = 0$.
- The effectiveness of the treatment for treated firms can be judged by asking whether the *treatment effect on treated*, (TT), equals zero, i.e., whether $E[(Y_E - Y_{NE})|C = E] = 0$.

- We can also test whether the *average treatment effect* (ATE) on the population as a whole, whether treated or not, is different from zero by examining whether $E(Y_E - Y_{NE}) = 0$.

The main issue in calculating any of the treatment effects discussed above, whether by selection or matching models, is the fact that unchosen counterfactuals are not observed. If a firm i chooses E , we observe outcome of its choice $Y_{E,i}$. However, because firm i chose E , we do not explicitly observe the outcome $Y_{NE,i}$ had it made the counterfactual choice NE . Thus, the difference $Y_{E,i} - Y_{NE,i}$ is never directly observed for any particular firm i , so its expectation—whether at the firm level, or across treated firms, or across treated and untreated firms—cannot be calculated directly. Treatment effects can, however, be obtained via selection models or by matching models, using different identifying assumptions. We discuss selection methods first and then turn to matching methods.

4.2 Treatment effects from selection models

Self-selection models obtain treatment effects by first estimating parameters of the system of equations Eqs. (24)-(26). Given the parameter estimates, it is straightforward to estimate treatment effects described in Section 4.1, as illustrated, e.g., in Section 3.1 for the switching regression model. The key identifying assumption in selection models is the specification of the variables entering selection and outcome equations, i.e., variables X and Z in Eqs. (24)-(26).

Two points deserve emphasis. The first is that the entire range of selection models discussed in Section 2 through Section 3.2 can be used to estimate treatment effects. This point deserves special mention because in received corporate finance applications, the tendency has been to report estimates of matching models and as a robustness check, an accompanying estimate of a selection model. With virtually no exception, the selection model chosen for the robustness exercise is the Heckman model of Section 2. However, there is no *a priori* reason to impose this restriction—any other model, including the switching regression models or the structural models, can be used, and perhaps ought to at least get a hearing. The second point worth mentioning is that regardless of the particular specification used, selection models always explicitly test for and incorporate the effect of unobservable private information, through the inverse Mills ratio term, or more generally, through *control functions* that model private information (Heckman and Navarro-Lozano (2004)).

4.3 Treatment effects from matching models

In contrast to selection models, matching models begin by assuming that private information is irrelevant to outcomes.¹⁴ Roughly speaking, this is equivalent to imposing zero correlation between private information η and outcome Y_E in Eqs. (24)-(26).

Is irrelevance of private information a reasonable assumption? It clearly depends on the specific application. The assumption is quite plausible if the decision to obtain treatment E is done through an exogenous randomization process. It becomes less plausible when the decision to choose E is an endogenous choice of the decision-maker, which is probably close to many corporate finance applications except perhaps for exogenous shocks such as regulatory changes.¹⁵ If private information can be ignored, matching methods offer two routes to estimate treatment effects: dimension-by-dimension matching and propensity score matching.

4.3.1 Dimension-by-dimension matching

If private information can be ignored, the differences in firms undergoing treatment E and untreated NE firms only depend on observable attributes X . Thus, the treatment effect for any firm i equals the difference between its outcome and the outcome for a firm $j(i)$ that matches it on all observable dimensions. Formally, the treatment effect equals $Y_{i,E} - Y_{j(i),NE}$, where $j(i)$ is such that $X_{i,k} = X_{j(i),k}$ for all K relevant dimensions, i.e., $\forall k, k = 1, 2, \dots, K$. Other measures such as TT and ATE defined in Section 4.1 follow immediately.¹⁶

Dimension-by-dimension matching methods have a long history of usage in empirical corporate finance, as explained in Chapter 1 (Kothari and Warner (2005)) in this book. Virtually all studies routinely match on size, industry, the book-to-market ratio, and so on. The “treatment effect” is

¹⁴See, e.g., Wooldridge (2002) for formal expressions of this condition.

¹⁵Of course, even here, if unobservable information guides company responses to such shocks, irrelevance of unobservables is still not a good assumption.

¹⁶One could legitimately ask why we need to match dimension by dimension when we have a regression structure such as Eqs. (25)-(26). The reason is that dimension-by-dimension matching is still consistent when the data come from the regressions, but dimension-by-dimension matching is also consistent with other data generating mechanisms. If one is willing to specify Eqs. (25)-(26), the treatment effect is immediately obtained as the difference between the fitted values in the two equations.

the matched-pair difference in outcomes. There is nothing inherently wrong with these methods. They involve the same economic assumptions as other matching methods based on propensity scores used in recent applications. In fact, dimension-by-dimension matching (such as those discussed in Chapter 1 for event-studies) imposes less structure and probably represents a first line of attack in empirical applications.

Matching on all dimensions and estimating the matched-pair differences in outcomes poses two difficulties. One is that characteristics are not always exactly matched in corporate finance applications. For instance, we often match firm size or book-to-market ratios with 30% calipers. When matches are inexact, substantial biases could build up as we traverse different characteristics being matched. A second issue that is commonly mentioned in applications is dimensionality. When the number of dimensions to be matched goes up and the matching calipers become fine (e.g., size and prior performance matched within 5% rather than 30%, and 4-digit rather than 2-digit SIC matches), finding matches becomes difficult or even infeasible. When dimension-by-dimension matching is not feasible, a convenient alternative is methods based on propensity scores. We turn to these next.

4.3.2 Propensity score (PS) matching

Propensity score (PS) matching methods handle the problems caused by dimension-by-dimension matching by reducing it to a problem of matching on a single one: the probability of undergoing treatment E . The probability of treatment is called the *propensity score*. Given a probability model such as Eq. (24), the treatment effect equals the outcome for the treated firm minus the outcome for an untreated firm with equal treatment probability. The simplicity of the estimator and its straightforward interpretation makes the propensity score estimator attractive.

It is useful to review the key assumptions underlying the propensity score method. Following Rosenbaum and Rubin (1983), suppose that the probability model in Eq. (24) satisfies

- **PS1:** $0 < pr(E|Z) < 1$.
- **PS2:** Given Z , outcomes Y_E, Y_{NE} do not depend on whether the firm is in group E (NE).

Assumption (PS1) requires that at each level of the explanatory variable Z , some firms should pick E and others pick NE . This constraint is frequently imposed in empirical applications by requiring that treated and untreated firms have common support.

Assumption (PS2) is the *strong ignorability* or conditional independence condition. It requires that unobserved private information should not explain outcome differentials between firms choosing E and those choosing NE . This is a crucial assumption. As Heckman and Navarro-Lozano (2004) show, even fairly mild departures can trigger substantial biases in treatment effect estimates.

Given assumptions (PS1) and (PS2), Rosenbaum and Rubin (1983) show that the treatment effect is the difference between outcomes of treated and untreated firms having identical treatment probabilities (or propensity scores). Averaging across different treatment probabilities gives the average treatment effect across the population.¹⁷

4.3.3 Implementation of PS methods

In light of Rosenbaum and Rubin (1983), the treatment effect is the difference between outcomes of treated and untreated firms with identical propensity scores. One issue in implementing matching is that we need to know propensity scores, i.e., the treatment probability $pr(E|Z)$. This quantity is not ex-ante known but it must be estimated from the data, using, for instance, probit, logit, or other less parametrically specified approaches. The corresponding treatment effects are also estimated with error and the literature develops standard error estimates (e.g., Heckman, Ichimura and Todd (1997), Dehejia and Wahba (1999), or Wooldridge (2002, Chapter 18)).

A second implementation issue immediately follows. What variables do we include in estimating the probability of treatment? While self-selection models differentiate between variables determining outcomes and variables determining probability of being treated (X and Z , respectively, in Eqs. (24)-(26)), matching models make no such distinction. Roughly speaking, either a variable determines the treatment probability, in which case it should be used in estimating treatment probability, or it does not, in which case it should be randomly distributed across treated and untreated firms

¹⁷This discussion points to another distinction between PS and selection methods. The finest level to which PS methods can go is the propensity score or the probability of treatment. Because many firms can have the same propensity score, PS methods do not estimate treatment effects at the level of the individual firm, while selection methods can do so.

and is averaged out in computing treatment effects. Thus, for matching models, the prescription is to use all relevant variables in estimating propensity scores.¹⁸

A third issue is estimation error. In principle, matching demands that treated firms be compared to untreated firms with the same treatment probability. However, treatment probabilities must be estimated, so exact matching based on the true treatment probability is usually infeasible. A popular approach, following Dehejia and Wahba (1999), divides the data into several probability bins. The treatment effect is estimated as the average difference between the outcomes of E and NE firms within each bin. Heckman, Ichimura and Todd (1997) suggest taking the weighted average of untreated firms. Weights decline inversely in proportion to the distance between the treated and untreated firms. For statistical reasons, Abadie and Imbens (2004) suggest that the counterfactual outcomes should be estimated not as the actual outcomes for a matched untreated firm, but as the fitted value in a regression of outcomes on explanatory variables.¹⁹

5 Panel data with fixed effects

In self-selection models, the central issue is that unobserved attributes that lead firms to self-select could explain variation in outcomes. In panel data settings, we have multiple observations on the same firm over different periods. If the unobservable attributes are fixed over time, we can control for them by including firm fixed effects. Applications of fixed effect models in corporate finance include Himmelberg, Hubbard and Palia (1999), Palia (2001), Schoar (2002), Bertrand and Mullainathan (2003), and Çolak and Whited (2005). There are undoubtedly many more. One question is whether the use of such fixed effect models alleviates self-selection issues. Not necessarily, as we discuss next.

There are two main issues with using firm fixed effects to rule out unobservables. One is that the unobservables should be time invariant. When time invariant effects exist and ought to be

¹⁸This statement is not, of course, a recommendation to engage in data snooping. For instance, in fitting models to estimate propensity scores, using quality of fit as a model selection criterion leads to difficulties, as pointed out by Heckman and Navorro-Lozano (2004).

¹⁹The statistical properties of different estimators has been extensively discussed in the econometrics literature, most recently in a review issue devoted to the topic (Review of Economics and Statistics (2004) vol. 86, issue 1, *Symposium on the Econometrics of Matching*.)

controlled for, fixed effect models are effective. However, time invariance is unlikely to be an appropriate modeling choice for corporate events. Here, unobservables are not only time varying but also related to the event under consideration. Furthermore, unobservables often have a causal role in precipitating the corporate finance event being studied. For instance, in the framework of Maksimovic and Phillips (2002), firms diversify or focus because they receive an unobserved shock that alters the optimal scope of the firm. Thus, in studying conglomerate diversification or spinoffs, the central unobservable of importance is the scope-altering shock. It is time varying and it leads to the event of interest—diversification. Including time-invariant firm fixed effects does nothing to address such event-related unobservable shocks. This point also applies to the difference-in-difference methods related to fixed effects. They do not account for event-related self-selection. Such methods are just not designed to capture time-varying and event-related unobservables, which are, in contrast, the central focus of selection models.

A second issue with fixed effect models is statistical power. Models with fixed effects rely on time variation in RHS variables and LHS outcomes for a given firm. Thus, fixed effect models often have limited power when the underlying variables vary slowly over time. In this scenario, causal effects, if any, are primarily manifested in the cross-section rather than time series. Zhou (2001) presents an argument on these lines with an empirical application. Thus, it appears especially important to take a more careful look at the lack of power as an explanation for insignificant results when using fixed effects. It should also be pointed out that regression R^2 is an inadequate and potentially misleading diagnostic for power.²⁰

6 Bayesian self-selection models

Thus far, our discussion covered inference via classical statistical methods. An alternative approach towards estimating selection models involves Bayesian methods. These techniques often represent an elegant way of handling selection models that are computationally too burdensome to be prac-

²⁰Most cross-sectional studies in corporate finance with reasonable sample sizes report a modest R^2 when there are no fixed effects. However when one adds fixed effects, there is often an impressive improvement in R^2 (e.g., Campa and Kedia (2002), and Villalonga (2004) for interesting illustrations of this point). The high R^2 should not be misattributed to the explanatory power of the included variables, because they often arise due to the (ultimately unexplained) fixed effects.

tical for classical methods. We review the Bayesian approach briefly and illustrate their potential value by discussing a class of selection models based on Markov Chain Monte Carlo (MCMC) simulations (see Poirier (1995) for a more in-depth comparison between Bayesian and classical statistical inferences).

6.1 Bayesian methods

The Bayesian approach begins by specifying a prior distribution over parameters that must be estimated. The prior reflects the information known to the researcher without reference to the dataset on which the model is estimated. In time series context, a prior can be formed by looking at out of sample historical data. In most empirical corporate finance applications, which are cross-sectional in nature, researchers tend to be agnostic and use non-informative diffuse priors.

Denote the parameters to be estimated by θ and the prior beliefs about these parameters by the density $p(\theta)$. If the observed sample is y , the posterior density of θ given the sample can be written as

$$p(\theta|y) = \frac{P(y|\theta)p(\theta)}{p(y)} \quad (27)$$

where $p(y|\theta)$ denotes the likelihood function of the econometric model being estimated. Given the prior and the econometric model, Eq. (27) employs Bayes rule to generate the posterior distribution $p(\theta|y)$ about parameter θ . The posterior density $p(\theta|y)$ summarizes what one learns about θ after seeing the data. It is the central object of interest that Bayesian approaches wish to estimate.

A key difficulty in implementing Bayesian methods is the computation of the posterior. Except for a limited class of priors and models, posteriors do not have closed-form analytic expressions, which poses computational difficulties in implementing Bayesian models. However, recent advances in computational technology and more importantly, the advent of the Gibbs sampler and the Metropolis-Hastings algorithm, which are specific implementations of MCMC methods, simplify implementation of fairly complex Bayesian models. In some cases, it even provides a viable route for model estimation where classical methods prove to be computationally intractable. Chib and Greenberg (1996) and Koop (2003) provide more detailed discussions of these issues.

6.2 Bayesian methods for selection models

To illustrate the implementation of the Bayesian approach to selection models, consider the switching regression model of Section 3.1. For notational convenience, rewrite this model as the system of equations

$$I = 1_{Z_i\gamma + \eta_i > 0} \tag{28}$$

$$Y_{E,i} = X_{E,i}\beta_E + \epsilon_{E,i} \tag{29}$$

$$Y_{NE,i} = X_{NE,i}\beta_{NE} + \epsilon_{NE,i} \tag{30}$$

where $1_{\{\cdot\}}$ denotes the indicator function and the other notation follows that in Section 3.1. As before, assume that the errors are trivariate normal with the probit error variance in Eq. (28) normalized to unity.

The critical unobservability issue, as discussed in Section 4, is that if a firm self-selects E , we observe the outcome $Y_{E,i}$. However, we do not observe the counterfactual $Y_{NE,i}$ that would have occurred had firm i chosen NE instead of E . Following Tanner and Wong (1987), a Bayesian estimation approach generates counterfactuals by augmenting the observed data with simulated observations of the unobservables through a “data augmentation” step. When augmented data are generated in a manner consistent with the structure of the model, the distribution of the augmented data converges to the distribution of the observed data. The likelihood of both the observed and the augmented data can be used as a proxy for the likelihood of the observed data. Conditional on the observed and augmented data and given a prior on parameters γ, β and the error covariances, approximate posteriors for the model parameters can be obtained by using standard simulation methods. The additional uncertainty introduced by simulating unobserved data can then be integrated out (Gelfand and Smith (1990)) to obtain posteriors conditional on only the observed data.

Explicitly modeling the unobserved counterfactuals offers advantages in the context of selection models. The counterfactuals that are critical in estimating treatment effects are merely the augmented data that are anyway employed in Bayesian estimation. The augmented data also reveal deficiencies in the model that are not identified by simple tests for the existence of selectivity bias. In addition, one can obtain exact small sample distributions of parameter estimates that are particularly useful when sample sizes are small to moderate, such as self-selection involving

relatively infrequent events. Finally, we can impose parameter constraints without compromising estimation. In later sections, we review empirical applications that employ the Bayesian approach towards estimating counterfactuals (Li and McNally (2004), and Scruggs (2005)). We also illustrate an application to a matching problem (Sørensen (2003)) in which the tractability of the conditional distributions given subsets of parameters leads to computationally feasible estimators in a problem where conventional maximum likelihood estimators are relatively intractable.

II. Empirical Applications

This part reviews empirical applications of self-selection models in corporate finance. We limit our scope to papers in which self-selection is an important element of the econometric approach or substantive findings. We begin with applications in event-studies applied to conversion-forcing calls, horizontal mergers, and stock splits. Here, the specifications are related to but differ from standard selection models. We then review applications in security offerings and financial intermediation, where selection models are used to characterize how private information affects debt issue pricing. We then turn to the diversification discount literature, where a range of methods have been used to address self-selection issues. The remaining sections include a collection of empirical applications based on selection and propensity score based matching methods. A last section covers Bayesian techniques. As will be clear from the review, most applications are relatively recent, involve a reasonably broad spectrum of approaches. In most cases, estimates of self-selection models reveal the importance of unobserved private information in influencing corporate finance choices.

7 Event studies

Event studies are a staple of empirical corporate finance. Hundreds of studies routinely report the stock market reactions to announcements such as mergers, stock splits, dividend announcements, equity issues, etc. Evidence in these studies has been used as a basis for testing and generating a wealth of theories, policies, and regulations. Chapter 1 in this volume (Kothari and Warner (2005)) overviews the literature.

Self-selection entered the event-study literature relatively recently. Its main use has been as a tool to model private information revealed in events. The basic idea is that when firms announce events, they reveal some latent “private” information. If the private information has value, it should explain the announcement effects associated with an event. Selection models are convenient tools to model the information revelation process and estimate “conditional” announcement effects.

7.1 Conditional announcement effects: Acharya (1988)

Acharya (1988) introduces the self-selection theme to event-studies, using a version of the standard Heckman specification to model calls of convertible bonds. In Acharya’s model, firms decide whether

to call an outstanding convertible bond (event E) or not (event NE) according to a probit model, viz.,

$$E \quad \text{if} \quad W_i = Z_i\gamma + \eta_i > 0 \quad (31)$$

$$NE \quad \text{if} \quad W_i = Z_i\gamma + \eta_i \leq 0 \quad (32)$$

where Z denotes known observables and η , the probit error term, is private information. Ex-ante, private information has zero mean, but ex-post, once the firm has announced E or NE , markets update expectations. If the private information affects stock prices, the stock price reaction y to the firm's choice should be related to the updated value of private information. Assuming that (η, y) are bivariate normal with mean, variances, and correlation equal to $(0, 0, 1, \sigma_y^2, \rho)$, we can write

$$E(y|E) = \pi E(\eta_i | \eta_i > -Z_i'\gamma) = \pi \lambda_E(Z_i'\gamma) \quad (33)$$

where $\pi = \rho\sigma_\epsilon$ and $\lambda_E(Z_i'\gamma) = \pi\phi(Z_i'\gamma)/\Phi(Z_i'\gamma)$, the inverse Mills ratio. Eq. (33) gives the *conditional* announcement effect associated with event E . It is a specialized version of the Heckman (1979) model (e.g., Eq. (10)) in which there are no regressors other than the inverse Mills ratio.²¹

The empirical application in Acharya (1988) is conversion-forcing calls of convertible bonds (event E) while NE denotes the decision to delay forced conversion. Acharya finds that the coefficient π in Eq. (33) is statistically significant, suggesting that the markets do react to the private information revealed in the call. The coefficient is negative, consistent with the Harris and Raviv (1985) signaling model. A legitimate question is whether testing for the significance of unconditional announcement effects and running a linear regression on characteristics Z could yield inferences equivalent to those from Acharya's model. Acharya (1993) offers simulation evidence and the question is formally analyzed in Prabhala (1997). The major conclusion appears to be that the simpler methods are often reasonable substitutes for Eq. (33). However, the self-selection model does add value if one has samples of firms that chose *not* to announce E . Self-selection methods offer a natural way of exploiting the information in samples of silent non-announcers.

²¹The absence of other regressors is dictated by the condition that announcement effects should not be related to ex-ante variables under the efficient markets hypothesis.

7.2 Two announcements on the same date: Nayak and Prabhala (2001)

In the Acharya model, there is one announcement on an event-date. Nayak and Prabhala (2001) analyze a specification in which two announcements are made on the same date. They present a model to recover the individual impact of each announcement from the observed announcement effects, which reflect the combined impact of both announcements made on one date.

The empirical application in Nayak and Prabhala is to stock splits, 80% of which are announced jointly with dividends. Nayak and Prabhala model the joint decisions about whether to split a stock and whether to increase dividends using a bivariate probit model, which can be specified as

$$SPL_i = \gamma_s Z_{si} + \psi_{si} \quad (34)$$

$$DIV_i = \gamma_d Z_{di} + \psi_{di} \quad (35)$$

If SPL_i exceeds zero, a firm splits, and if DIV_i exceeds zero, it increases dividends. The private information components of these two latent variables are ψ_{si} and ψ_{di} , and these have potentially non-zero correlation ρ_{sd} . The announcement effect from the two decisions is

$$E(AR_{sdi}) = \gamma_{sd} + \beta_d E(\psi_{di} | C, S) + \beta_s E(\psi_{si} | C, S) \quad (36)$$

The question of substantive interest is to decompose the joint split-dividend announcement effect into a portion due to the dividend information implicit in a split and the portion unrelated to the dividend information in the split. This decomposition cannot be inferred directly from Eq. (36) because the term relating to splits ($\beta_d E(\psi_{di} | C, S)$) incorporates both the dividend and the non-dividend portion of the information in splits. However, this decomposition is facilitated by writing the split information ψ_{si} into dividend and non-dividend components. Accordingly, write $\psi_{si} = \rho_{sd}\psi_{di} + \psi_{s-d,i}$, in which case the joint announcement effect is

$$E(AR_{sdi}|C, S) = \gamma_{sd} + (\alpha_d - \rho_{sd}\alpha_{s-d})E(\psi_{di} | C, S) + \alpha_{s-d}E(\psi_{si} | C, S) \quad (37)$$

where α_d and α_{s-d} denote the reaction to the dividend and pure split components of the information in splits. Given these, Nayak and Prabhala show that the market's reaction to a hypothetical "pure" split unaccompanied by a dividend is

$$E(AR_{si}) = (1 - \rho_{sd}^2)\alpha_{s-d}\psi_{si} + \rho_{sd}\alpha_d\psi_{si} \quad (38)$$

The first component in Eq. (38) represents the market’s reaction to pure split information orthogonal to dividends and the second represents the reaction to the dividend information implied by a split. Estimating the model is carried out using a two-step procedure.²² Using a sample of splits made between 1975 and 1994 divided into two sub-samples of ten years each, Nayak and Prabhala report that about 46% of split announcement effects are due to information unrelated to the dividend information in splits.

The Nayak and Prabhala analysis has interesting implications for sample selection in event studies. In many cases, an event is announced together with secondary information releases. For instance, capital expenditure, management, or compensation announcements may be made together with earnings releases, creating noisy samples. The conventional remedy for this problem is to pick samples in which the primary announcement of interest is not accompanied by a secondary announcement by firms. However, the analysis in Nayak and Prabhala suggests that this remedy may not cure the ill, since markets can form expectations about and price secondary announcements even when they are not explicitly announced on the event date. A different approach is to model both announcements and extract the information content of each. Selection methods are useful tools in this regard because they explicitly model and incorporate the latent information from multiple announcements.

7.3 Takeovers: Eckbo, Maksimovic and Williams (1990)

Eckbo, Maksimovic and Williams (1990)—henceforth EMW—propose variants of the “truncated regression” specification, rather than the Heckman selection model used in Acharya (1988) model to analyze announcement effects. Their empirical application is to takeovers, a subject reviewed extensively in Chapter 17 in this volume (Betton, Eckbo and Thorburn (2005)).

EMW develop two models for announcement effects. In both models, managers announce event E if the stock market gain, $y_i = x_i\gamma + \eta_i$ is positive. As before, η_i is private information, normally distributed with mean zero and variance ω^2 and x_i denotes publicly known variables. In model 1,

²²The parameter ρ_{sd} is obtained as the correlation coefficient in the bivariate probit model Eqs. (34)-(35). The inverse Mills ratios for Eq. (37) follow (they require modification from standard expressions to incorporate non-zero correlation between bivariate latent variables). The other coefficients can be estimated from regression Eq. (37).

event E completely surprises the capital markets. In this case, the bidder's announcement effect is

$$\begin{aligned} F(x_i) &= E(y_i | y_i = x_i\gamma + \eta_i > 0) \\ &= x_i\gamma + \omega \frac{\phi(x_i\gamma/\omega)}{\Phi(x_i\gamma/\omega)} \end{aligned} \quad (39)$$

In model 2, the market learns about the impending takeover on a prior rumor date. The probability that the takeover will be announced is the probability that the takeover gain is positive, i.e., $Pr(x_i\gamma + \eta_i > 0) = \Phi(x_i\gamma/\omega)$. If the takeover occurs, the gain is $F(x_i)$, while the absence of the takeover is assumed to lead to zero gain. Thus, the expected stock return on the rumor date is $F(x_i)\Phi(x_i\gamma/\omega)$. On the actual merger announcement date, the takeover probability rises to 1 and the announcement effect is

$$G(x_i) = [x_i\gamma + \omega \frac{\phi(x_i\gamma/\omega)}{\Phi(x_i\gamma/\omega)}][1 - \Phi(x_i\gamma/\omega)] \quad (40)$$

The EMW expression in Eq. (40) is different from the Acharya model because EMW assume that private information has value only conditional on the takeover E , but has no value if there is no takeover. Thus, EMW model the real gains specific to mergers rather than non-specific information modeled by Acharya. In the actual empirical application, EMW find that bidder gains decrease with the size of the bidder relative to the target, the concentration of firms in the industry, and the number of previous takeovers in the industry. As a model diagnostic, they show that OLS estimates differ from those of the non-linear model Eq. (40), which is supported by the Vuong (1989) test statistics. EMW also report that ω^2 is significant, indicating that bidders' private information is valued by capital markets.

The EMW framework has been widely applied in other event-studies with cross-sectional regressions. Eckbo (1990) examines the valuation effects of greenmail prohibitions and finds that the precommitment not to pay greenmail is value enhancing. Maksimovic and Unal (1993) estimate the after-market price performance of public offers in thrift conversions recognizing that management's choice of issue size reflects the value of growth opportunities and conflicts of interest between managers and investors. Servaes (1994) relates takeover announcement effects to excess capital expenditure. Hubbard and Palia (1995) find an increasing and then decreasing relation between merger announcement effects and managerial ownership levels. Bohren, Eckbo and Michaelsen (1997) use it to explain why rights flotations are not favored over public offerings despite the greater direct costs of the latter. Li and McNally (2006) apply the EMW method to

open market share repurchases in Canada and find evidence supporting a signaling interpretation of repurchase announcement effects. We study one particular extension, Eckbo (1992), in greater detail next.

7.4 Takeover deterrence: Eckbo (1992)

Eckbo (1992) extends the EMW framework to account for the fact that regulatory challenges and court decisions on these could affect merger gains. To the extent these decisions also involved unobserved private information, they introduce additional selection bias terms into the final specification. Eckbo develops these models and applies them to horizontal mergers and price effects of rivals not involved in takeovers.

As in EMW, horizontal mergers occur if the acquirer's share of the synergy gains, $y_j = x_j\gamma + \eta_j > 0$. Under the EMW assumptions, the model for the announcement effects is Eq. (40). Additionally, regulators can choose whether to initiate anti-trust actions or not, and subsequently courts can decide whether to stop a merger or not. These actions are modeled using additional probit models.

$$R = x_i\phi_r + \eta_r > 0 \quad (41)$$

$$C = x_i\phi_c + \eta_c > 0 \quad (42)$$

Merger gains are realized if mergers are not challenged or they are challenged but challenges are unsuccessful. Assuming that challenges have a cost c proportional to merger gains, conditional announcement effects of merger announcements can be written as

$$E(AR_i|E) = [(1 - p_{ri}p_{ci})(x_i\gamma + \omega \frac{\phi(x_i\gamma/\omega)}{\Phi(x_i\gamma/\omega)} - p_{ri}c)[1 - \Phi(x_i\gamma/\omega)] \quad (43)$$

Eckbo applies the truncated regression models to U.S. and Canadian data. For Canadian data, Eckbo uses the EMW models Eqs. (39)-(40) because there is no regulatory overhang, but uses Eq. (43) to U.S. horizontal mergers where regulatory overhang exists. The explanatory variables include the ratio of the market values of the bidder and target firms, the number of non-merging rival firms in the industry of the horizontal merger, the pre-merger level of and merger-induced change in industry concentration. Eckbo finds that bidder gains are positively related to the pre-merger concentration ratio and are negatively related to the merger-induced changes in the concentration ratio. These do not support the collusion explanation for merger gains. In an interesting innovation, Eckbo

also estimates the models for non-merging rivals. He reports similar and even sharper findings in challenged deals where court documents identify rivals more precisely. Changes in concentration are negatively related to rival gains in the regulatory overhang free environment in Canada, further refuting the collusion hypothesis.

8 The pricing of public debt offerings

Companies making a debt issue must make several decisions related to the offering such as the terms and structure of the offering, the type of the underwriter for the issue. Private information held by the issuer or the intermediaries participating in the offering could affect the choices made by firms. If such information has value, it affects the prices at which issues can be sold. A fairly wide range of self-selection models have been used to address the existence of private information and its effect on the pricing of debt issues. We review some of the applications and the key findings that emerge.

8.1 Bank underwritings and the Glass-Steagall Act: Puri (1996)

The choice of an underwriter is an area that has been extensively analyzed using self-selection models. An early application is Puri (1996), who investigates the information in a firm's choice between commercial banks and investment banks as underwriters of public debt offerings. Commercial banks are often thought to possess private information about their borrowers. If they use the private information positively, commercial bank underwritten offerings should be priced higher (the "certification" hypothesis). Alternatively, banks could use their information negatively to palm off their lemons to the market, in which case the markets should discount commercial bank underwritten offerings (the "conflicts of interest" hypothesis). Selection models are natural avenues to examine the nature of these private information effects.

Puri models the private information in the underwriter choice using a probit model, viz.,

$$C = CB \equiv W_i = Z_i\gamma + \eta_i > 0 \tag{44}$$

$$C = IB \equiv W_i = Z_i\gamma + \eta_i \leq 0 \tag{45}$$

where CB denotes a commercial bank, IB denotes an investment bank, and η_i is the private information in offering i . Markets price issue i at yield y_i where

$$y_i = x_i\beta + \epsilon_i \quad (46)$$

$$E(y_i|C) = X_i\beta + \pi\lambda_C(Z_i\gamma) \quad (47)$$

Eq. (47) follows from Eq. (46) and the assumption that ϵ and η are bivariate normal. The above system is, of course, the standard Heckman model of Section 2, so the sign of the covariance coefficient π determines the impact of private information on offer yields. If $\pi > 0$, markets demand higher yield for CB offerings, consistent with a conflicts of interest hypothesis, while $\pi < 0$ supports the certification hypothesis.

The data in Puri (1996) are debt and preferred stock issues prior to the passage of the 1933 Glass-Steagall Act. She includes issue size, credit rating, syndicate size, whether the security is exchange listed, whether it is collateralized, and the age of the issuer as determinants of the offer yield. She finds that $\pi < 0$, consistent with the certification hypothesis. Additionally, π is more negative for information sensitive securities, where the conflicts of interest hypothesis predicts the more *positive* coefficient.²³ Gande, Puri, Saunders and Walter (1997) and Gande, Puri, and Saunders (1999) report similar findings for debt issues offered after the 1989 relaxation of the Glass-Steagall Act. Underwritings by commercial banks convey positive information that improves the prices at which debt offerings can be sold.²⁴

8.2 Underwriting syndicate structure: Song (2004)

Song (2004) analyzes debt offerings as in Puri (1996) and Gande et al. (1997, 1999) but there are some important differences in specifications. Song uses a switching regression instead of the Heckman model. Second, she focuses on the effect of the syndicate structure rather than the commercial/investment banking dichotomy on debt issue spreads.

In Song's model, commercial banks could enter as lead underwriters or be part of a hybrid syndicate with investment banks. Alternatively, issues could be underwritten by a pure investment

²³Of course, it is possible that investors paid more for bank underwritten issues but were fooled into doing so. Puri (1994) rules out this hypothesis by showing that bank underwritten offerings defaulted less than non-bank issues.

²⁴Chiappori and Salanie (2000) use similar methods to analyze the role of private information in insurance markets.

bank syndicate. For each outcome, we observe the yield of the debt offering, which is modeled as a function of public information and (implicitly) the private information conveyed in the firm's choice of a syndicate structure. The resulting specification is a variant of the switching regression model of Section 3.1, and can be written as

$$A_i = 1 \quad \text{if} \quad -Z_{Ai}\gamma_A + \eta_{Ai} > 0 \quad (48)$$

$$B_i = 1 \quad \text{if} \quad -Z_{Bi}\gamma_B + \eta_{Bi} > 0 \quad (49)$$

$$C_i = 1 \quad \text{if} \quad -Z_{Ci}\gamma_C + \eta_{Ci} > 0 \quad (50)$$

$$Y_{1i} = X_{1i}\beta_1 + \eta_{1i} \quad (51)$$

$$Y_{2i} = X_{2i}\beta_2 + \eta_{2i} \quad (52)$$

$$Y_{3i} = X_{3i}\beta_3 + \eta_{3i} \quad (53)$$

where we have adapted Song's notation for consistency with the rest of this chapter.²⁵ In Eqs. (48)-(50), the counterfactuals are $A = 0$, $B = 0$, and $C = 0$, respectively.

In Song's model $A_i = 1$ if a lead investment bank invites a commercial bank to participate in the syndicate. $B_i = 1$ if the commercial bank joins the syndicate, and zero otherwise. $C_i = 1$ if a commercial bank led syndicate is chosen and $C_i = 0$ if a pure investment bank syndicate is chosen. Thus, a hybrid syndicate is observed (regime 1) when $A_i = 1$ and $B_i = 1$; a pure investment bank syndicate (regime 2) is observed when $A_i = 0$ and $C_i = 0$, while a commercial bank led syndicate (regime 3) is observed when $B_i = 0$ and $C_i = 1$.²⁶ Song assumes that the latent errors η are i.i.d. normal, correlated with yields Y with regression coefficients $\sigma_{\omega j}$ where $\omega \in \{A, B, C\}$ and $j \in \{1,$

²⁵Song's usage of signs for coefficients and error terms illustrates some confusing notation in the limited dependent variable literature. Her notation is in the spirit of Maddala (1983) where the selection criterion is often written as $Z\gamma - \eta > 0$, while the more modern textbook convention is to use $Z\gamma + \eta > 0$.

²⁶Song does not explicitly write out the extensive form of the model she estimates. It is unclear whether pure investment bank syndicates should also include the node at which an investment bank is awarded the mandate and chooses to invite a commercial bank but the bank declines to join.

2, 3}. The yields in each regime can be expressed in regression form as

$$E(y_{1i}|A_i = 1, B_i = 1) = X_{1i}\beta_1 + \sigma_{A1} \frac{\phi(Z_{Ai}\gamma_A)}{1 - \Phi(Z_{Ai}\gamma_A)} + \sigma_{B1} \frac{\phi(Z_{Bi}\gamma_B)}{1 - \Phi(Z_{Bi}\gamma_B)} \quad (54)$$

$$E(y_{2i}|A_i = 0, C_i = 0) = X_{2i}\beta_2 - \sigma_{A2} \frac{\phi(Z_{Ai}\gamma_A)}{\Phi(Z_{Ai}\gamma_A)} - \sigma_{C2} \frac{\phi(Z_{Ci}\gamma_C)}{\Phi(Z_{Ci}\gamma_C)} \quad (55)$$

$$E(y_{3i}|B_i = 0, C_i = 1) = X_{3i}\beta_3 - \sigma_{B3} \frac{\phi(Z_{Bi}\gamma_B)}{\Phi(Z_{Bi}\gamma_B)} + \sigma_{C3} \frac{\phi(Z_{Ci}\gamma_C)}{1 - \Phi(Z_{Ci}\gamma_C)} \quad (56)$$

Song's sample comprises 2,345 bond issues offered between January 1991 and December 1996. In the first step probit estimates, Song reports that compared to pure investment bank syndicates, hybrid syndicates underwrite small firms that have made smaller debt issues in the past, have low S&P stock rankings, invest less, and use more bank debt. These findings are reminiscent of those in Puri and Gande et al. (1997, 1999) that commercial banks underwrite informationally sensitive companies. Compared to commercial bank led syndicates, hybrid syndicates underwrite smaller firms with lower stock rankings that issue to refinance debt and lower ranked firms, consistent with the claim that these underwritings potentially alleviate conflicts of interest with commercial banks. Only two out of six private information coefficients in Eqs. (54)-(56) are significant. Pricing benefits are seen in pure investment banking syndicates (Eq. (55)) where excluding a commercial bank leads to higher yields, consistent with a certification hypothesis. On the other hand, picking an investment bank to run the syndicate increases yields, because the coefficient σ_{C2} in the same equation Eq. (55) is positive. Thus, the ex-ante effect of awarding a syndicate to an investment bank cannot be *a priori* signed.

Relative to prior work, Song (2004) has very different sample, sample period, and explanatory variables, not to mention the changes in underwriter classification, which is based on syndicate structure rather than on classification into commercial/investment bank or on bank reputation. Thus, it is hard to pinpoint the specific value added by her elaborate selection model. In addition, absent additional diagnostics, it is also difficult to interpret whether the general insignificance of most selection terms reflects coefficients that are truly zero, the lack of power, perhaps due to collinearity, or perhaps an unmodeled correlation between errors in Eqs. (48)-(50) that are assumed to be i.i.d. for the purposes of estimation. As Song points out, additional data may not help shed light on interpretation or robustness because there have been structural changes in the banking industry since 1996, due to several mergers and further relaxation of the Glass-Steagall Act.

8.3 Underwriter reputation: Fang (2005)

Like the other papers reviewed in this section, Fang (2005) also studies the role of underwriter choice in explaining at-issue bond spreads. Unlike the other papers in the section, however, Fang draws on an earlier literature and classifies underwriters by reputation rather than by organization into commercial or investment banks. Fang examines whether the information in the choice of a reputed underwriter impacts underwriting spreads and yields.

Fang uses a probit specification to model underwriter-issuer matching. If issue i is underwritten by a reputed underwriter, the yield is $Y_{E,i}$ and if not, the yield is $Y_{NE,i}$. Yields are specified as a function of regressors x_i with different regression coefficients across the two choices. Thus, Fang's model is exactly the switching regression of Section 3.1. Fang also estimates an auxiliary regression where the dependent variable is gross spread rather than offer yield.

Fang finds that reputed underwriters underwrite higher grade, less risky issues of large and frequent issuers, and are more likely to be associated with longer maturity callable issues that she interprets as being more complex. The self-selection term in the yield equation is negative. Thus, the unobserved information that leads firms to choose reputed underwriters leads to lower bond yields or better offer prices. In the specification analyzing gross spreads, Fang finds that issue size increases fees more rapidly but risk variables matter less for reputed underwriters, indicating greater marginal costs and superior risk bearing capacity of reputed underwriters. Most importantly, the coefficient for the inverse Mills ratio in the gross spread equation is positive, suggesting that reputed underwriters charge greater fees to issuers.

Taken together, the yield and gross spread specifications show that reputed underwriters charge issuers greater fees and lower the offer yields (i.e., increase the offer price) to borrowers. Fang shows that the benefit of lowered debt yields typically outweighs the higher commissions paid by issuers. The pattern of results is shown to strengthen in lower yield bonds, so that reputation matters more for more informationally sensitive issues.

8.4 Debt covenants: Goyal (2005)

While the papers reviewed in this section study and model information in underwriter choice, Goyal (2005) examines the information in the choice of covenants attached to debt issues. Goyal argues that commercial banks often enjoy franchise value because of regulations that deter free

entry. Banks with more valuable franchises are less likely to engage in excessive risk taking, so they should have less need to include covenants in their debt issues. This incentive is recognized and priced by the market, and the pricing differential again feeds back into firms' decisions about whether to include covenants. In other words, the decision to include covenants influences and is influenced by the expected pricing benefits from doing so. Goyal implements the structural self-selection model of Section 3.2 to model the simultaneity.

Goyal estimates the structural model on a sample of 415 subordinated debt issues made by firms between 1975 and 1994. He finds that yields are negatively related to franchise value. This finding is consistent with the hypothesis that banks with greater franchise value have less incentives to take risk, latent information that is recognized and priced by financial markets. The inverse Mills ratio term is significant in the no-covenant sub-sample but not in the sample with restrictive covenants. In the equation explaining whether firms use covenants or not, the coefficient for the yield differential with/without covenants is significant in explaining covenant choice, suggesting that anticipated pricing benefits do influence whether firms select covenants or not in their debt issues. Many of Goyal's results are more prominent in the 1981-1988 sub-period, when the risk-taking activity in the U.S. was more elevated.²⁷

8.5 Discussion

The public debt issue pricing literature is interesting for its wide range of selection models employed. One issue, however, is that it is a little difficult to place the literature in perspective because the source of self-selection modeled is quite different across papers. An additional issue is, of course, that there is probably self-selection on other dimensions as well, such as maturity, collateral, or the callability of an issue, not speaking of the decision to issue debt in the first place. This raises another thorny question, one that probably has no easy answer. What dimensions of self-selection should one control for in a given empirical application? Modeling every source of selection seems infeasible, while studying some sources of bias while ignoring others also seems a little *ad-hoc*. Embarking on a purely empirical search for sources of selection that matter is certainly undesirable, smacking of data snooping. A happy middle way is likely to emerge as the literature matures.

²⁷Reisel (2004) provides an interesting extension, a structural self-selection model applied to debt covenants included in industrial bonds.

9 Other investment banking applications

9.1 Underwriter compensation in IPOs: Dunbar (1995)

Dunbar (1995) presents an interesting application of a Roy (1951) style self-selection model to the study of underwriter compensation. Some IPO issuers offer warrants to compensate their underwriters while other issuers do not. Dunbar examines the role of self-selection in explaining this choice, and in particular, whether firms choose the alternative that minimizes their underwriting costs.

Let W denote the decision to use warrants to compensate underwriters and N if not, subscripts w and n denote the costs if warrants are used or not, respectively, U denote underpricing costs and C the other costs of going public. If firm i chooses underwriter warrant compensation, we observe the pair $\{U_{wi}, C_{wi}\}$ while we observe $\{U_{ni}, C_{ni}\}$ if it chooses just straight cash compensation. The key self-selection issue is that we observe the choice made by firm i but not the costs of the alternative not chosen by firm i . Without knowing the unchosen counterfactuals, we cannot tell how much a company saved by choosing to include or exclude warrants to compensate its underwriters.

Dunbar models the decision to use warrants using a probit model

$$W = \xi(U_{ni} + C_{ni} - U_{wi} - C_{wi}) - \varepsilon_i > 0 \quad (57)$$

$$N = \xi(U_{ni} + C_{ni} - U_{wi} - C_{wi}) - \varepsilon_i \leq 0 \quad (58)$$

The expression in parentheses in Eq. (57) is the reduction in offering costs if warrants are used as compensation instead of straight cash compensation. Each component of costs is written as a function of observables and unobservables as follows.

$$U_{ni} = X_{ni}\beta_n + \varepsilon_{uni} \quad (59)$$

$$U_{wi} = X_{wi}\beta_w + \varepsilon_{uwi} \quad (60)$$

$$C_{ni} = Z_{ni}\gamma_n + \varepsilon_{cni} \quad (61)$$

$$C_{wi} = Z_{wi}\gamma_w + \varepsilon_{cwi} \quad (62)$$

Assuming that the errors in Eqs. (59)-(62) are i.i.d. normal but potentially correlated with the probit error term, Dunbar's system is a version of the Roy (1951) self-selection model.

Dunbar reports that variables such as offering size, underwriter reputation, and a hot issue dummy explain underpricing in the warrant and cash compensation samples. The self-selection

term is significant in the non-warrant sample but not in the warrant compensation sample. Most interesting are Dunbar’s estimates of unobserved counterfactuals. For firms that do not use warrants, underpricing (other costs) would be 11.6% (19.2%) on average had warrants been used compared to actual underpricing (other costs) of 12.8% (9.8%). For firms that do use warrants, underpricing (other costs) would be 36.4% (14.6%) if warrants had not been used, compared to actual costs of 23.3% (23.9%). While warrants are associated with high underpricing in reduced form cross-sectional regressions, it is incorrect to conclude that warrants result in higher underpricing. Estimates of the self-selection model indicates that the use of warrants actually *reduces* underpricing compared to what it would be without warrants. Firms appear to use warrants to reduce underpricing costs.

9.2 Analyst coverage: Ljungqvist, Marston and Wilhelm (2005)

Ljungqvist, Marston and Wilhelm (2005) examine the relation between the decision to award an underwriting mandate to a bank and the coverage offered by the bank’s analyst. The self-selection issue in Ljungqvist et al. is that banks self-select on whether they cover a stock or not. If the bank covers a stock, we observe the nature of the stock recommendation and we can tie it to the decision to award an underwriting mandate. However, if a bank does not elect to cover a stock, we do not know what the nature of its recommendation might have been had it chosen to cover the stock. Ljungqvist et al. model this source of self-selection in testing whether a firm with more positive coverage of a firm is more likely to win the firm’s underwriting mandates.

Ljungqvist et al. model the probability that bank j covers firm i ’s stock as a probit model

$$\begin{aligned} y_C &= 1 & \text{if } & y_C^* = X_C\beta_C + u_C > 0 \\ y_C &= 0 & \text{if } & y_C^* = X_C\beta_C + u_C \leq 0 \end{aligned} \tag{63}$$

where all subscripts are suppressed for notational convenience. If there is coverage, the tie between coverage and the award of an underwriting mandate is established by the equations

$$\left. \begin{aligned} y_A &= \beta_A X_A + u_A \\ y_L &= I_{\beta_L X_L + \delta_L y_A + u_L > 0} \end{aligned} \right\} \text{if } y_C^* > 0 \tag{64}$$

If there is no coverage, we have

$$\left. \begin{aligned} y_A &= 0 \\ y_L &= I_{\beta_{LNC}X_L + u_{LNC} > 0} \end{aligned} \right\} \text{ if } y_C^* \leq 0 \quad (65)$$

where y_A is the nature of an analyst's recommendation, y_L is a 1/0 dummy for whether an underwriting mandate is awarded to a bank, I is the 1/0 indicator function, and X 's are explanatory variables. Eqs. (63)-(65) represent a switching regression system, similar to the type analyzed in Section 3.1. The difference here is that we have two recursive equations observed in each regime instead of just one regression in Section 3.1.

Ljungqvist et al. find that the decision to cover a stock is positively related to the type of coverage offered by an analyst for debt underwriting transactions. Prior relationships in the underwriting and loan markets are the other most significant explanatory variables. There is no evidence that the type of coverage influences the decision to award equity underwriting mandates. Even when it is significant, the coefficient for analyst recommendation β_A in Eq. (64) is negative. Ljungqvist et al. interpret this finding as evidence that even if analysts are overly biased, issuers refrain from using them for underwriting.

The analysis in Ljungqvist et al. has appealing features. The choice of instruments is carefully motivated, with both economic intuition and tests for instrument strength suggested by Steiger and Stock (1997). Their analysis also suggests some natural extensions. One issue is that the very decision to cover a stock—rather than the type of coverage—might affect the probability of winning an underwriting mandate. A second and perhaps more difficult issue is that of cross-sectional correlation. The 16,000+ transactions in the Ljungqvist et al. sample occur over overlapping periods, which leads to commonality across transactions and potential cross-sectional correlation in the disturbance terms.

10 Diversification discount

The scope of the firm is an issue that has occupied economists since Coase (1933). One issue in this literature has been whether firms should diversify or not. While the question can be examined from several perspectives, a now well developed literature in finance investigates the diversification question from a valuation perspective. Does diversification impact firm value, and if so, in what

direction, and why does diversification have this effect? Our review of this literature focuses on self-selection explanations for diversification. Chapter 9 in this volume (Maksimovic and Phillips (2005)) provides a more complete review of the now vast literature on diversification.

The recent finance literature on diversification begins with the empirical observation that diversified firms trade below their imputed value, computed as the weighted average value of stand-alone firms in the same businesses as the divisions of the diversified firm (see, e.g., Lang and Stulz (1994), Berger and Ofek (1995), and Servaes (1996)). The difference between the actual and imputed values is called the diversification discount. The existence of a diversification discount is frequently interpreted as a value destroying consequence of diversification, although there is no consensus on the issue (e.g., Chevalier (2000), and Graham, Lemmon and Wolf (2002)). We review three papers that discuss the role of self-selection in explaining the source of diversification discount.

10.1 Unobservables and the diversification discount: Campa and Kedia (2002)

Campa and Kedia (2002) argue that firms self-select into becoming diversified and that self-selection explains the diversification discount. They model the decision to become diversified using a probit model

$$D_{it} = 1 \quad \text{if} \quad Z_{it}\gamma + \eta_{it} > 0 \quad (66)$$

$$D_{it} = 0 \quad \text{if} \quad Z_{it}\gamma + \eta_{it} \leq 0 \quad (67)$$

where D_{it} is a diversification dummy that takes the value of 1 if the firm operates in more than one segment, and 0 otherwise, and Z_{it} is a set of explanatory variables. The notations are adapted to match that in Section 2. Excess value V_{it} is specified as

$$V_{it} = d_0 + d_1 X_{it} + d_2 D_{it} + \epsilon_{it} \quad (68)$$

where X_{it} is a set of exogenous observable characteristics of firm i at time t . Coefficient d_2 is the key parameter of interest. If it is negative, becoming diversified causes the diversification discount. If not, the diversification discount could not be due to diversification. Under the assumption that

the error terms in Eqs. (67)-(68) are bivariate normal, the system is akin to and is estimated just like the basic Heckman selection model.²⁸

In the empirical application, Campa and Kedia measure the LHS variable in Eq. (68), as the difference between the actual value of the firm and the sum of the imputed value of each of its segments. Segment imputed values are estimated using multipliers based on market value to sales or market value to book value of assets of peer firms. The explanatory variables for Eq. (68) include profitability, size, capital expenditure, and leverage. The additional instruments used in the probit specification Eqs. (66)-(67) include industry instruments such as the fraction of firms (or their sales) in an industry that are diversified, time instruments, macroeconomic indicators such as the overall economic growth and economic expansion/contraction, and firm specific instruments such as being listed on a major exchange or being included in a stock index. Campa and Kedia extensively discuss their choices for instruments.

Campa and Kedia show that in OLS specifications, d_2 is negative, so that diversified firms do appear to trade at a discount. However, once they include the inverse Mills ratio to correct for self-selection, the coefficient d_2 becomes positive. The negative sign seen in OLS estimates is soaked up by the coefficient for the inverse Mills ratio. This indicates that diversified firms possess private information that makes them self-select into being diversified. The information is negatively associated with value and leads to the diversification discount. After accounting for unobserved private information, there is no diversification discount: in fact, there is a premium, implying that diversification may well be in shareholders' best interests.

The flip in the sign of d_2 when the selection term is introduced does raise the question of robustness of results, particularly with respect to potential collinearity between the dummy variable for diversification and the inverse Mills ratio that corrects for selection. Campa and Kedia address this issue by reporting several other models, including a simultaneous questions system that instruments the diversification dummy D_{it} and evidence based on a sample of refocusing firms. The main results are robust: there is indeed a diversification discount as found by Lang and Stulz (1994) or Berger and Ofek (1995) when using OLS estimates. However, this discount is not due

²⁸Compared to the standard Heckman model, there is one additional variable in the second stage equation (68), specifically, the dummy variable D_{it} . The Heckman model with the additional dummy variable is called a "treatment effects" model.

to diversification, but by private information that leads firms to become diversified. In fact, the Campa and Kedia self-selection estimates suggest that diversified firms trade at a premium relative to their value had they not diversified.

10.2 Observables and the discount: Villalonga (2004)

While Campa and Kedia (2002) attribute the diversification discount to unobservables causing firms to diversify, Villalonga (2004) offers an explanation based on differences in observables. Villalonga uses a longitudinal rather than cross-sectional analysis, focusing on *changes* in excess value around diversification rather than the level of the excess value itself.

Villalonga’s main sample comprises 167 cases where firms move from being one segment to two segments. She tracks the changes in the diversification discount around the diversification event compared to a control group of non-diversifying firms, using propensity score (PS) based matching to construct matched control firms. Following the methods discussed in Section 4.3.2, Villalonga fits a probit model to estimate the probability that a given firm will diversify using variables similar to those in Campa and Kedia (2002). She matches each diversifying firm with a non-diversifying firm with a similar propensity score, i.e., diversifying probability. Her final sample has five quintiles of firms based on their estimated propensity scores and having a common support.

Villalonga estimates the “treatment effect” caused by diversification as the difference between the change in excess value of a diversifying firm and the excess value change of a comparable non-diversifying firms with the closest propensity score. She reports that while the treatment effect is negative, it is not significant whether she uses the Dehajia and Wahba (1999) or the Abadie and Imbens (2004) technique for estimation. Villalonga also reports similar findings when using a Heckman correction, presumably a treatment effect model on the lines of Campa and Kedia (2002).²⁹

²⁹In reviewing applications, we often found references to “the” Heckman model or “standard” Heckman models to be quite confusing. Campa and Kedia (2002) and Çolak and Whited (2005) use it to denote a treatment effects model, and focus on the coefficient for the diversification dummy variable. However, the Heckman (1979) model is not a treatment effects model. Also, it is not clear whether this reference is to the coefficient for the dummy variable in a treatment effects model or for the inverse Mills ratio term. It is perhaps a better practice not to use labels but instead describe fully the specification being estimated.

Two aspects of Villalonga’s results deserve comment. One issue is perhaps semantic, the use of the term *causal inference*. In reading the work, one could easily come away with the impression that matching methods somehow disentangle causality from correlation. This is incorrect. Matching methods rule out correlation by arbitrary fiat: causality is an *assumption* rather than a statistically tested output of these methods. This fact is indeed acknowledged by Villalonga but easy to overlook given the prominence attached to the term “causal inference” in the paper.

A second issue is that some point estimates of treatment effects are insignificant but not very different in economic magnitude from those in Lang and Stulz (1994) and Berger and Ofek (1995)—and indeed, from the baseline industry-adjusted estimates that Villalonga herself reports. Thus, in fairness to Lang and Stulz and Berger and Ofek, Villalonga’s results do not necessarily refute their earlier work. Nevertheless, Villalonga’s work does make an important point. Specifically, the statistical significance of discount based on industry/size matching methods is not a given fact, but is an open question in light of her results.

10.3 Refocusing and the discount: Çolak and Whited (2005)

If one accepts the diversification discount as a fact, then the question is what causes the discount. One view is that conglomerates (i.e., diversified firms) follow inefficient investment policies, subsidizing inefficient divisions with cash flow from the efficient divisions. Çolak and Whited (2005) evaluate the efficiency of investment in conglomerate and non-conglomerate firms by comparing investments made by focusing firms with those made by firms that do not focus. The focusing sample in Çolak and Whited (2005) consists of 267 divestitures and 154 spinoffs between 1981 and 1996. Control non-focusing firms are multi-segment firms in similar businesses that do not focus in years -3 through +3 where year 0 is the focusing event for a sample point.

The main specification used in Çolak and Whited (2005) employs propensity scores to match focusing and non-focusing firms. As in standard propensity score method implementations, Çolak and Whited (2004) estimate the propensity score as the probability that a given firm will focus in the period ahead. The probit estimates broadly indicate that firms are more likely to focus if they are larger, have less debt, diversity in segments (entropy), and have had recent profit shocks.

The central issue in Çolak and Whited is, of course, on change in investment efficiency after a focusing activity. Çolak and Whited use several measures of change in investment efficiency,

including investment-Q sensitivity, the difference in adjusted investment to sales ratio between high and low growth segments, and the relative value added, which is akin to weighted investment in high minus low Q segments. Çolak and Whited find that the changes in these measures are not significant relative to changes in firms that do not focus and that have similar propensity scores, using the Dehejia and Wahba (1999) matching procedure and the Abadie and Imbens (2004) implementation. There is no evidence that post-spinoff efficiency improves once the focusing firms are matched by propensity score to the non-focusing firms.

For robustness, Çolak and Whited also report estimates of a treatment effects model, Eq. (68) of Campa and Kedia (2002). There is little evidence for efficiency gains, except for one case in which the investment efficiency has a significance level of 10% for focusing firms. This could, however, arise due to pure chance given the wide number of dependent variables and specifications examined. While the paper does not report the coefficient for the inverse Mills ratio in the treatment effects model, Toni Whited confirms to us in private communication, that this selection term *is* significant. This suggests that self-selection is the main explanation for why firms experience efficiency gains after focusing. The unobserved private information that leads firms to focus explains post-focusing improvements in efficiency; controlling for self-selection, there is little evidence of any additional efficiency gains.

10.4 Discussion

A key advantage of the diversification discount literature is that it has reasonably similar datasets, so it is easier to see the changes due to different econometric approaches. By the same token, it becomes easier to raise additional questions on model choice. We raise these questions here for expositional convenience, but emphasize that the questions are general in nature and not particular to the diversification discount literature.

One issue is statistical power. The diversification discount is significant using conventional industry-size matching but it is insignificant using PS based matching methods. Is this because the latter lack power? Çolak and Whited offer some welcome Monte Carlo evidence with respect to their application, simulating data with sample sizes, means, covariance matrix, and covariates with third and fourth moments equal to that observed in the actual data. They confirm that their tests have appropriate size, and at the level of the treatment effects in the sample, there is a better

than 20% chance of detecting the observed treatment effect. More on these lines would probably be useful.

A second issue is the use of PS based matching methods as primary means of inference about treatment effects. There are good reasons to be uncomfortable with such an approach. The main issue is that propensity score methods assume that private information is irrelevant. However, this assumption is probably violated to at least some degree in most corporate finance applications, as illustrated, for instance, by Campa and Kedia (2002) and Çolak and Whited (2005). Thus, using PS methods as the primary specification seems inappropriate without strong arguments as to why firms' private information is irrelevant. As Heckman and Navarro-Lozano (2004) stress, even small deviations from this assumption can introduce significant bias. Thus, the practice of reporting private information specifications in conjunction with matching models, a feature of all three papers discussed above, is appropriate. It is probably a good idea to provide more complete discussions of private information models than typically seen in applications.

A final comment is about the self-selection specifications used to control for private information. While the literature has used versions of the baseline Heckman (1979) model, we emphasize that this restriction is neither necessary nor desirable. Other models, such as switching regressions and structural models are viable alternatives for modeling self-selection and private information. Because these models come with their own additional requirements, it is not clear that they would always be useful, but these issues are ultimately empirical.

11 Other applications of selection models

11.1 Accounting for R&D: Shehata (1991)

Shehata (1991) applies self-selection models to analyze the accounting treatment of research and development (R&D) expenditures chosen by firms during the period of the introduction of FASB ruling SFAS No. 2. This ruling pushed firms to expense rather than defer R&D expenditures. Other studies examined the issue by comparing observed changes in R&D expenditures for a sample of capitalizing firms with those of expensing firms. If firms self-select into the choice they prefer, it is inappropriate to treat the choice as exogenous and assess its impact by comparing differences between capitalizers and expensers. Shehata uses a switching regression instead.

Shehata uses a probit specification to model how firms choose an accounting method, and two regressions to determine the level of the R&D expenditure, one for each accounting choice. This is, of course, the switching regression system of Section 3.1. Shehata estimates the system using standard two-step methods. As discussed in Section 3.1, one useful feature of the system is the estimation of counterfactuals: what the *R&D* spending would be for firms that expensed had they elected to defer and vice-versa. Shehata reports that capitalizers are small, highly leveraged, have high volatility of R&D expenditures, more variable earnings, and spend a significant portion of their income on R&D activities. The second stage regression shows that the two groups of firms behave differently with respect to R&D spending. For instance, R&D is a non-linear function of size and is related to the availability of internally generated funds for capitalizers but the size relation is linear and internally generated funds do not matter for expensers. Thus, it is more appropriate to use a switching regression specification rather than the Heckman (1979) setup to model selection.

The inverse Mills ratio that corrects for self-selection matters in the second stage regression for both groups. Thus, standard OLS estimates tend to understate the impact of SFAS No. 2 on R&D expenditures. Finally, Shehata (1991) reports predictions of the expected values of R&D expenditures for both expensing and capitalizing samples had they elected to be in the other group. The mean value of R&D for each group is lower under the unchosen alternative. The decline is more pronounced for the capitalizing group, where it declines from \$ 0.69 mm to \$ 0.37 mm, while the decline is from \$ 0.85 mm to \$ 0.79 mm for the expensing group.

11.2 Bankruptcy costs: Bris, Welch and Zhu (2005)

Bris, Welch and Zhu (2005) analyze the relative costs of bankruptcy under the Chapter 11 and Chapter 7 procedures in the U.S., codes that are discussed more fully in Chapter 6 of this volume (John, Hotchkiss, Mooradian, and Thorburn (2005)). The sample consists of close to 300 bankruptcy filings in Arizona and Southern New York, the largest sample in the literature as of this writing.

The specification is the basic Heckman model of Section 2, with treatment effects in some specifications. Step 1 is a probit specification that models the choice between Chapter 11 and Chapter 7, conditional on deciding to file for bankruptcy. Bris et. al show that the procedural choice is related to firm characteristics such as size, managerial ownership, and the structure of

debt including variables such as the number of creditors, whether the debt is secured or not, and the presence of banks as a company creditor. Step 2 involves modeling the costs of bankruptcy. Bris et al. analyze four metrics to specify the LHS dependent variable: the change in value of the estate during bankruptcy; the time spent in bankruptcy; the expenses submitted to and approved by the bankruptcy court; and the recovery rates of creditors. These are modeled as a function of a comprehensive set of regressors that include linear and non-linear functions of firm size, various proxies for the structure of the filing firm and managerial ownership. Because the variables in the two stages are similar, the study essentially relies on non-linearity for identification.

Bris et al. find no evidence that firms that were more likely to self-select into Chapter 11 were any faster or slower in completing the bankruptcy process. Controlling for self-selection, Chapter 11 cases consumed more fees, not because Chapter 11 is intrinsically the more expensive procedure, but because of intrinsic differences in firms that choose to reorganize under this code. After controlling for self-selection, Chapter 11 emerges as the cheaper mechanism, and Bris et al. report that self-selection explains about half of the variation in bankruptcy expenses. With self-selection controls, Chapter 11 cases had higher recovery rates than Chapter 7 cases. In sum, selection has a significant impact on estimates of reorganization costs under different bankruptcy codes. After controlling for selection, Chapter 7 takes almost as long, consumes no less and probably more in professional fees, and creditors rarely receive as much, so there is little evidence that it is more efficient than Chapter 11 reorganizations.

11.3 Family ownership and value: Villalonga and Amit (2004)

Villalonga and Amit (2004) examine the effect of family ownership, control, and management on value for a sample of Fortune 500 firms from 1994 to 2000. The specification is a standard Heckman style selection model of Section 2 with a treatment effect.

The first step is a probit specification that models whether a firm remains family owned or not. Family ownership is defined as firms in which the founding family owns at least 5% of shares or holds the CEO position. In the second step, value, proxied by Tobin's Q, is regressed on a dummy variable for family ownership, industry dummy variables, the Gompers, Ishii and Metrick (2003) shareholder rights index, firm-specific variables from COMPUSTAT, outside block ownership and proportion of non-family outside directors, and, of course, the inverse Mills ratio that corrects for

self-selection. To assist in identification, Villalonga and Amit include two additional instruments in the selection equation lagged Q and idiosyncratic risk. Idiosyncratic risk is presumably related to family ownership but not to Q if only systematic risk is priced by the market.

Villalonga and Amit report that family ownership has a positive effect on value in the overall sample and in sub-samples in which the founder is the CEO. Interestingly, the sign is negative when the founder is not the CEO. Villalonga and Amit interpret their findings as evidence that the benefits of family ownership are lost when the family retains control in the post-founder generation. Their results strengthen when they incorporate a control for self-selection. In the self-selection specification, the inverse Mills ratio is significant and negative in the overall specification and sub-samples in which the CEO is the founder. In these samples, family ownership appears to be associated with unobserved attributes that are negatively related to value. These unobserved attributes positively impact value if the founder is not the CEO.³⁰

12 Other applications of matching methods

12.1 Bank debt versus bonds: Bharath (2004)

Debt financing by a corporation gives rise to conflicts of interest between creditors and shareholders that can reduce the value of the firm. Such conflicts are limited more effectively in bank loans than in public debt issues if banks monitor. Bharath (2004) measures the size of agency costs by calculating the yield spread between corporate bonds and bank loans (the Bond-Bank spread) of the same firm at the same point in time. To quantify the difference, Bharath needs to match bonds with bank loans of the same firm at the same point in time and having substantively identical terms. The matching problem is complicated by the fact that bank loans and public bonds are contractually very different on multiple dimensions such as credit rating, seniority, maturity, and collateral.

Bharath argues that because bank loans and bonds are matched at the same point of time and for the same firm, matching based on observables should adequately control for differences between

³⁰An interesting question raised by this study is survivorship (e.g., Brown, Goetzmann and Ross (1995)). Perhaps family owned firms that survived and made it to Fortune 500 status are of better quality, and hence these firms are valued more. This question can perhaps be resolved by looking at broader samples that incorporate smaller firms outside the Fortune 500 universe.

bank debt and public debt. Thus, propensity score based matching methods are appropriate tools to control for differences between bank loans and public debt. Bharath uses the propensity score matched difference between bank and bond credit spreads as the treatment effect, or the value added by banks. The spread can be interpreted as the value added by banks in enforcing better investment policies, or more generally, as the price of the “specialness” of banks due to their ability to monitor, generate information, or better renegotiate loans, or even perhaps other explanations such as monopoly rents.

Using a sample of over 15,000 yield observations, Bharath finds that the Bond-Bank spread is negative for high credit quality firms and positive for low credit quality firms. He interprets his findings as being consistent with the view that for high quality firms, the benefits of bank monitoring are outweighed by the costs of bank hold-up. This causes the spread to be negative, indicating that bank debt offers few benefits for high quality firms. For low quality firms, the opposite is true, causing the spread to be positive. The magnitude of the potential agency costs mitigated by banks is more important for poor quality firms, justifying the decision to borrow from banks.

12.2 Matching and long-run performance: Cheng (2003), Li and Zhao (2005)

A vast literature on market efficiency examines the long-run stock return after events such as IPOs, SEOs, share repurchases, listing changes, etc. The semi-strong version of the efficient markets hypothesis predicts that long-run returns should be zero on average. However, several papers report empirical evidence against the efficiency hypothesis (Fama (1998)). In most studies, post-event buy-and-hold returns are systematically positive or negative relative to benchmarks over periods of three to five years. Chapter 1 in this volume (Kothari and Warner (2005)) offers an overview of this literature. We focus on applications of matching models to assess long-run performance.

To test whether abnormal returns are zero or not, one needs a model of benchmark returns. As discussed in Chapter 1, the standard approach, is to match an event firm with a non-event firm on between two and four characteristics that include size, book-to-market, past returns, and perhaps industry. This method runs into difficulties when the number of dimensions becomes large and the calipers become fine, when it becomes difficult to generate matching firms. Propensity score (PS) based matching methods reviewed in Section 4.3.2 are potentially useful alternatives in this

scenario. Two recent papers, Cheng (2003) and Li and Zhao (2005) use PS methods to reexamine the long-term performance of stock returns after SEOs. Both papers find that while characteristic-by-characteristic matching results in significant long-term abnormal returns after SEOs, abnormal returns are insignificant if one uses propensity score based matching methods instead.

Cheng (2003) studies SEOs offered between 1970 and 1997 for which necessary COMPUSTAT data are available on firm characteristics. She finds significant buy-and-hold abnormal returns of between -6% and -14% over three to five years in the full sample and various sub-samples when matches are constructed on size, industry and book-to-market. She then uses three logit models, one for each decade, to predict the probability of issuance. Several firm characteristics such as size, book-to-market, industry, R&D, exchange, as well as 11-month past returns predict the issuance decision. Cheng matches each issuer with a non-issuer in the SEO year with a similar propensity score (i.e., predicted probability). She finds little evidence of significant abnormal returns except for one sub-sample in the 1970s.

Li and Zhao undertake an exercise similar to that in Cheng (2003) for issuers from 1986 to 1997. They show that characteristic-by-characteristic matching produces inadequate matches between issuers and non-issuers in terms of average size.³¹ They estimate propensity scores with size, book-to-market, and past returns in three quarters prior to issuance, one model per year, and add interaction terms for better predictions and delete firms as necessary to have a common support. In their final sample, conventional matching gives average three-year buy-and-hold abnormal returns of -16%, but this drops to an insignificant -4% with PS matching.

Cheng (2003) and Li and Zhao (2005) emphasize that PS methods are merely substitutes for characteristic-by-characteristic matching of observables. This perspective is probably appropriate. The main issue in these applications is the data driven nature of the exercise in fitting probit models. Characteristics and interaction terms are added as needed to achieve balance in characteristics and propensity scores. While we recognize that a reasonable probit model seems necessary to place faith in treatment effect estimates, the search required to achieve balance, however transparent, nevertheless raises data dredging concerns and even inconsistency of estimates (Heckman and Navarro-Lozano (2004)). The general use of PS methods in studies of long-term stock return

³¹Medians are not reported, so it is hard to assess the role of outliers.

or operating performance as an alternative to methods studied in Barber and Lyon (1996, 1997), Barber, Lyon and Tsai (1999), and Kothari and Warner (1997) remains an open question.

13 Bayesian methods

13.1 Matching: Sørensen (2003)

Investors differ in their abilities to select good investments, and in their ability to take a given investment and monitor and manage it so as to add value to what they invest in. A key question in the venture capital literature is the differentiation of selection from value-addition. To what extent are better performing venture capitalists more successful because of their ability to select good investments rather than their ability to value-add to their investments? Sørensen (2003) employs a matching-selection model to separate these two influences, using Bayesian MCMC (Markov Chain Monte Carlo) methods to estimate it.

In Sørensen's model, there is a set of venture capital investors indexed by i . Each investor evaluates a set of potential investments indexed by j and ultimately invests (i.e., becomes the lead investor) in a subset of these. Once an investment occurs, its outcome is specified as the variable IPO which equals one if the investment results in a public offering and zero otherwise. In Sørensen's model, feasible investments for each investor are partly determined by the characteristics of the other agents in the market. These characteristics are related to the investment decision but unrelated to the investment outcome, so they provide the exogenous variation used for identification of the model. On the other hand, this type of sorting also causes interaction between investment decisions by different venture capitalists, which leads to a dimensionality problem and considerable numerical difficulties in estimation. Bayesian methods offer feasible routes for estimation.

Sørensen specifies normally distributed and diffuse prior beliefs with prior variances that are over 300 times the posterior variance. He assumes that error terms for different deals are independent. There are three sets of exogenous variables. The characteristics of the company includes the stage of development of the company and industry dummies. The characteristics of the venture capital investor include his experience and amount of capital he has available. The characteristic of the market is the year of the investment. There are two parameters of central interest. One is the access of better venture capitalists to deal flow, which is captured by the experience of the venture

capitalist. The other is the synergy between venture capitalists and their target investments or the value added by venture capitalists, which is captured by the correlation between the private information in the decision to invest and the probability of going public.

Sørensen's final sample includes 1,666 investments made by 75 venture capitalists between 1975 and 1995 in the states of California and Massachusetts. Experience is proxied by the total and stage-of-life-cycle-specific number of deals done since 1975. Sørensen reports a number of interesting findings. He finds evidence for sorting. Experienced investors are more likely to have access to the better deals whose probability of going public (and doing so faster) increases by about two-thirds. This type of sorting explains about 60% of the increased probability of success, leaving about 40% for the synergies, or the value added by venture capital investors. Sørensen explains why one might get different results from estimating a standard selection model compared to one with sorting.

13.2 Switching regressions: Li and McNally (2004), Scruggs (2005)

Li and McNally (2004) and Scruggs (2005) offer interesting applications of Bayesian methods to estimate switching regression models of self-selection. Both papers emphasize that the value of the Bayesian approach is not merely the difference in philosophy or technique; rather, the techniques offer insights not readily available through classical methods. The application in Li and McNally (2004) is the choice of a mechanism to effect share repurchases, while the application in Scruggs relates to whether convertibles are called with or without standby underwriting arrangements. For convenience, we focus on Li and McNally, but substantially similar insights on methodology are offered in the work by Scruggs.

Share repurchases started becoming popular in the 1980s as a way to return excess cash to shareholders in lieu of dividends. Repurchases tend to be more flexible in timing and quantity relative to the fixed cash flow stream expected by markets when companies raise dividends. Share repurchases can be implemented in practice as a direct tender offer or more open-ended open market repurchases. Li and McNally (2004) investigate the choice between the two mechanisms and their impact on share price reactions to announcements of repurchases using Bayesian self-selection methods.

Li and McNally propose the following system of equations to analyze the choice of a repurchase mechanism

$$I^* = Z_i\gamma + \eta_i \quad (69)$$

$$p_1^* = X_1\beta_1 + \epsilon_1 \quad (70)$$

$$p_2^* = X_2\beta_2 + \epsilon_2 \quad (71)$$

$$y_1^* = W_1\alpha_1 + v_1 \quad (72)$$

$$R_1^* = V_1\theta_1 + u_1 \quad (73)$$

$$R_2^* = V_2\theta_2 + u_2 \quad (74)$$

where I^* is an unobserved latent variable representing the incremental utility of tender offers over open market repurchases, p_1^*, y_1^*, R_1^* are the percentage of shares sought, tender premium and announcement effects under the tender offer regime, and p_2^*, R_2^* are the proportion sought and announcement effects in an open market repurchase regime. The error terms in Eqs. (69)-(74) are assumed to have a multivariate normal distribution.

The system of equations (69)-(74) represents a switching regression system discussed in Section 3.1, but with more than one regression in each regime. The key issue in estimating the system is the lack of information on unobserved counterfactuals. We observe outcomes in the repurchase technique actually chosen by a firm but do not explicitly observe what would happen if the firm had chosen the alternative technique instead. Li and McNally employ MCMC methods that generate counterfactuals as a natural by-product of the estimation procedure. This approach involves a data augmentation step in which the observed data are supplemented with counterfactuals generated consistent with the model structure. The priors about parameters are updated and posteriors obtained using standard simulation methods after which the additional uncertainty due to the data augmentation step can be integrated out. Observations on counterfactual choices and outcomes are generated as part of the estimation procedure. These can be directly used to examine the impact of choosing a given type of repurchase mechanism not just in isolation, but also relative to the impact of choosing the unchosen alternative.

The sample in Li and McNally comprises 330 fixed price tender offers, 72 Dutch auction tender offers, and 1,197 open market repurchases covering time periods from 1962 to 1988. In terms of findings, Li and McNally report that firms choose the tender offer mechanism when they have

financial slack and large shareholders that monitor management. Firms prefer the open market repurchase in times of market turbulence or weak business conditions. Unobserved private information affects both the type of the repurchase program and the repurchase terms and is reflected in the stock market announcement effects. The estimates of counterfactuals are quite interesting. For instance, if the open market repurchasers had opted for tender offers, the proportion of shares sought would have been 36% (versus actual of about 7%) and the tender premium would have been 33% compared to 0% actuals, and the five-day announcement effect would be 16% compared to the actual announcement effect of 2.2%. Likewise, tender offer firms would have repurchased 10.6% (actual = 19.7%) and experienced announcement effects of 3.7% (actual= 10.2%). Firms appear to have a comparative advantage in their chosen repurchase mechanisms.

14 Conclusions

Our review suggests that self-selection is a growth area in empirical corporate finance. The rapidly expanding number of applications undoubtedly reflects the growing recognition in the finance profession that self-selection is an important and pervasive feature of corporate finance decisions. The range of econometric models in use is also growing as techniques diffuse from the econometrics literature to finance. However, the key issue in implementing self-selection models still remains the choice of specification, particularly the economic assumptions that make one model or another more appropriate for a given application. One size does not fit all. Each self-selection model addresses a different kind of problems, places its own demands on the type of data needed, and more importantly, carries its own baggage of economic assumptions. The plausibility of these assumptions is perhaps the primary criterion to guide what is used in empirical applications.

References

- [1] Abadie, A., and G. Imbens, 2004, Simple and bias-corrected matching estimators for average treatment effects, Harvard University working paper.
- [2] Acharya, S., 1988, A generalized econometric model and tests of a signaling hypothesis with two discrete signals, *Journal of Finance*, 43, 413-429.
- [3] Acharya, S., 1993, Value of latent information: Alternative event study methods, *Journal of Finance*, 48, 363-385.
- [4] Barber, B., and J. Lyon, 1996, Detecting abnormal operating performance: The empirical power and specification of test statistics, *Journal of Financial Economics*, 41, 359-400.
- [5] Barber, B., and J. Lyon, 1997, Detecting long-run abnormal stock returns: The empirical power and specification of test statistics, *Journal of Financial Economics*, 43, 341-372.
- [6] Barber, B., J. Lyon, and C. Tsai, 1999, Improved methods for tests of long-run abnormal stock returns, *Journal of Finance*, 54, 165-201.
- [7] Berger, P.G., and E. Ofek, 1995, Diversification's effect on firm value, *Journal of Financial Economics*, 37, 39-65.
- [8] Bertrand, M., and S. Mullainathan, 2003, Enjoying the quiet life? Corporate governance and managerial preferences, *Journal of Political Economy*, 111, 1043-1075.
- [9] Betton, S., E. Eckbo, and K. Thorburn, 2005, Takeovers, restructurings, and corporate control, *Handbook of Corporate Finance: Empirical Corporate Finance*, B. Espen Eckbo, ed., Elsevier/North-Holland.
- [10] Bharath, S., 2004. Agency costs, bank specialness and renegotiation, University of Michigan Stephen M. Ross School of Business working paper.
- [11] Bohren, O., B.E. Eckbo, and D. Michalsen, 1997, Why underwrite rights offerings? Some new evidence, *Journal of Financial Economics*, 46, 223-261.
- [12] Bris, A., N. Zhu, and I. Welch, 2005, The cost of bankruptcy: Chapter 7 cash auction vs. Chapter 11 bargaining, *Journal of Finance*, forthcoming.

- [13] Brown, S., W. Goetzmann, and S. Ross, 1995, Survival, *Journal of Finance* 50, 853-873.
- [14] Campa, J.M., and S. Kedia, 2002, Explaining the diversification discount, *Journal of Finance*, 57, 1731-1762.
- [15] Cheng, Y., 2003, Propensity score matching and the new issues puzzle, Florida State University working paper.
- [16] Chevalier, J., 2000, Why do firms undertake diversifying mergers? An examination of the investment policies of merging firms, Yale School of Management working paper.
- [17] Chiappori, P., and B. Salanie, 2000, Testing for asymmetric information in insurance markets, *Journal of Political Economy* 108, pp. 56-78.
- [18] Chib, S., and E. Greenberg, 1996, Markov Chain Monte Carlo simulation methods in econometrics, *Econometric Theory*, 12, 409-431.
- [19] Coase, R.H., 1933, The nature of the firm, *Economica*, 4, 386-405.
- [20] Çolak, G., and T.M. Whited, 2005, Spin-offs, divestitures, and conglomerate investment, University of Wisconsin working paper.
- [21] Dehejia, R.H., and S. Wahba, 1999, Casual effects in non-experimental studies: Re-evaluating the evaluation of training programs, *Journal of American Statistical Association*, 94, 1053-1062.
- [22] Dunbar, C.G., 1995, The use of warrants as underwriter compensation in initial public offerings, *Journal of Financial Economics*, 38, 59-78.
- [23] Eckbo, B.E., 1990, Valuation effects of greenmail prohibitions. *Journal of Financial and Quantitative Analysis*, 25, 491-505.
- [24] Eckbo, B.E., 1992, Mergers and the value of antitrust deterrence, *Journal of Finance*, 47, 1005-1029.
- [25] Eckbo, B.E., V. Maksimovic, and J. Williams, 1990, Consistent estimation of cross-sectional models in event studies, *Review of Financial Studies*, 3, 343-365.

- [26] Fama, E., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283-306.
- [27] Fang, L.H., 2005, Investment bank reputation and the price and quality of underwriting services, *Journal of Finance*, forthcoming.
- [28] Gande, A., M. Puri, and A. Saunders, 1999, Bank entry, competition, and the market for corporate securities underwriting, *Journal of Financial Economics*, 54, 165-195.
- [29] Gande, A., M. Puri, A. Saunders, and I. Walter, 1997, Bank underwriting of debt securities: Modern evidence, *Review of Financial Studies*, 10, 1175-1202.
- [30] Gelfand, A.E., and A.F.M. Smith, 1990, Sampling based approaches to calculating marginal densities, *Journal of American Statistical Association*, 85, 398-409.
- [31] Goldberger, A., 1983, Abnormal selection bias, in: S. Karlin, T. Amemiya and L. Goodman eds., *Studies in Econometrics, Time Series, and Multivariate Statistics*, New York: Academic Press.
- [32] Gompers, P.A., J.L. Ishii, and A. Metrick, 2003, Corporate governance and equity prices, *Quarterly Journal of Economics*, 118, 107-155.
- [33] Goyal, V.K., 2005, Market discipline of bank risk: Evidence from subordinate debt contracts, *Journal of Financial Intermediation*, forthcoming.
- [34] Graham, J.R., M.L. Lemmon, and J.G. Wolf, 2002, Does corporate diversification destroy value? *Journal of Finance* 57, 695-720.
- [35] Greene, W., 1981, Sample selection bias as a specification error: Comment, *Econometrica*, 49, 795-798.
- [36] Greene, W., 2003, *Econometric Analysis*, 5th edition, Prentice Hall.
- [37] Harris, M., and A. Raviv, 1985, A sequential signaling model of convertible debt call policy, *Journal of Finance*, 40, 1263-1282.
- [38] Heckman, J.J., 1979, Sample selection as a specification error, *Econometrica*, 47, 153-161.

- [39] Heckman, J.J., H. Ichimura, and P. Todd, 1998, Matching as an econometric evaluation estimator, *Review of Economic Studies*, 65, 261-294.
- [40] Heckman, J.J., and S. Navarro-Lozano, 2004, Using matching, instrumental variables, and control functions to estimate economic choice models, *Review of Economics and Statistics*, 86, 30-57.
- [41] Himmelberg, C., R. Hubbard, and D. Palia, 1999, Understanding the determinants of managerial ownership and the link between ownership and performance, *Journal of Financial Economics*, 53, 353-384.
- [42] Hubbard, R.G., and D. Palia, 1995, Benefits of control, managerial ownership, and the stock returns of acquiring firms, *RAND Journal of Economics*, 26, 782-792.
- [43] John, K., E. Hotchkiss, R.C. Mooradian, and K. Thorburn, 2005, Bankruptcy and financial distress, *Handbook of Corporate Finance: Empirical Corporate Finance*, B. Espen Eckbo, ed., Elsevier/North-Holland.
- [44] Koop, G., 2003, *Bayesian Econometrics*, John Wiley & Sons Ltd.
- [45] Kothari, S.P., and J.B. Warner, 1997, Measuring long-horizon security price performance, *Journal of Financial Economics*, 43, 301-339.
- [46] Kothari, S.P., and J.B. Warner, 2005, Econometrics of event studies, *Handbook of Corporate Finance: Empirical Corporate Finance*, B. Espen Eckbo, ed., Elsevier/North-Holland.
- [47] Lang, L.H.P., and R.M. Stulz, 1994, Tobin's Q, corporate diversification, and firm performance, *Journal of Political Economy*, 102, 1248-1280.
- [48] Lee, L.F., 1983, Generalized econometric models with selectivity, *Econometrica*, 51, 507-512.
- [49] Li, K., and W. McNally, 2004, Open market versus tender offer share repurchases: A conditional event study, Sauder School of Business UBC working paper, available at <http://finance.sauder.ubc.ca/~kaili/buyback.pdf>.
- [50] Li, K., and W. McNally, 2006, The information content of Canadian open market repurchase announcements, *Managerial Finance (A Special Issue on Payout Policy)*, forthcoming.

- [51] Li, X., and X. Zhao, 2005, Is there an SEO puzzle? York University working paper.
- [52] Ljungqvist, A., F. Marston, and W.J. Wilhelm, Jr., 2005, Competing for securities underwriting mandates: Banking relationships and analyst recommendations, *Journal of Finance*, forthcoming.
- [53] Maddala, G.S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.
- [54] Maksimovic, V., and G. Phillips, 2002, Do conglomerate firms allocate resources efficiently? *Journal of Finance* 57, 721-767.
- [55] Maksimovic, V., and G. Phillips, 2005, Internal capital markets, *Handbook of Corporate Finance: Empirical Corporate Finance*, B. Espen Eckbo., ed., Elsevier/North-Holland.
- [56] Maksimovic, V., and H. Unal, 1993, Issue size choice and “underpricing” in thrift mutual-to-stock conversions, *Journal of Finance* 48, 1659-1692.
- [57] Murphy, K., and R. Topel, 1985, Estimation and inference in two step econometric models, *Journal of Business and Economic Statistics*, 3, 370-379.
- [58] Myers, S.C., and N. Majluf, 1984, Corporate financing and investment decisions when firms have information investors do not have, *Journal of Financial Economics*, 13, 187-221.
- [59] Nayak, S., and N.R. Prabhala, 2001, Disentangling the dividend information in splits: A decomposition using conditional event-study methods, *Review of Financial Studies*, 14, 1083-1116.
- [60] Newey, W., J. Powell, and J. Walker, 1990, Semi-parametric estimation of selection models, *American Economic Review*, 80, 324-328.
- [61] Palia, D., 2001, The endogeneity of managerial compensation in firm valuation: A solution, *Review of Financial Studies* 14, 735-764.
- [62] Poirier, D.J., 1995, *Intermediate Statistics and Econometrics*, Cambridge: MIT Press.
- [63] Prabhala, N.R., 1997, Conditional methods in event studies and an equilibrium justification for standard event-study procedures, *Review of Financial Studies*, 10, 1-38.

- [64] Puri, M., 1994, The long-term default performance of bank underwritten securities issues, *Journal of Banking and Finance*, 18, 397-418.
- [65] Puri, M., 1996, Commercial banks in investment banking: Conflict of interest or certification role? *Journal of Financial Economics*, 40, 373-401.
- [66] Reisel, N., 2004, On the value of restrictive covenants: An empirical investigation of public bond issues, Rutgers University working paper.
- [67] Rosenbaum, R., and D. Rubin, 1983, The central role of the propensity score in observational studies for causal effects, *Biometrika* 70, 41-55.
- [68] Roy, A.D., 1951, Some thoughts on the distribution of earnings, *Oxford Economic Papers*, 3, 135-146.
- [69] Schoar, A., 2002, The effect of diversification on firm productivity, *Journal of Finance*, 62, 2379-2403.
- [70] Scruggs, J.T., 2005. Estimating the cross-sectional market response to an endogenous event: Naked vs. underwritten calls of convertible bonds, University of Georgia Terry College of Business working paper.
- [71] Servaes, H., 1994, Do takeover targets overinvest? *Review of Financial Studies*, 7, 253-277.
- [72] Servaes, H., 1996, The value of diversification during the conglomerate merger wave, *Journal of Finance*, 51, 1201-1225.
- [73] Shehata, M., 1991, Self-selection bias and the economic consequences of accounting regulation: An application of two-stage switching regression to SFAS No. 2, *Accounting Review*, 66, 768-787.
- [74] Song, W., 2004, Competition and coalition among underwriters: The decision to join a syndicate, *Journal of Finance*, 59, 2421-2444.
- [75] Sørensen, M., 2003, How smart is smart money? An empirical two-sided matching model of venture capital, Stanford School of Business working paper.

- [76] Staiger, D., and J.H. Stock, 1997, Instrumental variable regression with weak instruments, *Econometrica*, 65, 557-586.
- [77] Tanner, M.A. and W.H. Wong, 1987, The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, 82, 528-550.
- [78] Villalonga, B., 2004, Does diversification cause the “diversification discount”? *Financial Management*, 33, 5-27.
- [79] Villalonga, B., and R. Amit, 2004, How do family ownership, control, and management affect firm value? *Journal of Financial Economics*, forthcoming.
- [80] Vuong, Q., 1989, Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, 57, 307-333.
- [81] Wooldridge, J., 2002, *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.
- [82] Zhou, X., 2001, Understanding the determinants of managerial ownership and the link between ownership and performance: Comment, *Journal of Financial Economics*, 62, 559-571.