View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Maastricht University Research Portal

Maastricht University

Microeconometric evidence of financing frictions and innovative activity - a revision

Citation for published version (APA):

Tiwari, A. K., Mohnen, P., Palm, F. C., & Schim van der Loeff, S. (2013). Microeconometric evidence of financing frictions and innovative activity - a revision. (UNU-MERIT Working Papers; No. 027). Maastricht: UNU-MERIT, Maastricht Economic and Social Research and Training Centre on Innovation and Technology.

Document status and date:
Published: 01/01/2013

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 04 Dec. 2019

Working Paper Series

#2013-027

Microeconometric evidence of financing frictions and innovative activity – a revision

Amaresh K Tiwari, Pierre Mohnen, Franz C Palm and Sybrand Schim van der Loeff

This paper is a revision of working paper #2012-62

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://mgsog.merit.unu.edu

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

UNU-MERIT Working Papers

ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology, UNU-MERIT

Maastricht Graduate School of Governance MGSoG

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSoG to stimulate discussion on the issues raised.

MICROECONOMETRIC EVIDENCE OF FINANCING FRICTIONS AND INNOVATIVE ACTIVITY[1]

AMARESH K TIWARI[2], PIERRE MOHNEN[3], FRANZ C PALM[4] AND SYBRAND SCHIM VAN DER LOEFF[5]

Using Dutch data we empirically investigate how financing and innovation vary across firm characteristics. We find that when firms face financial constraints, debt financing and innovation choices are not independent of firm characteristics, and R&D slows down. In the absence of financial constraints, however, as they raise debt, firms become less inclined to innovate and the change in the propensity to innovate no longer varies with firm characteristics. We find that financing constraints faced, propensity to innovate, and R&D intensity are not uniform across firm characteristics. A new "Control Function" estimator to account for heterogeneity and endogeneity has been developed.

KEYWORDS: Innovation, R&D, Capital Structure, Financial Constraints, Firm Characteristics, Correlated Random Effects, Control Function, Expected a Posteriori.

JEL Classification: G30, O30, C30

1. INTRODUCTION

In this paper we empirically investigate how incentives to innovate interact with financing frictions that are related to innovative activity. We show that financing and innovation choices vary with firm characteristics such as size, age, and leverage for financially con-

[1]The authors gratefully acknowledge the financial support from The Netherlands Organization for Scientific Research (NWO) under grant 400-03-109. The authors wish to thank Statistic Netherlands (CBS) for providing the data; in particular, they would like to thank Gerard Meinen for helping with the different aspects of the data. The authors would also like to thank Lionel Artige, Aline Muller, seminar participants at HEC of the University of Liege, participants at ZEW at Mannheim, and participants at the 66^{th} European Meeting of the Econometric Society, Malaga, for their helpful comments. All remaining errors are our own.

[2]University of Liege, A.Tiwari@ulg.ac.be

[3] Maastricht University, UNU-MERIT and CIRANO, P.Mohnen@maastrichtuniversity.nl

[4]Maastricht University and CESifo fellow, F.Palm@maastrichtuniversity.nl

[5] Maastricht University, S.Loeff@maastrichtuniversity.nl

strained and unconstrained firms. This implies that the incentives to innovate and the extent and the nature of frictions are not uniform across firm characteristics. Our results, thus, inform theory that in modeling firm dynamics, investment in R&D along with investment in physical capital and the financing decisions of the firm must be taken into account, especially since, given the nature of R&D activity, the associated financing frictions can be acute.

For our empirical analysis we use a unique data set where firms report if they have faced financial constraint due to which some of their R&D projects were hampered. To study how financing and innovation policy vary with firm characteristics and to establish the extent of impact due to existence of financing frictions on innovative activity, for our empirical analysis we write a fully specified econometric model of R&D investment with endogenous financial constraint, endogenous decision to innovate, and endogenous financial choices. This entails estimating a system of structural equations pertaining to (a) a model for decision to innovate, where we study the financing choices of innovative firms, (b) a model for financial constraint, where we try to explain why certain firms report they are financially constrained, (c) model for R&D investment, where we try to assess the impact of financial constraint, as reported by the firms, on R&D investment, and (d) a system of reduced form equations of financing choice and other endogenous variables.

Firstly, in our study of innovation and financing choice we find results that are in congruence with the papers that provide empirical evidence that R&D intensive firms are less leveraged than those that are not. Brown et al. (2009) (henceforth BFP) studying a panel of R&D performing US firms draw out a financing hierarchy for R&D intensive firms, where equity – when more easily available, as during the boom in the supply of internal and external equity finance in the mid and late 1990's in US – might be preferred to debt as a means of financing R&D. Corroborating Brown et al. (2012), we find that, ceteris paribus, innovative firms are likely to maintain higher levels of internal liquidity reserve. We also find that firms that pay out dividends are less likely to take up innovative activity, suggesting that external financing could be more costly for innovative firms.

Secondly, given that firms themselves report if they are financially constrained with respect to innovative activity, in our empirical model of endogenous financing and innovation choices and endogenous financial constraint we are able to assess if certain financing choices, as reflected in the balance sheets of the firms, determine whether a firm is financially constrained. This allows us to assess the relevance of the classification criteria (see Hennessy and Whited, 2007, henceforth HW) that distinguish firms as financially constrained or unconstrained, and which have been motivated by the theories of financial contracting. We find that small and young firms, firms that are highly levered, and firms that pay less dividends are more likely to face financial constraint. The finding is in line with prediction made by Albuquerque and Hopenhayn (2004) (henceforth AH) and Clementi and Hopenhayn (2006) (henceforth CH) in models of firm growth and survival with endogenous borrowing constraint. We also find that firms that maintain high level of liquidity reserve and those whose asset base includes more tangible assets are less likely to face financial constraint.

Our third and important set of findings are from the investigation of innovation and financing choices under financial constraint and under no constraint. We find that (i) under financial constraint, extent of which varies with firm characteristics such as age, size, and existing leverage, the change in propensity to innovate by employing more long-term debt also varies with the firm characteristics. However, (ii) when financial constraint do not bind the change in propensity to innovate by increasing leverage does not, or vary little with firm characteristics, and is uniformly lower as compared to the situation when financial constraint bind. Some other important findings that underscore the fact that innovation and financing decision are not uniform across firm characteristics are (iii) that large and young firms are more likely to engage in innovative activity, (iv) that large and mature firms are less R&D intensive, and (v) that small and younger firms are more financially constrained. These third set of findings suggest that decisions to innovate, financing choices, and firm dynamics are not independent.

Now, while there are models of efficient firm and industry dynamics where R&D activity and uncertainty in innovation explain some of the stylized facts related to R&D investment, productivity, firm dynamics, and firm size distribution (see for eg. Klette and Kortum (2004) (henceforth KK), Klepper and Thompson (2006)), none to our knowledge has explored the interaction of financing frictions with innovative activity in shaping up firm and industry dynamics. Some of our results, for example, that not all firms are innovative, that under financial constraint financing and innovation policy are not independent of firm characteristics, and that R&D intensity is not independent of firm size are contrary to what KK purport to explain. Our findings suggest that in modeling firm dynamics with R&D and innovation, financial consideration too must also be taken into account.

On the other hand there are models, such as by Cooley and Quadrini (2001) (henceforth CQ), of financial market inefficiency, where financial frictions introduced in a standard model of firm and industry dynamics generate results that match the empirical regularities of the financial and company level investment characteristics of firms that are related to their size and age. AH and CH in their respective papers develop models of endogenous borrowing constraint and study its implication for firm dynamics such as growth and survival. However, R&D and innovation do not feature in these models. In AH and CH borrowing constraint is hinged on the capital structure of the firm, where state contingent equity value determines borrowing constraint, exit probability, and expansion. Our results suggest that capital structure also matters for the exercise of growth options that are related to R&D. While a successful completion and implementation of a R&D project enhances firm productivity and chances of survival, given the nature of R&D and the fact that it is affected by various kinds of uncertainties (see Berk et al., 2004), engaging in R&D will also affect the evolution of equity, thereby affecting borrowing constraint and firm growth and survival. Our results suggests that modeling firm dynamics with R&D and innovation that incorporates borrowing constraint in a dynamic financial contract framework could be an important area of research.

Fourthly, our paper contributes to the empirical literature that seeks to test for financing frictions and quantifying the extent of market failure due to existence of financing frictions. The small number of empirical studies on testing for financing frictions for R&D investment are documented in Hall and Lerner (2010). More recent papers such as, Whited (2006), Bayer (2006), and Bayer (2008), studying company level investment, show how financing frictions interact with adjustment costs to alter the timing of company level lumpy investment. HW find that existence of costly external funds depresses the path of investment. Hajivassiliou and Savignac (2011), using a similar data set for France, find that financial constraint do adversely affect innovation output. Our objective here is to assess by how much R&D investment is hampered given that a firm faces financial constraint.

Empirical analysis in corporate finance, as discussed in Roberts and Whited (2010), is marred with issues of endogeneity. Our estimation strategy combines the method of "correlated random effect" and "control function" (see Blundell and Powell, 2003) to account for unobserved heterogeneity and endogeneity of regressors in the structural equations. We estimate the fully specified model, stated earlier, in three steps. In the first step we estimate the system of reduced form equations, the estimates of which are then used to construct the control functions that correct for the bias that can arise due the presence of endogenous regressors in the structural equations. With the control functions in place, in the second stage we jointly estimate the structural model of financial constraint faced by the firms and the decision to innovate, and finally in the third stage, conditional on the decision to innovate, we estimate the switching regression model of R&D investment to assess the impact of financial constraint on R&D investment.

Typically, in a control function approach the structural parameters are estimated conditional on unobserved heterogeneity and unobserved idiosyncratic errors that appear in reduced form equations of a simultaneous triangular system of equations. In such an approach residuals obtained from the first stage reduced form estimates that proxy for the idiosyncratic errors are used as control variables in the structural equations to account for the endogeneity of the regressors in the structural equations. However, in panel data models, the residuals of the reduced form regression, which are defined as the observed value of the response variable minus the expected value of the response conditional on exogenous regressors and the individual effects, are functions of unobserved individual effects. Since the individual effects are unobserved, the residuals remain unidentified. The novelty of our approach lies in integrating out the unobserved individual effects. The integration is performed with respect to the conditional distribution of the individual effects approximated by the posterior distribution of the individual effects obtained from the first stage reduced form estimation. This leaves us with the expected a posteriori (EAP) values of the individual effects, which can then be used to get the residuals. The paper also provides the theoretical foundations for such a procedure.

The rest of the paper is organized as follows. In Section 2 we present the economic framework, in Section 3 we discuss the empirical strategy employed, in Section 4 the data used and the definition of the variables are discussed, in Section 5 we present the results, and finally in Section 6 we conclude. In Appendix A we discuss the identification of the structural parameters. The details of the econometric methodology are provided in appendices B, C, D, and E. All the appendices, for reasons of space, have not included in the core of the paper, but can be made available upon request.

2. FINANCING FRICTIONS AND INNOVATIVE ACTIVITY

A. *Financing and Innovation Decision*

Holmstrom (1989) points out that from the perspective of investment theory R&D has a number of characteristics that make it different from ordinary investment: it is long-term in nature, high risk in terms of the probability of failure, unpredictable in outcome, labor intensive, and idiosyncratic. The high risk involved and unpredictability of outcomes are potential sources of asymmetric information that give rise to agency issues in which the inventor frequently has better information about the likelihood of success and the nature of the contemplated innovation project than the investors. Leland and Pyle (1977) point out that investors have more difficulty distinguishing good or low risk projects from bad ones when they are long-term in nature. Besides, due to the ease of imitation of inventive ideas, as pointed out by Hall and Lerner (2010), firms are reluctant to reveal their innovative ideas to the marketplace, and there could be a substantial cost to revealing information to their competitors. Thus the implication of asymmetric information coupled with the costliness of mitigating the problem is that firms and inventors will face a higher cost of external capital for R&D.

Because the knowledge asset created by R&D investment is intangible, partly embedded in human capital, and ordinarily very specialized to the particular firm in which it resides, the capital structure of R&D intensive firms customarily exhibits considerably less leverage than that of other firms, see Titman and Wessels (1988). The logic is that the lack of a secondary market for R&D and the non-collaterability of R&D activity mitigates against debt-financed R&D activity. Aboody and Lev (2000) argue that because of the relative uniqueness of R&D, which makes it difficult for outsiders to learn about the productivity and value of a given firm's R&D from the performance and products of other firms in the industry, the extent of information asymmetry associated with R&D is larger than that associated with investment in tangible (e.g., property, plant, and equipment) and financial assets. Bond holders, ceteris paribus, may be unwilling to hold the risks associated with greater R&D activity. BFP studying a panel of R&D intensive firms, find that equity, when more easily available, might be preferred to debt as a means of financing R&D.

Brown et al. (2012), Hall and Lerner (2010) and BFP point out that most of the R&D spending is in the form of payments to highly skilled workers, who often require a great deal of firm-specific knowledge and training. The effort of the skilled workers create the knowledge base of the firm, and is therefore embedded in the human capital of the firms. This knowledge base is lost once workers get laid off. The implication of this is that R&D intensive firms behave as if they faced large adjustment costs and therefore chose to smooth their R&D spending. Thus R&D intensive firms that face financing frictions, to smooth R&D relative to transitory finance shocks, build and manage internal buffer stocks of liquidity (e.g., cash reserves). Gamba and Triantis (2008) point out that cash balances, which give financial flexibility to firms, are held when external finance is costly and/or income uncertainty is high. With higher liquidity reserve firms can counter bad shocks by draining it.

Now, given the nature of R&D activity that makes borrowing costly, internal funds may be more preferable. Therefore, innovative firms, ceteris paribus, are less likely to distribute cash as dividends. Both Carpenter and Petersen (2002) and Chan et al. (2001) studying R&D intensive firms from COMPUSTAT files find that R&D intensive firms pay little or no dividend, indicating that most firms retain essentially all of their internal funds. In our data set we too find that, on average, innovating firms pay less dividends than non-innovating firms.

In this paper we study a firm's decision to innovate and the financing choices of a panel of Dutch firms observed over three waves. While there are many studies that have explored a firm's choice to innovate in the Schumpeterian tradition, few have considered how financing and innovation choices are related. We formally model the decision to innovate as

$$I_t = 1\{I_t^*(\textit{Long-term Debt}, \textit{Liquidity Reserve}, \textit{Dividend}, \textit{Controls}, \tilde{\alpha}, \upsilon_t) > 0\}, \tag{2.1}$$

where $I = 1\{.\}$ is an indicator function that takes value 1 if the latent variable $I_t^*(.) > 0$. $\tilde{\alpha}$ is the unobserved heterogeneity, υ_t the idiosyncratic term, and $\textit{Controls}$ being the traditional control variables. We term equation (2.1) as the Innovation equation. Given the above discussion, we should expect that, ceteris paribus, firms with higher long-term debt in their capital structure, firms that maintain low liquidity reserve, and firms that pay out dividends to be less likely to engage in innovative activity. We do not contend that other consideration such as taxes or issuance cost do not affect financial decisions. We also know that financing and investment decisions are history dependent and are forward looking. However, ceteris paribus, across time and firms one should expect the above hypothesized relationships to hold on an average.

B. *Financial Constraints and Innovation*

Papers, such as CQ, AH, and CH, studying firm dynamics look at how financial constraint and capital structure affect firm growth and survival. These papers have shown that financing constraint and financing and investment decisions are not uniform across firm characteristics such as size and age. Now, it is well known that innovation too affects growth and survival of firms (see KK), and that R&D effort is marred by various kinds uncertainties (see Berk et al., 2004) unique to the innovation process. Hence, a firm engaging in R&D will have its equity value affected, with implications for borrowing constraint, state contingent growth trajectory and future financing and innovation decision.

Therefore, while the unconditional relation between financing and innovation, discussed in the last subsection, could be expected to be true, under financial constraint, firms could, depending on the extent of constraint, opt for a innovation and financing policy different from when they are unconstrained. This could be ascertained by looking at how the decision of a firm to engage in innovative activity changes by changing the financial policy of the firm under varying degrees of financial constraint. To achieve this end, we start by studying how financial constraint arise for firms that report that they are financially constrained.

To formalize, we denote by F_{it}, which takes value 1 if the firm i reports that it is financially constrained in time period t. Now, see HW, a firm may be constrained both because of high cost of external funds and/or because of high need for external funds. Thus, when a firm reports that it is financially constrained, $F_{it} = 1$, it could be because it is required to pay a high premium, which could be higher for firms engaging in R&D

activity, on scarce external finance or because it is unable to access external funds. The premium, for example, could reflect bankruptcy cost (see Gale and Hellwig, 1985) or the cost of floating equity as in HW and CQ. In AH and CH this premium is formalized as higher repayment schedule to lenders as a fraction of its profits during such time as when the firm faces borrowing constraint and short-term capital advancement are low. Also, for a given financial state of a firm, higher expectation of profits from R&D activity will drive up the demand for R&D investment, creating a gap between desired and available funds, which in turn will cause the firm to report itself as being financially constrained. Hence, in our explanation of how financial constraint arise, we will need to control for future expected profitability.

Barring a few that have been documented in Hall and Lerner (2010), most papers in empirical corporate finance study corporate financing and firm level investment. Now, financing frictions with respect to R&D activity, which for reasons discussed earlier, can be acute when compared to financing investment in physical capital. Consequently, innovative firms might find themselves more constrained than those that are not. To test this, like Almeida and Campello (2007), we test if asset intangibility, which is higher for innovating firms and which limits the debt capacity of firms, have a bearing on the firms reporting financial constraint.

Formally, we model financial constraint as

$$F_t = 1\{F_t^*(Financial\ State\ \ Variables, Expected\ Profitability, Controls, \tilde{\alpha}, \zeta_t) > 0\}, \tag{2.2}$$

where $\tilde{\alpha}$ is unobserved heterogeneity and ζ_t is the idiosyncratic component of the Financial Constraint equation. As in Whited and Wu (2005) and Gomes et al. (2006), where the shadow price of scarce external finance in the firm's intertemporal optimization problem is assumed to be a function of observable variables, we hypothesize that the latent variable F_t^*, which captures the premium on external finance and the gap in financing, to be a function of observable and endogenously determined financial state variables. HW give a detailed discussion on constraint proxies that reflect high cost or high need for external finance. Our specification, discussed later, to explain financial constraint is rich enough to capture both aspects, high cost as well as high need for external finance.

Now, to return to the question of innovation and financing policy under financial constraint across firm characteristics, we look at how the propensity to innovate under financial constraint, both of which are determined endogenously, changes with endogenous financing policy, say an increase in long-term debt, of the firm. To put it formally, we look at how $\Pr(I = 1|F = 1)$ and $\Pr(I = 1|F = 0)$ changes with debt policy at different level of firm characteristics, such as size of the firm. These firm characteristics also indicate the extent of constraint the firm faces, so in effect by studying how $\Pr(I = 1|F = 1)$ changes with the financing policy of the firm at different level of firm characteristics, we are looking at how $\Pr(I = 1|F = 1)$ changes with the financing policy at different level of constraint.

C. *Financial Constraints and R&D Investment*

Beginning with Fazzari et al. (1988) there has been a huge amount of literature that has sought to test for financing frictions and quantifying the extent of market failure in company level investment due to the presence of financing frictions. A survey of this literature is beyond the scope of this paper. However, as Brown et al. (2012) point out there aren't many papers that have looked at financing frictions and R&D investment. Few papers that have studied the implication of financial constraint for R&D investment have been surveyed in Hall and Lerner (2010).

Empirical study of the effect of financing frictions on investment has broadly followed two approaches. One approach is to ad hoc classify firms into those that are financially constrained and those that are not, and specify a reduced form accelerator type model for the constrained and unconstrained firms. The extent of financing frictions, controlling for the investment opportunity, is judged by the sensitivity of investment to cash flow. Another approach, which is more structural, is to estimate Euler equations derived from standard intertemporal investment model augmented with financial state variables to account for financial frictions, where external financing constraint affect the intertemporal substitution of investment today for investment tomorrow, via the shadow value of scarce external funds, (see Whited and Wu, 2005). The few empirical studies on financing frictions and R&D investment, broadly speaking, follow these two approaches.

In this paper, besides studying financing and innovation decisions of firms under financial constraint across firm characteristic, we also study how financial constraint affect R&D investment, which is observed conditional on firms choosing to innovate, $I_t = 1$. We posit that the observed R&D intensity, measured as a ratio of R&D investment to total capital asset, for a firm i, can be explained by estimating the following R&D equation:

$$R_t = R_t(\textit{Financial Constraint}, \textit{Expected Profitability}, \textit{Controls}, \tilde{\alpha}, \eta_t) \text{ if } I_t = 1, \tag{2.3}$$

where $\tilde{\alpha}$ is the unobserved heterogeneity, η_t the idiosyncratic component. The specification is motivated by the fact that financing frictions, which could be either due to high cost of external funds or due to lack of access to it, is summarized by the reported financial constraint, F_t. Thus, given future expected profitability and other controls, we can gauge the extent of market failure for R&D investment due the presence of financing frictions by estimating the metric,

$$\mathrm{E}[R_t(F_t = 0)|I_t = 1] - \mathrm{E}[R_t(F_t = 1)|I_t = 1].$$

This metric could be construed as the difference between first best R&D investment and optimal R&D investment under financing constraint.

Using firm's assessment of being financially constrained avoids the need to ad hoc classify the firms into constrained and unconstrained firms. Moreover, papers that a priori classify firms as constrained and unconstrained assume financial constraint faced by firms to be exogenous to investment decisions. In assessing the impact of reported financial constraint, $F_{it} = 1$, on R&D expenditure, ours is a departure from the reduced form accelerator type models, about which questions have been raised as to whether such a procedure can indeed identify the extent of financing frictions, (see Kaplan and Zingales 1997; Gomes 2001; and HW). We address the issue of endogeneity of financial constraint by estimating simultaneously the Innovation equation (2.1), the Financial Constraint equation (2.2) and the R&D equation along with the equations for the financing choice made by the firms. Thus, in contrast to reduced form models, ours is a more structural approach.

Our frame work for studying the effect of financing constraint on R&D in essence is a static one. Though one could derive a dynamic empirical model for R&D investment from a firm's dynamic optimization problem with adjustment cost where the firm is subject to external financing constraints, or employ indirect inference approach as in Whited (2006) and HW to test for financing frictions and its implication for R&D investment, we avoid this route for two reasons. The first being, as we explain when discussing our data, that in our data set we observe R&D investment every alternative year, which precludes us from estimating a dynamic empirical model of R&D investment, at least in the classical regression framework. The second reason is that, since firms tend to smooth R&D investment over time, adjustment costs, for firms that have decided to engage in R&D in the past, is unlikely to be a substantial factor in explaining R&D investment[1]. We believe that, given our comprehensive treatment of heterogeneity and endogeneity, a misspecification due to omission of adjustment cost should be taken care of.

Also, using the binary indicator on financial constraint as reported by firms allows us to generalize the R&D equation (2.3) to a switching regression model, where the endogenous financial constraint equation sorts the firms over the two different regimes, financially constrained and unconstrained. This allows us to investigate how firms with different characteristics, such as maturity and size, invest in R&D under financial constraint and under no constraint. In doing so we are able to underscore that financing frictions condition firm dynamics, which are brought about through R&D investment.

3. EMPIRICAL MODEL

The usual problem faced in any empirical exercise is that of accounting for heterogeneity and endogeneity. For the problem at hand, we know that the decision to innovate, the financial choices made, the financial constraint faced, and the amount to invest in R&D are all endogenously determined. In this paper we develop a control function approach to address the issue of heterogeneity and endogeneity. In this section we introduce our empirical model, the model assumptions, and some results. Technical details on identification of

[1] It is also possible that new innovators bear sunk cost of investment and that starting to innovate involves costly learning, giving rise to non-convex adjustment cost, which can interact with financing friction to alter the timing of R&D investment. However, estimating parameters of interest of a model that allows for sunk cost of investment that interacts with financing frictions to affect R&D investment would most likely involve a different econometric approach, such as in Cooper and Haltiwanger (2006) or HW, and this is beyond the scope of our paper.

structural parameters of interest has been discussed in the Appendix.

To study the effect of endogenous financial constraint on R&D expenditure, the endogenous decision to innovate, and to account for the fact that R&D expenditure is observed only for firms that opt to innovate, the three structural equations – Innovation, Financial Constraint, and R&D – introduced in section 2 are

$$I_{it} = 1\{I_{it}^{*} = \mathcal{X}_{it}^{I\prime}\boldsymbol{\gamma} + \theta\tilde{\alpha}_i + \upsilon_{it} > 0\}, \tag{3.1}$$

$$F_{it} = 1\{F_{it}^{*} = \mathcal{X}_{it}^{F\prime}\boldsymbol{\varphi} + \lambda\tilde{\alpha}_i + \zeta_{it} > 0\}, \tag{3.2}$$

$$\begin{aligned} R_{it} &= F_{it}(\beta_f F_{it} + \mathcal{X}_{it}^{R\prime}\boldsymbol{\beta}_1 + \mu_1\tilde{\alpha}_i + \eta_{1it}) + (1 - F_{it})(\mathcal{X}_{it}^{R\prime}\boldsymbol{\beta}_0 + \mu_0\tilde{\alpha}_i + \eta_{0it}) \text{ if } I_{it} = 1 \\ &= F_{it}R_{1it} + (1 - F_{it})R_{0it} \text{ if } I_{it} = 1, \end{aligned} \tag{3.3}$$

where I_t is an indicator variable that takes value 1 if the firm i decides to innovate, F_t takes value 1 if firm i experiences financial constraint, and R_t is the observed R&D intensity, defined as the ratio of total R&D expenditure to total capital assets (tangible + intangible), if the firm decides to innovate[2]. To allow for the effect of $\mathcal{X}_t^R$ to be different in the two regimes, financially constrained and unconstrained, we model equation (3.3) as an endogenous switching regression model, where the Financial Constraint equation sorts the firms over the two different regimes. That is,

$$R_t = R_{1t} = \beta_f F_t + \mathcal{X}_t^{R\prime}\boldsymbol{\beta}_1 + \mu_1\tilde{\alpha} + \eta_{1t} \text{ if } F_t = 1 \text{ and } I_t = 1$$

and

$$R_t = R_{0t} = \mathcal{X}_t^{r\prime}\boldsymbol{\beta}_0 + \mu_0\tilde{\alpha} + \eta_{0t} \text{ if } F_t = 0 \text{ and } I_t = 1.$$

In the above set of equations $\mathcal{X}_t^I = \{\mathbf{z}_t^{I\prime}, \mathbf{x}_t^{I\prime}\}'$, $\mathcal{X}_t^F = \{\mathbf{z}_t^{F\prime}, \mathbf{x}_t^{F\prime}\}'$, and $\mathcal{X}_t^R = \{\mathbf{z}_t^{R\prime}, \mathbf{x}_t^{R\prime}\}'$, where conditional on unobserved heterogeneity $\tilde{\alpha}_i$, each of the $\mathbf{z}_t$ is a vector of exogenous variables. That is, $\upsilon_t|\tilde{\alpha}_i, \mathbf{z}_t^I \sim \upsilon_t|\tilde{\alpha}_i$; the same being true for the Financial Constraint and R&D equation. Each of the $\mathbf{x}_t$, is a vector of endogenous variables, that is, $\mathrm{E}(\upsilon_t|\tilde{\alpha}, \mathbf{x}_t^I) \neq 0$, the same holds for the Financial Constraint and R&D equation.

Simultaneity in the decision to innovate, the financial constrained faced, and the amount to expend in R&D investment is captured by the fact that unobserved heterogeneity that affects the decision to innovate also affects the constraint faced and R&D investment, that

[2] In the rest of the paper unless otherwise needed we drop the firm script i.

the unobserved idiosyncratic components in each of the equations are correlated with each other, and certain observable variables are common to the structural and reduced form equations. However, because $\mathbf{x}_t$s are endogenous, estimating the system of equations will give inconsistent results.

To obtain the consistent estimates for the structural equations we adopt a control function approach, which involves a multi-step procedure. In the first step we estimate

$$\mathbf{x}_{it} = \mathbf{Z}_{it}'\boldsymbol{\delta} + \tilde{\alpha}_i\boldsymbol{\kappa} + \boldsymbol{\epsilon}_{it}, \tag{3.4}$$

which is the system of 'm' equations written in a reduced form for the endogenous variables $\mathbf{x}_t = (x_{1t}, \ldots, x_{mt})'$, where every component of $\mathbf{x}_t^I$, $\mathbf{x}_t^F$, and $\mathbf{x}_t^R$ is also a component of $\mathbf{x}_t$. $\mathbf{Z}_t = \text{diag}(\mathbf{z}_{1t}, \ldots, \mathbf{z}_{mt})$ is the matrix of exogenous variables or instruments appearing in each of the m reduced form equations in (3.4) and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_m')'$. Let $\mathbf{z}_t$ be the union of all exogenous variables appearing in each of $\mathbf{z}_t^I$, $\mathbf{z}_t^F$, and $\mathbf{z}_t^R$. For every $l \in (1, \ldots, m)$, $\mathbf{z}_{lt} = \mathcal{Z}_t = (\mathbf{z}_t', \tilde{\mathbf{z}}_t')'$, where the dimension of vector of instruments, $\tilde{\mathbf{z}}$, is greater than or equal to the dimension $\mathbf{x}$. This is the crucial identifying condition, see Blundell and Powell (2003) for details. Also define $\mathbf{X}_i = \{\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT_i}'\}'$ and $\mathcal{Z}_i = (\mathcal{Z}_{i1}' \ldots \mathcal{Z}_{iT_i}')'$.

$\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})'$ is the vector of idiosyncratic component and $\tilde{\alpha}$, the unobserved individual effect for firm i, which we model as a random effect, is correlated with $\mathcal{Z}_i$. But, conditional on $\tilde{\alpha}$, $\mathcal{Z}_i$ is assumed to be independent of η_{1t}, η_{0t}, ζ_t, υ_t, and $\boldsymbol{\epsilon}_t$. Since the unobserved individual specific effect affects the endogenous regressors as well as the firm's innovation decision and it being financially constrained, to account for simultaneity that arises due to unobserved heterogeneity, we therefore have different factor loadings, such as, $\{\kappa_1 \ldots, \kappa_m\}$, that appear in the reduced form equations, and θ, λ, μ_0, and μ_1, that appear in the structural equations.

The above structural equations – (3.1), (3.2), and (3.3) – can be succinctly written as

$$\mathbf{y}_t^* = \mathbb{X}_t'\mathbf{B} + \tilde{\alpha}\mathbf{k} + \Upsilon_t, \tag{3.5}$$

where $\mathbf{y}_t^* = \{I_t^*, F_t^*, I_tF_tR_{1t}, I_t(1-F_t)R_{0t},\}'$. $\mathbb{X}_t = \text{diag}(\mathcal{X}_t^I,\ \mathcal{X}_t^F,\ \mathcal{X}_{1t}^R,\ \mathcal{X}_{0t}^R)$, where $\mathcal{X}_{1t}^R = \{F_t, I_tF_t\mathcal{X}_t^{R\prime}\}'$ and $\mathcal{X}_{0t}^R = I_t(1-F_t)\mathcal{X}_t^R$. $\mathbf{B}$ in (3.5) is given by $\mathbf{B} = \{\boldsymbol{\gamma}', \boldsymbol{\varphi}', \beta_f, \boldsymbol{\beta}_1', \boldsymbol{\beta}_0'\}'$. Finally, $\mathbf{k} = \{\theta, \lambda, \mu_1, \mu_0\}'$ and $\Upsilon_t = \{\upsilon_t, \zeta_t, \eta_{1t}, \eta_{0t}\}'$.

Some of the distributional assumptions that will eventually allow us to construct the control functions, which correct for the bias due to the endogeneity of $\mathbf{x}_t$ and help us identify the structural parameters of interest are:

$\mathcal{A}1$. $\Upsilon_{it}|\tilde{\alpha}_i, \mathcal{Z}_i \sim \Upsilon_{it}|\tilde{\alpha}_i$ and $\boldsymbol{\epsilon}_{it}|\tilde{\alpha}_i, \mathcal{Z}_i \sim \boldsymbol{\epsilon}_{it}$,

$\mathcal{A}2$. $\Upsilon_{it}|\tilde{\alpha}_i, \boldsymbol{\epsilon}_i \sim \Upsilon_{it}|\boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i = \{\boldsymbol{\epsilon}'_{i1} \ldots \boldsymbol{\epsilon}'_{iT_i}\}'$, and

$\mathcal{A}3$. The error terms Υ_{it} and $\boldsymbol{\epsilon}_{it}$ are i.i.d.[3] and

$$\begin{pmatrix} \Upsilon_{it} \\ \boldsymbol{\epsilon}_{it} \end{pmatrix} \sim \mathrm{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \Sigma_{\Upsilon\Upsilon} & \Sigma_{\Upsilon\epsilon} \\ \Sigma_{\epsilon\Upsilon} & \Sigma_{\epsilon\epsilon} \end{pmatrix} \right].$$

According to assumption $\mathcal{A}1$, conditional on $\tilde{\alpha}$, $\mathcal{Z}$ is independent of Υ_t, which is a standard assumption made in the literature. In $\mathcal{A}2$ by assuming Υ_t to be independent of $\tilde{\alpha}$ conditional on reduced form errors $\boldsymbol{\epsilon}$, we weaken the standard assumption of independence of Υ_t and $\tilde{\alpha}$.

As stated earlier, to estimate the structural parameters of interest in equation (3.5), a multi-step estimation procedure has been proposed. In the first stage the parameters, Θ_1, of the system of reduced form equations, equation (3.4), is estimated. In the subsequent stages additional correction terms or "control variables", obtained from the first stage reduced form estimates, correct for the bias due to endogeneity of the $\mathbf{x}_t$. We study the identification and estimation of structural parameters for nonlinear response models and show the construction of correction terms in subsection B and, in detail, in Appendix A. But before we discuss identification of structural parameters, we first discuss the estimation of the parameters of the reduced form equation.

A. *Estimation of the First Stage Reduced Form Equations*

In the first stage we estimate the system of reduced form equations (3.4). Since $\tilde{\alpha}_i$ and $\mathcal{Z}_i$ are correlated in order to estimate $\boldsymbol{\delta}$, $\Sigma_{\epsilon\epsilon}$, and $\boldsymbol{\kappa}$ consistently, we use Mundlak's (1978) correlated random effects formulation. We assume that

$$\mathcal{A}4. \quad \mathrm{E}(\tilde{\alpha}_i|\mathcal{Z}_i) = \bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}, \tag{3.6}$$

[3] Though the i.i.d. assumption is not strictly necessary, and can be relaxed.

where $\bar{\mathcal{Z}}_i$, is the mean of time-varying variables in $\mathcal{Z}_{it}$. We also assume that

$$\mathcal{A}5. \quad \tilde{\alpha}_i|\mathcal{Z}_i \sim \mathrm{N}\left[\mathrm{E}(\tilde{\alpha}_i|\mathcal{Z}_i), \sigma_\alpha^2\right], \tag{3.7}$$

so that the tail, $\alpha_i = \tilde{\alpha}_i - \mathrm{E}(\tilde{\alpha}_i|\mathcal{Z}_i) = \tilde{\alpha}_i - \bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}}$, is distributed normally with mean zero and variance σ_α^2, and is also assumed to be independent of $\mathcal{Z}_i$. Given the above, equation (3.4) can now be written as

$$\mathbf{x}_{it} = \mathbf{Z}_{it}'\boldsymbol{\delta} + (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \alpha_i)\boldsymbol{\kappa} + \boldsymbol{\epsilon}_{it}. \tag{3.4a}$$

To consistently estimate the reduced form parameters, $\Theta_1 = \{\boldsymbol{\delta}', \bar{\boldsymbol{\delta}}', \mathrm{vech}(\Sigma_{\epsilon\epsilon})', \boldsymbol{\kappa}', \sigma_\alpha\}'$, we employ the technique of step-wise maximum likelihood method in Biørn (2004). However, our model differs from Biørn. While Biørn estimates the covariance matrix Σ_α of $\boldsymbol{\alpha}_i = \{\alpha_{1i}, \ldots, \alpha_{mi}\}'$, where each of the α_{li}, $l \in \{1, \ldots, m\}$, is unrestricted, we place the restriction $\alpha_{li} = \kappa_l \alpha_i$. This implies that

$$\Sigma_\alpha = \sigma_\alpha^2 \Sigma_\kappa = \sigma_\alpha^2 \begin{pmatrix} \kappa_1^2 & & & \\ \kappa_1\kappa_2 & \kappa_2^2 & & \\ \vdots & \vdots & & \\ \kappa_1\kappa_m & \kappa_2\kappa_m & \ldots & \kappa_m^2 \end{pmatrix}.$$

Moreover, as can be seen from the modified equation (3.4a), we also impose the restriction that $\bar{\boldsymbol{\delta}}$ remains the same across each of the m reduced form equations. In Appendix B we provide a note on the estimation strategy employed to estimate the parameters of the reduced form equations.

B. *Identification and Estimation of the Structural Parameters*

Given the above set of assumptions we have

$$\begin{aligned} \Upsilon_t|\mathbf{X}, \mathcal{Z}, \tilde{\alpha} &\sim \Upsilon_t|\mathbf{X} - \mathrm{E}(\mathbf{X}|\mathcal{Z}, \tilde{\alpha}), \mathcal{Z}, \tilde{\alpha} \\ &\sim \Upsilon_t|\boldsymbol{\epsilon}, \mathcal{Z}, \tilde{\alpha} \\ &\sim \Upsilon_t|\boldsymbol{\epsilon}, \tilde{\alpha} \\ &\sim \Upsilon_t|\boldsymbol{\epsilon}, \end{aligned} \tag{3.8}$$

where the second equality in distribution follows from the fact that $\mathbf{X}_i - \mathrm{E}(\mathbf{X}_i|\mathcal{Z}_i, \tilde{\alpha}_i) = \boldsymbol{\epsilon}_i$, the third follows from $\mathcal{A}1$, and the fourth from assumption $\mathcal{A}2$. According to the above, the dependence of the structural error term Υ_t on $\mathbf{X}$, $\mathcal{Z}$, and $\tilde{\alpha}$ is completely characterized by the reduced form errors $\boldsymbol{\epsilon}$. The expectation of Υ_t given $\boldsymbol{\epsilon}$ is given by

$$E(\Upsilon_t|\boldsymbol{\epsilon}) = E(\Upsilon_t|\boldsymbol{\epsilon}_t) = \Sigma_{\Upsilon\epsilon}\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t = \tilde{\Sigma}_{\Upsilon\epsilon}\Sigma_{\epsilon}\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t = \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t, \tag{3.9}$$

where the first equality follows from the assumption that conditional on $\boldsymbol{\epsilon}_{it}$, Υ_{it} is independent of $\boldsymbol{\epsilon}_{i_{-t}}$. This assumption has also been made in Papke and Wooldridge (2008), and Semykina and Wooldridge (2010). The $(4 \times m)$ matrices $\tilde{\Sigma}_{\Upsilon\epsilon}$ in the fourth equality is

$$\tilde{\Sigma}_{\Upsilon\epsilon} = \begin{pmatrix} \rho_{\eta_1\epsilon_1}\sigma_{\eta_1} & \dots & \rho_{\eta_1\epsilon_m}\sigma_{\eta_1} \\ \rho_{\eta_0\epsilon_1}\sigma_{\eta_0} & \dots & \rho_{\eta_0\epsilon_m}\sigma_{\eta_0} \\ \rho_{\zeta\epsilon_1}\sigma_{\zeta} & \dots & \rho_{\zeta\epsilon_m}\sigma_{\zeta} \\ \rho_{\upsilon\epsilon_1}\sigma_{\upsilon} & \dots & \rho_{\upsilon\epsilon_m}\sigma_{\upsilon} \end{pmatrix}$$

and the $(m \times m)$ matrix Σ_ϵ is $\mathrm{diag}(\sigma_{\epsilon 1}, \dots, \sigma_{\epsilon m})$, so that $\tilde{\Sigma}_{\Upsilon\epsilon}\Sigma_\epsilon = \Sigma_{\Upsilon\epsilon}$. Finally, in the last equality $\tilde{\Sigma}_{\epsilon\epsilon}^{-1} = \Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}$. We prefer to write the above conditional expectation as $\mathrm{E}(\Upsilon_t|\boldsymbol{\epsilon}_t) = \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t$ because the elements of $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}$ are obtained from the estimates of the first stage reduced form estimation of our sequential estimation procedure, and the formulation in (3.9) helps us distinguish the parameters that are estimated in the first stage from those that are estimated in the subsequent stages. Also, as we will see, it is the elements of $\tilde{\Sigma}_{\Upsilon\epsilon}$, which are estimated in the subsequent stages that give us the test of exogeneity of $\mathbf{x}_t$ with respect to Υ_t .

Given assumptions $\mathcal{A}4$ and $\mathcal{A}5$ and equations (3.8) and (3.9), we can write the expectation of $\mathbf{y}_t^*$ given $\mathbf{X}$, $\mathcal{Z}$, and $\tilde{\alpha}$

$$\begin{aligned} \mathrm{E}(\mathbf{y}_t^*|\mathbf{X}, \mathcal{Z}, \tilde{\alpha}) &= \mathbb{X}_t'\mathbf{B} + \tilde{\alpha}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t \\ &= \mathbb{X}_t'\mathbf{B} + (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \alpha)\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t = \mathrm{E}(\mathbf{y}_t^*|\mathbf{X}, \mathcal{Z}, \alpha) \end{aligned} \tag{3.10}$$

To estimate the system of equations in (3.10) the standard technique is to replace $\boldsymbol{\epsilon}_t$ by the residuals from the first stage reduced form regression, here equation (3.4a). However, the residuals $\mathbf{x}_t - \mathrm{E}(\mathbf{x}_t|\mathcal{Z}, \alpha) = \mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \alpha)\boldsymbol{\kappa}$, remain unidentified because the α's are unobserved even though the reduced form parameters, $\Theta_1 = \{\boldsymbol{\delta}', \bar{\boldsymbol{\delta}}', \boldsymbol{\kappa}', \sigma_\alpha\}'$, can be consistently estimated for the first stage estimation. From the results on identification of structural parameters derived in Appendix A, it can be shown that

$$\mathrm{E}(\mathbf{y}_t^*|\mathbf{X},\mathcal{Z}) = \int \mathrm{E}(\mathbf{y}_t^*|\mathbf{X},\mathcal{Z},\alpha)f(\alpha|\mathbf{X},\mathcal{Z})d\alpha = \mathbb{X}_t'\mathbf{B} + (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha})\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t, \tag{3.11}$$

where $\hat{\alpha}_i(\Theta_1,\mathbf{X}_i,\mathcal{Z}_i) = \mathrm{E}(\alpha_i|\mathbf{X}_i,\mathcal{Z}_i)$, as discussed in Appendix A, is the Expected a Posteriori (EAP) value of α_i and $\hat{\boldsymbol{\epsilon}}_{it}(\Theta_1,\mathbf{X}_i,\mathcal{Z}_i) = \mathbf{x}_{it} - \mathrm{E}(\mathbf{x}_{it}|\mathbf{X}_i,\mathcal{Z}_i) = \mathbf{x}_{it} - \mathbf{Z}_{it}'\boldsymbol{\delta} - \boldsymbol{\kappa}(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)$. $\hat{\tilde{\alpha}}_i = \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$ are the "control functions" that correct for the bias which arises due to the correlation of $\mathbf{x}_t$ with α and Υ_t. The correlation of the exogenous variables $\mathcal{Z}_t$ with $\tilde{\alpha}$, is accounted by $\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha}$. In Appendix A we show how to construct $\hat{\alpha}_i$. Given (3.11) we can write the projection of $\mathbf{y}_{it}^*$ given $\mathbf{X}_i$, $\mathcal{Z}_i$ in error form as

$$I_t = 1\{I_t^* = \mathcal{X}_t^{I\prime}\boldsymbol{\gamma} + \theta\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \tilde{\upsilon}_t > 0\}, \tag{3.12}$$

$$F_t = 1\{F_t^* = \mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \tilde{\zeta}_t > 0\}, \tag{3.13}$$

$$\begin{aligned} R_t =& F_t(\beta_f F_t + \mathcal{X}_t^{R\prime}\boldsymbol{\beta}_1 + \mu_1\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\eta_1\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \tilde{\eta}_{1t}) \\ &+ (1-F_t)(\mathcal{X}_t^{R\prime}\boldsymbol{\beta}_0 + \mu_0\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\eta_0\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \tilde{\eta}_{0t}) \text{ if } I_t = 1 \end{aligned} \tag{3.14}$$

where $\tilde{\Upsilon}_t = \{\tilde{\upsilon}_t, \tilde{\zeta}_t, \tilde{\eta}_{1t}, \tilde{\eta}_{0t}\}'$, defined in Appendix A, is normally distributed with mean 0, variance $\Sigma_{\tilde{\Upsilon}\tilde{\Upsilon}}$, and is independent of $\mathbf{X}$ and $\mathcal{Z}$. We would like to state here that in the modified Innovation equation (3.12),

$$\tilde{\Sigma}_{\upsilon\epsilon} = \{\rho_{\upsilon\epsilon_1}\sigma_\upsilon, \ldots, \rho_{\upsilon\epsilon_m}\sigma_\upsilon\}',$$

where $\rho_{\upsilon\epsilon_1}\sigma_\upsilon$, for example, gives a measure of correlation between x_1 and υ, thus providing us a test of exogeneity of x_1 in the Innovation equation. Similarly, the estimates of $\tilde{\Sigma}_{\zeta\epsilon}$ and $\tilde{\Sigma}_{\eta\epsilon}$ give us a test of exogeneity of $\mathbf{x}$ in the Financial Constraint and the R&D equation respectively.

Given $\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}+\hat{\alpha}$ and $\hat{\boldsymbol{\epsilon}}_t$, it may be possible to consistently estimate the structural parameters of interest by specifying a joint likelihood for I_t, F_t, and R_t. However, given the presence of nonlinearities in the model, the likelihood function will be difficult to optimize. Given this fact, we estimate the structural parameters of interest in equations (3.12) to (3.14) in two steps after the first stage reduced form estimation. In the second stage, given the estimates of the control functions $\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha}$ and $\hat{\boldsymbol{\epsilon}}_t$, we estimate jointly the structural parameters, Θ_2, of the Innovation equation (3.12) and the Financial Constraint equation (3.13). Then in the third stage, given the control function and second stage estimates, we the estimate the R&D equation (3.14).

Estimating the parameters of the second, Θ_2, and third, Θ_3, stage, given the first stage consistent estimates $\hat{\Theta}_1$, is asymptotically equivalent to estimating the subsequent stage parameters had the true value of Θ_1 been known. To obtain correct inference about the structural parameters, Θ_2 and Θ_3, one has to account for the fact that instead of true values of first stage reduced form parameters, we use their estimated value. In Appendix D we provide analytical expression for the error adjusted covariance matrix for the estimates of the structural parameters.

B.1. *The Second Stage: Estimation of the Innovation and the Financial Constraint Equations*

Given the modified Innovation equation (3.12) and the modified Financial Constraint equation (3.13), the conditional log likelihood function for firm i in period t given $\mathbf{X}$, $\mathcal{Z}$, if the time period t corresponds to CIS3 and CIS3.5[4], is given by

$$\begin{aligned} \mathcal{L}_{t2}(\Theta_2|\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) &= I_t F_t \ln(\Pr(I_t = 1, F_t = 1)) + I_t(1 - F_t)\ln(\Pr(I_t = 1, F_t = 0)) \\ &+ F_t(1 - I_t)\ln(\Pr(I_t = 0, F_t = 1)) + (1 - F_t)(1 - I_t)\ln(\Pr(I_t = 0, F_t = 0)). \end{aligned} \tag{3.15}$$

For CIS2.5, since we do not observe whether a firm is financially constrained or not for the non-innovating firms, for time period t corresponding to CIS2.5, we have

$$\begin{aligned} &\mathcal{L}_{t2}(\Theta_2|\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \\ &F_t I_t \ln(\Pr(I_t = 1, F_t = 1)) + (1 - F_t) I_t \ln(\Pr(I_t = 1, F_t = 0)) + (1 - I_t)\ln(\Pr(I_t = 0)). \end{aligned} \tag{3.16}$$

[4]For our empirical analysis, as discussed in next Section on data, we use three waves of Dutch Community Innovation Survey (CIS). For CIS3 and CIS3.5 we observe if the firm is financially constrained for both the innovating and the non-innovating firms, but for CIS2.5 the information on financial constraint is given for only the innovating firms.

In the above two equations

$$
\begin{aligned}
&\Pr(I_t = 1, F_t = 1) = \Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{\upsilon}}), \quad \Pr(I_t = 1, F_t = 0) = \Phi_2(-\varphi_t, \gamma_t, -\rho_{\tilde{\zeta}\tilde{\upsilon}}), \\
&\Pr(I_t = 0, F_t = 1) = \Phi_2(\varphi_t, -\gamma_t, -\rho_{\tilde{\zeta}\tilde{\upsilon}}), \quad \Pr(I_t = 0, F_t = 0) = \Phi_2(-\varphi_t, -\gamma_t, \rho_{\tilde{\zeta}\tilde{\upsilon}}), \\
&\text{ and } \Pr(I_t = 0) = \Phi(-\gamma_t),
\end{aligned}
$$

where Φ_2 is the cumulative distribution function of a standard bivariate normal, $\rho_{\tilde{\zeta}\tilde{\upsilon}}$ is the correlation of $\tilde{\zeta}_t$ and $\tilde{\upsilon}_t$,

$$
\gamma_t = (\mathcal{X}_t^{I\prime}\boldsymbol{\gamma} + \theta\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t)\frac{1}{\sigma_{\tilde{\upsilon}}} \quad \varphi_t = (\mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t)\frac{1}{\sigma_{\tilde{\zeta}}}, \tag{3.17}
$$

and $\Theta_2 = \{\boldsymbol{\varphi}', \lambda, \tilde{\Sigma}_{\zeta\epsilon}, \boldsymbol{\gamma}', \theta, \tilde{\Sigma}_{\upsilon\epsilon}, \rho_{\tilde{\zeta}\tilde{\upsilon}}\}'$. The log likelihood of the second stage parameters is given by

$$
\mathcal{L}_2(\Theta_2) = \sum_{i=1}^{N}\sum_{t=1}^{T_i} \mathcal{L}_{it2}(\Theta_2|\hat{\tilde{\alpha}}_i, \hat{\boldsymbol{\epsilon}}_{it}). \tag{3.18}
$$

It should be noted that all the parameters of the structural equations (3.12) and (3.13) can only be identified up to a scale, the scaling factor for the financial constraint equation and selection equation being respectively $\sigma_{\tilde{\zeta}}$ and $\sigma_{\tilde{\upsilon}}$. In what follows, with a slight abuse of notation, we will denote the scaled parameters of the second stage estimation by their original notation. Computation of the standard errors of the structural parameters of this second stage, Θ_2, has been discussed in Appendix D.

Now, given the first stage estimates $\hat{\Theta}_1$, we can obtain the estimates of the control functions $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$, which can then be used in the above likelihood function to obtain consistent estimates for the second stage parameters. The true measure, however, of the effect of a certain variable, w, on the probability of engaging in innovation or the probability of being financially constrained is the Average Partial Effect (APE) of a variable. In Appendix A we show that

$$
\int \frac{\partial \Pr(I_t = 1|\hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)}{\partial w} dF_{\hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t} \text{ and } \int \frac{\partial}{\partial w}\left(\frac{\Pr(I_t = 1, F_t = 1|\hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)}{\Pr(F_t = 1|\hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)}\right) dF_{\hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t}
$$

are the true measure of the effect of w on the probability of being an innovator and the probability of being an innovator conditional on being financially constrained. We discuss tests for the estimates of APE in Appendix E.

B.2. *The Third Stage: Estimation of the R&D Switching Regression Model*

The structural parameters of interest, Θ_3, of the R&D switching regression equation in (3.14) are estimated in the third stage, which is an extension of Heckman's classical two step estimation to multivariate selection problem. Here we are dealing with two kinds of selection problems: (1) R&D investment conditional on being financially constrained or not, and (2) R&D investment conditional on being an innovator, where being an innovator determines if R&D expenditure needs to be declared or not. To consistently estimate the parameters of equation (3.14), in Appendix D we derive the correction terms that correct for the bias due to endogenous switching and the bias due to endogenous sample selection. These correction terms are obtained for each firm-year observation. Adding these extra correction terms, in addition to the estimates of $\hat{\tilde{\alpha}}$ and $\hat{\boldsymbol{\epsilon}}_t$, for each observation, we obtain consistent estimates of Θ_3.

To this effect, consider the following conditional mean:

$$\begin{aligned}\mathrm{E}(R_t|F_t^*, I_t^* > 0, \hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = F_t\Big(&\beta_f + \mathcal{X}_t^{R\prime}\boldsymbol{\beta}_1 + \mu_1\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\eta_1\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \mathrm{E}(\tilde{\eta}_{1t}|F_t^* > 0, I_t^* > 0, \mathbf{X}, \mathcal{Z})\Big)\\ &+(1 - F_t)\Big(\mathcal{X}_t^{R\prime}\boldsymbol{\beta}_0 + \mu_0\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\eta_0\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \mathrm{E}(\tilde{\eta}_{0t}|F_t^* \leq 0, I_t^* > 0, \mathbf{X}, \mathcal{Z})\Big).\end{aligned} \tag{3.19}$$

Now, we know that

$$\mathrm{E}(\tilde{\eta}_{1t}|F_t^* > 0, I_t^* > 0, \hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \mathrm{E}[\tilde{\eta}_{1t}|\tilde{\zeta}_t > -\varphi_t, \tilde{\upsilon}_t > -\gamma_t],$$

and

$$\mathrm{E}(\tilde{\eta}_{0t}|F_t^* \leq 0, I_t^* > 0, \hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \mathrm{E}[\tilde{\eta}_{0t}|\tilde{\zeta}_t \leq -\varphi_t, \tilde{\upsilon}_t > -\gamma_t],$$

where φ_t and γ_t have been defined in (3.17). In Appendix C we show that

$$\mathrm{E}[\tilde{\eta}_{1t}|\tilde{\zeta}_t > -\varphi_t, \tilde{\upsilon}_t > -\gamma_t] = \sigma_{\tilde{\eta}_1}\rho_{\tilde{\eta}_1\tilde{\zeta}}C_{11t} + \sigma_{\tilde{\eta}_1}\rho_{\tilde{\eta}_1\tilde{\upsilon}}C_{12t} \tag{3.20}$$

and

$$\mathrm{E}[\tilde{\eta}_{0t}|\tilde{\zeta}_t \leq -\varphi_t, \tilde{\upsilon}_t > -\gamma_t] = \sigma_{\tilde{\eta}_0}\rho_{\tilde{\eta}_0\tilde{\zeta}}C_{01t} + \sigma_{\tilde{\eta}_0}\rho_{\tilde{\eta}_0\tilde{\upsilon}}C_{02t}, \tag{3.21}$$

where

$$C_{11t} \equiv \phi(\varphi_t)\frac{\Phi\left((\gamma_t-\rho_{\tilde{\zeta}\tilde{v}}\varphi_t)/\sqrt{1-\rho^2_{\tilde{\zeta}\tilde{v}}}\right)}{\Phi_2(\varphi_t,\gamma_t,\rho_{\tilde{\zeta}\tilde{v}})}, \quad C_{12t} \equiv \phi(\gamma_t)\frac{\Phi\left((\varphi_t-\rho_{\tilde{\zeta}\tilde{v}}\gamma_t)/\sqrt{1-\rho^2_{\tilde{\zeta}\tilde{v}}}\right)}{\Phi_2(\varphi_t,\gamma_t,\rho_{\tilde{\zeta}\tilde{v}})},$$

$$C_{01t} \equiv -\phi(\varphi_t)\frac{\Phi\left((\gamma_t-\rho_{\tilde{\zeta}\tilde{v}}\varphi_t)/\sqrt{1-\rho^2_{\tilde{\zeta}\tilde{v}}}\right)}{\Phi_2(-\varphi_t,\gamma_t,-\rho_{\tilde{\zeta}\tilde{v}})}, \text{ and } C_{02t} \equiv \phi(\gamma_t)\frac{\Phi\left((-\varphi_t+\rho_{\tilde{\zeta}\tilde{v}}\gamma_t)/\sqrt{1-\rho^2_{\tilde{\zeta}\tilde{v}}}\right)}{\Phi_2(-\varphi_t,\gamma_t,-\rho_{\tilde{\zeta}\tilde{v}})}.$$

In the above ϕ is the standard normal density function, Φ the cumulative distribution function of a standard normal, and Φ_2 is the cumulative distribution function of a standard bivariate normal.

Given estimates of $\hat{\tilde{\alpha}}$, $\hat{\boldsymbol{\epsilon}}_t$, φ_t, γ_t, and $\rho_{\tilde{\zeta}\tilde{v}}$, we can construct the above control functions, which control for the bias that arises due to endogeneity of financial constraint faced and the bias due to endogenous selection. With the above defined, we can now write the R&D switching equations in (3.14), conditional on F_t, $I_t = 1$, $\mathbf{X}$, $\mathcal{Z}$ as

$$\begin{aligned} R_t = F_t&\left(\beta_f + \mathcal{X}_t^{R\prime}\boldsymbol{\beta}_1 + \mu_1\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\eta_1\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \sigma_{\tilde{\eta}_1}\rho_{\tilde{\eta}_1\tilde{\zeta}}C_{11t} + \sigma_{\tilde{\eta}_1}\rho_{\tilde{\eta}_1\tilde{v}}C_{12t} + \underline{\eta}_{1t}\right) \\ &+ (1-F_t)\left(\mathcal{X}_{it}^{r\prime}\boldsymbol{\beta}_0 + \mu_0\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\eta_0\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \sigma_{\tilde{\eta}_0}\rho_{\tilde{\eta}_0\tilde{\zeta}}C_{01t} + \sigma_{\tilde{\eta}_0}\rho_{\tilde{\eta}_0\tilde{v}}C_{02t} + \underline{\eta}_{0t}\right), \end{aligned} \tag{3.22}$$

where $\underline{\eta}_{1t}$ and $\underline{\eta}_{0t}$ conditional on F_t^*, I_t^*, $\mathbf{X}$, $\mathcal{Z}$ is distributed with mean zero. With the additional correction terms – C_{11}, C_{12}, C_{01}, and C_{02} – constructed for every firm year observation, the parameters of the R&D switching regression model can be consistently estimated by running a simple pooled OLS for the sample of selected/innovating firms. Analytical expression for the error adjusted covariance matrix for the estimates of the the third state structural parameters, Θ_3, has been derived in Appendix D.

To measure the magnitude by which R&D intensity is affected due to the presence of financial constraint we have to compute the average partial effect (APE) of F_t. For a firm, i, in time period, t, given $\mathcal{X}_t = \bar{\mathcal{X}}$, where $\mathcal{X}_t$ is the union of elements appearing in $\mathcal{X}_t^R$, $\mathcal{X}_t^F$, and $\mathcal{X}_t^I$, the APE of financial constraint on R&D intensity is computed as the difference in the expected R&D expenditure between the two regimes, non-financially constrained and financially constrained, averaged over $\hat{\hat{\alpha}}$ and $\hat{\boldsymbol{\epsilon}}$. The APE of financial constraint on R&D expenditure, conditional on $I_t = 1$, is given by

$$\Delta_F \mathrm{E}(R_t|\bar{\mathcal{X}}) = \int \mathrm{E}(R_{1t}|\bar{\mathcal{X}}, F_t = 1, I_t = 1, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}) dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}} - \int \mathrm{E}(R_{0t}|\bar{\mathcal{X}}, F_t = 0, I_t = 1, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}) dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}}. \tag{3.23}$$

In Appendix E we discuss the estimation and the testing of the above measure.

4. DATA AND DEFINITION OF VARIABLES

For our empirical analysis we had to merge two data sets, one containing information on R&D related variables and the other on the financial status of the firms. The data on information related to R&D is obtained from the Dutch Community Innovation Surveys (CIS), which are conducted every two years by the Central Bureau of Statistics (CBS) of The Netherlands. The Innovation Survey data are collected at the enterprise level. Information on financial variables is available at the firm/company level, which could be constituted of many enterprises consolidated within the firm. The financial data, known as Statistiek Financiën (SF), is from the balance sheet of the individual firms.

A combination of a census and a stratified random sampling is used to collect the CIS data. A census of large (250 or more employees) enterprises, and a stratified random sample for small and medium sized enterprises from the frame population is used to construct the data set for every survey. The stratum variables are the economic activity and the size of an enterprise, where the economic activity is given by the Dutch standard industrial classification. For our empirical analysis we use three waves of innovation survey data: CIS2.5, CIS3, and CIS3.5 pertaining respectively to the years 1996-98, 1998-2000, and 2000-02.

However, since not all enterprises belonging to the firm have been surveyed in the CIS data the problem when merging the SF data and the CIS data is to infer the size of the relevant R&D variables for each firm. To do this we use the information on the sampling design used by CBS.

For any given year, let N be the total population of R&D performing enterprises in the Netherlands. From this population a stratified random sampling is done. These strata are again based on size and the activity class. Let S be the total number of strata, and each stratum is indexed by $s = 1, 2, \cdots, S$. Then, $\sum_{s=1}^{S} N_s = N$, where N_s is the population size of R&D performing enterprise belonging to stratum s. Let n_s be the sample size of each stratum and let $\Theta_s = \{1, 2, \cdots, i, \cdots, i_s\}$ be the set of enterprises for the s^{th} stratum, that is $|\Theta_s| = n_s$.

Let x be the variable of interest and x_i the value of x for the i^{th} enterprise. The average value of x for an enterprise belonging to the s^{th} stratum is $\bar{x}_s = (\sum_{i\in\Theta_s} x_i)/n_s$. Now consider a firm f. Let N_{fs} be the total number of enterprises belonging to the firm f and stratum s and n_{fs} be the number of enterprises belonging to firm f and stratum s that have been surveyed.

Then the estimated value of x for the firm f, $\hat{x}_f$ is given by

$$\hat{x}_f = \sum_{s=1}^{S}(N_{fs} - n_{fs})\bar{x}_s + \sum_{s=1}^{S}\sum_{k=1}^{n_{fs}} x_{fsk}, \tag{4.1}$$

where x_{fsk} is the value of x for the k^{th} enterprise belonging to stratum s and firm f that has been surveyed, and $N_{fs} - n_{fs}$ is the number of enterprises of the f^{th} firm in stratum s that have not been surveyed. It can be shown under appropriate conditions that $\hat{x}_f$ is an unbiased estimator of the expected value of x for firm f[5]. Table I below gives, based on size class and 2 digit Dutch Standard Industry Classification (SBI), the number of strata between which the enterprises surveyed in the CIS surveys were divided.

For our analysis $N_f = \sum_{s=1}^{S} N_{fs}$ was obtained from the Frame Population constructed by the CBS and $n_f = \sum_{s=1}^{S} n_{fs}$ was obtained from the CIS surveys. The exact count of firms for which $N_f = n_f$ and for which $(N_f - n_f) > 0$ can be found in Table III. The sample of firms used in the estimation is, however, much smaller than shown in Table III. The sample

[5]Proof:

The proof is based on the assumption that the expected value of x is the same for each enterprise in a particular stratum. Let μ_{xf} be the population mean of x for the firm f and let μ_{xs} be the population mean of x for an enterprise belonging to stratum s. Given our assumption, we know that $\bar{x}_s$ is an unbiased estimator of μ_{xs}, that $\mu_{xf} = \sum_{s=1}^{S} N_{fs}\mu_{xs}$, and that the expected value of $\sum_{s=1}^{S}\sum_{k=1}^{n_{fs}} x_{fsk}$, the second term on the RHS of equation (4.1), is $\sum_{s=1}^{S} n_{fs}\mu_{xs}$. Taking expectations in (4.1) and substituting the expected value of $\mathrm{E}(\sum_{s=1}^{S}\sum_{k=1}^{n_{fs}} x_{fsk}) = \sum_{s=1}^{S} n_{fs}\mu_{xs}$ and noting that $\mathrm{E}(\sum_{s=1}^{S} n_{fs}\bar{x}_s) = \sum_{s=1}^{S} n_{fs}\mu_{xs}$, we get $\mathrm{E}(\hat{x}_f) = \mu_{xf} = \sum_{s=1}^{S} N_{fs}\mu_{xs}$.

TABLE I

Number of Enterprises and Number of Strata

	CIS2.5	CSI3	CIS3.5
Total no. of enterprises	13465	10750	10533
Total no. of strata	240	249	280

These figures are from the original/raw data set.

of firms used for the analysis are only those for which we had financial information, that is, those in the SF data. For these firm we required that at least one of their potentially R&D performing enterprises be present in the innovation surveys. Enterprises in the innovation survey belonging to firms not present in the SF data had to be dropped. The percentage of firms *in the sample* for which imputation, using equation (4.1), had to be done was 18.06% in CIS2.5, 24.62% in CIS3 and 23.75% in CIS3.5. The majority of the firms happened to be single enterprises: 78.97%, 74.01%, and 73.87% respectively for CIS2.5, CIS3, and CIS3.5.

The two variables of interest for which the aggregating exercise in equation (4.1) was done are the R&D expenditure and the share of innovative sales in the total sales ($SINS$) of the enterprise. Here we would like to mention that we do not have any information on these two variables for those firms that have been categorized as non-innovators. An enterprise is considered to be an innovator if either one of the following conditions is satisfied: (a) it has introduced a new product to the market, (b) it has introduced a new process to the market, (c) it has some unfinished R&D project, and (d) it has begun an R&D project, and abandoned it during the time period that the survey covers. Given that the criteria, classifying an enterprise as an innovator, are exhaustive, we, for the purpose of aggregation, reasonably assumed that if an enterprise meets none of the above criteria, it has no R&D expenditure and no new products.

We consider a firm to be financially constrained as soon as any one of its enterprises declares to be financially constrained. When $N_f > n_f$, a firm is characterized as an innovator if one the constituent enterprises surveyed has innovated or if anyone of the enterprises that have not been surveyed is found in a stratum that is classified as an innovating stratum[6],

[6] An example could help illustrate. Suppose there is a firm that has three enterprises: E_1, E_2, and E_3. Assume that of the three enterprises only E_3 has been surveyed, and has been found not to innovate. Now, we know to which stratum E_1 and E_2 respectively belong to. Let E_2 belong to the stratum s and E_1 to

where a stratum is defined to be innovative if $\bar{x}_s > 0$.

The total number of employees as a measure of the size of the firm was also constructed using information from the CIS data and the General Business Register. As far as the number of employees in a firm is concerned, if all the enterprises belonging to a firm are surveyed, that is if $N_f = n_f$, then we simply add up the number employees of each of the constituent enterprises. However, when $N_f > n_f$, for those enterprises that have not been surveyed we take the mid point of the size class of those enterprises that have not been surveyed. The size class to which an enterprise belongs to is available from the General Business Register for every year.

In Table II below we tabulate the number of innovating and non-innovating firms for each of the three waves, and the number of firms that declare to be financially constrained in their innovation activities. As can be seen from the table, for CIS2.5 information on financial constraint is available only for the innovators. It can be noticed that the number of financially constrained firms is much lower than the number of unconstrained firms. In our sample we find that the number of financially constrained firms is larger for the innovating firms than for the non-innovating ones.

As mentioned earlier the CIS survey is conducted every two years. The question on being innovative or being financially constrained pertains to all the years of the survey. However, the variables, share of innovative sales in the total sales ($SINS$) and R&D expenditure are reported only for the last year. The stock variables – long-term debt, liquidity reserve, assets of the firms, and the number of employees, indexed t – are the values of the variables as recorded at the beginning of period t. The flow variables are the observed values as recorded during period t. R&D expenditure and $SINS$ are reported only for the last year of the periods that any CIS covers.

Below we provide the definition and the list of the variables that were used in the empirical exercise.

1. R_t: R&D intensity defined as the ratio of R&D expenditure to total (tangible+ intangible) capital assets
2. F_t: Binary variable equal to one if the firm is financially constrained

stratum s'. If we find that $\bar{x}_s > 0$ and that $\bar{x}_{s'} = 0$, we will still regard the firm to be an innovator, with R&D expenditure $\bar{x}_s$.

TABLE II

Innovating/Non-Innovating and Financially Constrained/Unconstrained Firms

	Financially Constrained	Financially Unconstrained	Total
CIS2.5 (1996-98)			
Innovators	525	2,422	2,947
Non-Innovators			2,416
Total			5,363
CIS3 (1998-00)			
	Financially Constrained	Financially Unconstrained	Total
Innovators	336	1,508	1,844
Non-Innovators	75	1,504	1,579
Total	411	3,012	3,423
CIS3.5 (2000-02)			
	Financially Constrained	Financially Unconstrained	Total
Innovators	154	1,826	1,980
Non-Innovators	32	2,234	2,266
Total	186	4,060	4,246

These figures are for the data set used in estimation.

In CIS 2.5, non-innovating firms do not report if they are financially constrained.

3. I_t: Binary variable equal to one if the firm is an innovator
4. $DEBT_t$: Long-term debt constituted of the book value of long-term liabilities owed to group companies, members of cooperative society and other participating interests, plus subordinated loans and debentures
5. LQ_t: Liquidity reserve including cash, bills of exchange, cheques, deposit accounts, current accounts, and other short-term receivables
6. DIV_t: Dividend payments to shareholders, group companies, and cooperative societies
7. $SIZE_t$: Logarithm of the number of people employed
8. $RAINT_t$: Ratio of intangible assets to total (tangible+ intangible) capital assets
9. $SINS_t$: Share of sales in the total sales of the firm which is due to newly introduced products
10. CF_t: Cash flow defined as operating profit after tax, interest payment, and preference dividend plus the provision for depreciation of assets
11. $MKSH_t$: Market share defined as the ratio of firms sales to the total industry sales
12. $DNFC_t$: Dummy variable that takes value one for negative realization of cash flow
13. $DMULTI_t$: Dummy that takes value one if a firm has multiple enterprises
14. AGE_t: Age of the firm[7].
15. Industry dummies
16. Year dummies

To minimize heteroscedasticity we scale long-term debt ($DEBT_t$), cash flow (CF_t), liquidity reserve (LQ_t), and dividend payout DIV_t by total capital assets. Henceforth whenever we refer to these variables, it would mean the scaled value of these variables.

[Table IV about here]

A. *Endogenous Explanatory Variables*

The set of endogenous regressors, $\mathbf{x}_t$, that appear in the structural equations, and for which we construct control functions to account for their endogeneity are:

1. Long-term debt ($DEBT_t$)

[7]We do not the age of the firms that existed prior to 1967 as the General Business Register, from which we calculated the age of the firms, was initiated in 1967. For such cases we assume that the firm began in 1967.

2. Liquidity reserve (LQ_t)
3. Dividend payout (DIV_t)
4. Logarithm of the number of people employed ($SIZE_t$)
5. Ratio of intangible assets to total assets ($RAINT_t$)
6. Share of innovative sales in the total sales of the firm ($SINS_t$)

While we know that financial variables, debt, liquidity reserve, and dividends, are endogenously determined, we also test whether size of the firm is endogenously determined along with financing and innovation decisions. Both AH and CH point out that under endogenous borrowing constraint, debt and equity value of the firm are together endogenously determined with size of the firm. We take ratio of intangible assets to total assets. $RAINT_t$ as endogenous because it could be determined by the decision to innovate and investment in R&D.

Share of innovative sales in the total sales of the firm, $SINS_t$, will most likely be endogenous because it could be determined by current investment decision. $SINS_t$ is only observed for innovators. For the purpose of estimating the reduced form equation we assume that $SINS_t$ is zero for the non-innovators. Given that the classification criteria, classifying firms as innovators, is fairly exhaustive, we believe that this is not a strong assumption.

B. *Exogenous Explanatory Variables*

The vector of exogenous variables, $\mathbf{z}_t$, that appear in the structural and reduced form equation are:

1. Cash flow of the firm (CF_t)
2. Dummy for negative realization of cash flow ($DNFC_t$)
3. Market share of the firm ($MKSH_t$)
4. Age of the firm (AGE_t)
5. Dummy that takes value 1 if the firm consists of multiple enterprises ($DMULTI_t$)
6. Industry dummies
7. Year dummies

Cash flow is assumed to be exogenous because cash flow, which as Moyen (2004) points out is highly correlated with the income shock, is largely driven by exogenous shocks.

However, it should be pointed out it is exogenous conditional on unobserved heterogeneity, $\tilde{\alpha}_i$. Hence, any component of cash flow that is endogenous to the system of equations has been accounted for by allowing it to be correlated with the unobserved heterogeneity. Similarly, while market share, $MKSH_t$, and dummy for multiple enterprise, $DMULTI_t$, may not be strictly exogenous, they are likely to be given unobserved heterogeneity[8].

C. *Additional Instruments*

Our additional set of instruments, $\tilde{\mathbf{z}}_t$, needed to identify the structural parameters through the control functions constructed from the first stage reduced estimates are:

1. Cash flow in period $t-1$ (CF_{t-1})
2. Dummy for negative cash flow ($DNFC_{t-1}$)
3. Square of cash flow in period $t-1$ (CF^2_{t-1})
4. Square of cash flow in period t (CF^2_t)
5. Market share in period $t-1$ ($MKSH_{t-1}$)
6. Dummy that takes value 1 if the firm consists of multiple enterprises in period $t-1$ ($DMULTI_{t-1}$)
7. Dummy if the firm existed prior to 1967 ($DAGE_t$)

We include past realization of cash flow in the set of instruments because, as argued earlier, cash flow is strongly correlated with exogenous revenue shocks experienced by the firm. To the extent that financing decisions of the firms are state contingent, current and past realizations will influence all financing decision. For example, AH have shown that better realization of past revenue shocks imply a higher leverage and long-term debt. Reddick and Whited (2009) show that saving and cash flow are negatively correlated because firms optimally lower liquidity reserves to invest after receiving a positive cash flow shocks. Hence, liquidity holdings of the firm and past level of income shocks are expected to be correlated. Similarly, a higher dividend payout could be expected with better realization of past revenue shocks.

[8]Most paper studying nonlinear panel data models assume all regressors to be exogenous conditional on unobserved heterogeneity. In this paper we have relaxed this assumption to allow certain variables, $\mathbf{x}_t$, to be correlated with the idiosyncratic component even after having accounted for their correlation with unobserved heterogeneity.

It has found that firms with monopoly and those that are multiple enterprise firm are more likely to engage in innovative activity. Hence, firms that have had a higher degree of monopoly in the past or have been a multiple enterprise firm in the past could be expected to have a higher share of innovative sales, $SINS_t$, today, and a higher ratio of intangible assets to total capital assets, $RAINT_t$. Finally, given that age and size of a firm are correlated, $DAGE_t$ of the firm has been assumed to instrument size. We interact cash flow and market share in period $t-1$ with $DMULTI_{t-1}$ and $DAGE_t$.

We stress again that variables included in $\mathcal{Z}_t = \{\mathbf{z}'_t, \tilde{\mathbf{z}}'_t\}'$ may or may not be strictly exogenous, but, conditional on unobserved individual effects, these variables are unlikely to be correlated with idiosyncratic component in the structural equations. To the extent that we take into account the correlation betweeen $\mathcal{Z}_t$ and $\tilde{\alpha}_i$, the presence of these variables in the specification of the structural equations or as instruments will not lead to inconsistent results.

5. RESULTS

A. *Financing and Innovation Decision*

As discussed earlier, in the second stage we jointly estimate the structural parameters of the financial constraint and the innovation equation. The results of the second stage estimation results are shown in Table V and Table VI. While Table V has the coefficient estimates, in Table VI the Average Partial Effects (APE) of the covariates are reported. In Specification 2 and Specification 3 shown in Table V and VI we do not have dummies for multiple enterprises in the financial constraint equation, and while the specification for the innovation equation in Specification 1 and 2 are same, in Specification 3 we remove the control function/correction term for share of innovative sales.

We begin by discussing the results of the Innovation equation[9]. We find that firms with

[9]In the innovation equation, unlike Hajivassiliou and Savignac (2011), we do not include the financial constraint variable F_t. This is because our aim in this paper is to study innovation and financing decision of firms unlike Hajivassiliou and Savignac (2011), who look at how financial constraints affect the innovation of potentially innovating firms. Given their objective, they exclude firms that have no wish to innovate. Excluding such firms helps them identify the impact of F_t on I_t, which takes value 1 for firms that innovate and 0 for those who want to innovate but cannot. In our data set, as discussed earlier, for CIS2.5 we can

higher long-term debt, $DEBT$, in their capital structure are less likely to take up innovative activity. This is consistent with the theoretical prediction, as discussed earlier, that bond holders are unwilling to hold the higher risks associated with R&D activity, and also with the findings of empirical papers, such as BFP and others, who find that equity rather than debt may be more suitable to finance innovative activity.

[Table V about here]

[Table VI about here]

We also find that firms that take up innovative activity maintain higher amount of liquidity reserve, LQ. Again, as pointed out earlier, because R&D intensive firms behave as if they faced large adjustment cost, they choose to smooth their R&D spending. This necessitates that innovative firms maintain a higher level of cash reserve to counter periods of negative revenue shocks.

As far as dividend pay out is concerned, in Specification 3, where we remove the correction term for $SINS$ in the innovation equation, we find a significant negative coefficient for dividends, DIV. We remove the correction term for $SINS$ in the selection because $SINS$, which is observed only for the innovators, is not included in the specification for the innovation equation[10]. This suggests that firms that pay out dividends are less likely to innovate. Now, given the nature of R&D activity that makes borrowing costly, internal funds may be more preferable. Therefore, innovative firms, ceteris paribus, are less likely to distribute cash as dividends.

We find that large firms are more likely to be ones taking up innovative activity. While the finding is consistent with the Schumpeterian view that large firms have a higher incentive to

not distinguish between those firms that want to innovate but due to constraints cannot innovate and those who have no wish to innovate. That is, in CIS2.5 only innovators report if they are financially constrained. Hence in our data set we cannot identify if innovation is hampered due to the presence of financing constraints. Moreover, our aim is to study how financing and innovation choices are related and how $\Pr(I = 1|F = 1)$ and $\Pr(I = 1|F = 0)$ changes with the financing policy of firms with different characteristics.

[10] As stated earlier, since $SINS$ is not observed for non-innovators, we assumed $SINS$ to be zero for the non-innovators when estimating the system of reduced form equation. Therefore, like $SINS$, the correction term for $SINS$ will be highly correlated with I, the decision to innovate. This could be the reason for the very high significance of correction term/control function for $SINS$ in Specification 1 and Specification 2.

engage in innovative activities because they can amortize the large fixed costs of investing by selling more units of output, we also know that large firms, as shown in AH and CH, are less likely to face constraint in accessing external capital and therefore more likely to engage in R&D activity.

We find that younger firms are more innovative. This corroborates with the findings of other studies that find that young firms in their bid to survive and grow take up more innovative activity. Entry (see Audretsch, 1995; Huergo and Jaumandreu, 2004) is envisaged as the way in which firms explore the value of new ideas in an uncertain context. Entry, the likelihood of survival and subsequent growth are determined by barriers to survival, which differ by industries according to technological opportunities. In this framework entry is innovative and increases with uncertainty. Also, firms with large market share, $MKSH$, are found to be engaging more in innovative activity. This result confirms the fact that to prevent entry of potential rivals a firm is more incited to innovate if it enjoys a monopoly position, as has been argued in the Schumpeterian tradition.

The ratio of intangible assets to total capital assets, $RAINT$, has been found to be significantly positive in the innovation equation. Since firms that engage in innovative activity have more intangible assets in their asset base, this should be expected. Besides, as Raymond et al. (2010) point out, there is persistence in innovation activity of a firm, or in other words, innovation decision exhibits a certain degree of path dependency. To the extent that $RAINT$ is the outcome of past innovation activity, it captures the persistence in the innovation decision of the firm. We also find that firms that have many enterprises consolidated within them, $DMULTI$, are more likely to be innovative. Cassiman et al. (2005) argue that entreprises merged or acquired may realize economies of scale in R&D, and therefore have bigger incentive to perform R&D than before. Also, when merged entities are technologically complementary they realize synergies and economies of scope in the R&D process through their merger, and become more active R&D performers after being merged or acquired.

We also find that factor loading, θ, which is the coefficients of $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ in the Innovation equation is significant. This and the fact that the control functions to correct for the bias in the structural equations due to the presence of endogenous regressors are all significant suggest a strong simultaneity in the decision to innovate and the financing choices made.

B. *Financial constraint and Innovation*

In this subsection we discuss the specification and the results of the Financial Constraint equation. Here we will also discuss how financing and innovation decisions under financial constraints vary over the distribution of firm characteristics.

To begin with, as discussed earlier, given the financial state of a firm, higher expected profitability from R&D investment could lead to a firm being financially constrained. Therefore we need to control for the investment opportunity of the firm. To this end, we include cash flow of the firm in specification for Financial Constraint equation. However, the realized cash flow of the firm may not be only from the firm's R&D activity. A measure to control for the investment opportunity for R&D related activity should be based on a measure such as Tobin's "q" for R&D related activity or cash flows that result from R&D output. However, in the absence of any such measure, we use the share of innovative sales in the total sales of the firm, $SINS$, which can potentially signal demand for R&D related activity. Besides, Moyen (2004) finds that Tobin's "q" is a poor proxy for investment opportunities, cash flow is an excellent proxy, and that cash flow is an increasing function of the income shock. We find that both CF and $SINS$ have a significant positive sign in the Financial Constraint equation[11]. This suggests that both cash flow and the share of innovative sales are correlated with the R&D investment opportunity set and, ceteris paribus, are indicative of the financing gap that firms face. We note here that while CF, which is largely driven by exogenous shocks and is exogenous conditional on $\tilde{\alpha}$, $SINS$ is an outcome of current and past R&D efforts. Therefore we endogenise $SINS$. The coefficient of the control function for $SINS$ suggest that financial constraints and $SINS$ are determined endogenously.

In our specification we also have a dummy for negative cash flow, $DNCF$, which is found to have a significantly positive coefficient. It seems that variations over time from negative

[11] While it may be desirable to include a measure of expected profitability from R&D investment in the Innovation equation, we do not include cash flow, CF, and share of innovative sales in the total sales, $SINS$, in the Innovation equation. We do not include $SINS$ because it is observed only for innovators. We do not include cash flow in the Innovation equation because, as explained in section 4, in our data the decision to innovate precedes the realization of cash flow. Hence, cash flow can not identify a firm's decision to innovate.

to positive cash flow are more indicative of positive "shifts" in the supply of internal equity finance that relax financial constraint than variation in cash flow itself.

For all the specifications we obtain a significant positive sign on debt to assets ratio, $DEBT$, indicating that highly leveraged firms are more likely to be financially constrained. This is consistent with the prediction in AH and CH, who show that firms with higher long-term debt in their capital structure are more likely to face tighter short-term borrowing constraint. This could also reflect the debt overhang problem studied in Myers (1977). It is also possible that, ceteris paribus, firms with higher leverage face a threat of default and therefore a higher premium on additional borrowing due to bankruptcy costs. Also, as can be evinced from the APE's in Table VI, for an average firm, the likelihood of experiencing higher financial constraint is quite high for a firm that has higher long-term debt in its capital structure.

We find that firms that maintain higher liquidity reserve, LQ, are less likely to be constrained. Gamba and Triantis (2008) point out that cash balances, which give financial flexibility to firms, are held when external finance is costly and/or income uncertainty is high. With higher liquidity reserve firms can counter bad shocks by draining it. Hence, when a firm is not sure about a steady supply of positive cash flow it is likely to practice precautionary savings to reduce its risks of being financially constrained during periods of bad shocks. Besides, R&D intensive firms behave as if they face large adjustment costs, and therefore chose to smooth their R&D spending. Hence, the need of financing flexibility could be important for innovation firms.

Our results suggest that dividends DIV paying firms are less likely to be financially constrained. HW also find low dividend paying firms face high costs of external funds. Besides, AH and CH show that when a firm faces borrowing constraint, and all profits are reinvested or paid to the lenders so that the burden of debt is reduced and the firm grows to its optimal size, no dividends are paid. Since the APE of dividends, as shown in Table V, is very high, our results lend credence to papers that employ dividend pay out as a criterion for classifying firms as financially constrained or unconstrained.

We find that large and mature firms are less likely to be financially constrained. HW also find large differences between the cost of external funds for small and large firms. AH and CH show that over time as the firm pays off its debt, it reduces its debt burden and increases its equity value. This increase in the value of equity reduces the problem of threat of default in AH and the problem of moral hazard in CH, with the result that the extent of borrowing constraint decreases, the advancement of working capital from the lender increases and the firm grows in size. Consequently larger and mature firms are less likely to face financial constraint. On the other hand, old firms having survived through time have built a reputation over the years and are therefore less likely to face adverse information asymmetry problems as compared to young firms.

We include the ratio of intangible assets to total capital assets, $RAINT$, in the specification for financial constraint. Since secondary markets for intangible asset is fraught with more frictions and generally does not exist, firms with a higher percentage of intangible assets have a lower amount of pledgable support to borrow, and are thus expected to be more financially constrained. Almeida and Campello (2007) also find that firms with lower levels of asset tangibility are more financially constrained, and that investments in intangible assets do not generate additional debt capacity. Our results suggest that firms that have a higher percentage of intangible assets are indeed more likely to be financially constrained. Since a large part of the capital of an R&D intensive firm resides in the knowledge base of the firm, which is intangible, innovating and R&D intensive firms, as can be evinced in Table IV, have a higher intangible asset base. Given this fact, innovating firms are thus more likely to face financial constraint.

We do not, however, find firms with a high market share, which serves as a proxy for monopoly power, and firms with multiple enterprises to be significantly less or more financially constrained.

In Table V we find λ, which are the coefficients of $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ in the Financial Constraint equation and all correction terms to be significant, suggesting that the share of innovative sales, long-term debt, liquidity reserve, dividends, size, and the ratio of intangible assets to total assets are endogenously determined.

In Figure 1 we plot the average partial effect of long-term debt on the propensity to innovate conditional on being financially *constrained* $\left(\int \frac{\partial \Pr(I=1|F=1,\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}})}{\partial DEBT} dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}} \right)$ and conditional being financially *unconstrained* $\left(\int \frac{\partial \Pr(I=1|F=0,\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}})}{\partial DEBT} dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}} \right)$. We plot the APE of $DEBT$ against size, age and leverage. These plots of APE against age, size and leverage are based on Specification 2 of the second stage estimation. The APE plots based on other specifications are almost exactly same.

[Figure 1 about here]

We find that conditional on *not* being financially constrained, the APE of $DEBT$ on innovation to be negative and almost constant over the distribution of size, age and leverage. In contrast, the APE of $DEBT$ on innovation conditional on being financially constrained varies widely over the distribution of age, size and leverage, and is less negative and sometime positive when compared to the APE of $DEBT$ on innovation conditional on not being financially constrained. This indicates that under no financial constraint innovative firms, regardless of size, maturity, and existing level of debt, would almost *uniformly* be less inclined to innovate by financing themselves with debt. In other words, when borrowing constraint do not bind and debt is accessible on easier terms, and if for some reason the firm has to finance itself with debt, then it is very unlikely the debt financing will be used for engaging in or starting an innovative activity. The following scenario can elucidate this: suppose there is a profitable firm, that has a substantial amount of cash holdings, that it can distribute to its shareholders. Being profitable, it is likely that it has a rather large debt capacity and suppose its existing debt levels are such that it has not reached its debt capacity. In such a situation, the firm can distribute cash and borrow more to finance its investment. However, if it decides to innovate or spend more on R&D related activity, then as our results suggests, it would be less inclined to distribute cash as dividends, be more inclined to maintain a high cash reserves and not borrow more, in other words, finance itself with cash flow or retained earnings. This is in congruity with the findings of BFP, who show that in the absence of constraint, when internal and external equity are easily available, the preferred means for financing innovation is not debt.

When financial constraints set in, innovating firms, though still averse to debt financing, do innovate by borrowing as is reflected in the relatively higher change in propensity or willingness to innovate by increasing $DEBT$ as compared to when firms are unconstrained. Now, under financial constraint, as Lambrecht and Myers (2008) explain, there can be two possibilities: (a) postpone investment or (b) borrow more to invest. Given that most of the firms that report being financially constrained are innovators, it is true that these firms have not entirely abandoned innovative activity. Therefore, the fact that the change in propensity to innovate by increasing $DEBT$ is relatively higher than under no financial constraint, suggests that some projects might have been valuable enough to be pursued by borrowing, even if that implied a higher cost.

However, under financial constraint, the change propensity to innovate by increasing $DEBT$ varies with size, age, and existing leverage. This is because under financial constraint, the relative cost of, or access to, external financing depends on firm's age, size, and the existing levels of debt.

Consider the plot of APE of $DEBT$ on innovation conditional on financial constraint against size of the firm. We see that under financial constraint large firms are more likely to innovate by increasing their leverage as compared to small firms. This is because as firms become large the extent of constraint weakens, and if some R&D projects are valuable enough to be pursued, large firms have more leeway to finance their project by borrowing than small firms. Both AH and CH show that a firm with a given need of external financing to fund an initial investment and working capital, for a given level of growth opportunity and profitability, over time, during which firms face borrowing constraint and dividend payment is restricted, firms by paying off debt reduces its debt and increases its equity value. As the firm increases its equity value, with the result that the problem of threat of default in AH and the problem of moral hazard in CH decreases, the advancement of working capital from the lender increases and the firm grows in size. Thus if a large firm sees an investment opportunity in some R&D project it will be in a better position to borrow than a small firm. Also, HW find that large firms face lower bankruptcy and equity flotation costs as compared to small firms, which gives an advantage to large firms when it comes to borrowing for R&D. While the above argument explains, through the role of finance, why, for a given investment opportunity, large firms under financial constraint are more likely to be willing to engage in innovation by borrowing more, it is also true that large firms, by Schumpeterian argument, have a higher incentive to innovate, and, given that large firms have a higher stock of knowledge, they are able to find more valuable R&D investment projects.

Incentives to innovate also explain the plot of APE of $DEBT$ on the conditional probability to innovate against age of the firms. We know that even though younger firms are more likely to be financially constrained, it is the young firms that are more likely to take up innovative activity. This is because, as discussed earlier, survival and subsequent growth of young firms, especially those that are in the high-tech sector, depend on their innovation. Hence, under financial constraint young firms are more willing to finance themselves by increasing their $DEBT$ than matured firms. Consequently, we find the willingness to innovate by increasing $DEBT$ of young firms is higher compared to a matured firm. This also makes the young firms more prone to default as discussed in CQ and more likely to be financially constrained, which our results too suggests. However, the difference in APE of $DEBT$ on innovation conditional on being financially constrained for young and old is not large as compared to the same for small and large firms. This could be due to the fact that once conditioned on size, here at the mean value of all firm-year observations, APE of $DEBT$ on engaging in innovation does not vary much with age.

Lastly, under financial constraint, we find that change in propensity to innovate by increasing $DEBT$ declines with higher leverage, which only shows that, ceteris paribus, for reasons stated earlier, the borrowing constraint get tighter with higher long-term debt in the capital structure, and the firm becomes more reluctant to engage in innovative activity by increasing long-term debt.

C. *Financing Constraints and R&D Investment*

In the third stage we estimate the R&D switching regression model, given in equation (3.22), to assess the impact of financial constraint, as reported by the firms, on R&D investment. The distinguishing feature of our R&D model is that it takes into consideration the fact that R&D investment is determined endogenously along the decision to innovate and other financial choices. To the extent that the latent variable, F_t^*, underlying F_t reflects high premium on external finance and the high financing need of firms, the switching regression model for R&D investment allows us test whether financing frictions affect R&D activity adversely.

[Table VII about here]

The results of the third stage switching regression estimates are presented in Table VII. The additional correction terms – C_{11}, C_{12}, C_{01}, C_{02} – that correct for the bias that can arise due to endogeneity of selection, I_t, and financial constraint, F_t, are constructed out of the estimates of the Specification 2 of the second stage estimates. Results of the third stage that are based on the other specification of the second stage estimates are almost exactly the same, the coefficients differing at the third or fourth decimal places. The results in Table VI has two specifications; in Specification 2 the correction term for size, not being significant in Specification 1, has been dropped.

In order to see the effect of financial constraint, F_t, on R&D investment, we have to fix the firm's investment opportunity. Since we do not have any information on the market valuation of the firms, we can not construct average "q" for our firms or any such measure related to the firm's R&D investment. Hence, for reasons stated in Section 5.2, where we discussed the results of the second stage estimation, we include cash flow, CF, and share of innovative sales, $SINS$, which are indicative of demand signals and are thus correlated with the R&D investment opportunity set.

The specification for the R&D equation does not include any financial state variables such as long-term debt or cash reserves. This is because in the structural model for R&D investment, R&D investment is determined only by the degree of financial constraint a firm faces and the expected profitability from R&D investment. Therefore, it seems unlikely that leverage and cash holdings will have an independent effect, other than through the financial constraint affecting the firm. Now, we know that conditional on control functions, $\hat{\hat{\alpha}}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$, the financial state variables become exogenous to Innovation, Financial Constraint, and R&D investment. Hence, excluding the financial variables from the R&D equation helps us to identify the parameters of the R&D equation when going from the second and the third stage. This is similar to the exclusion restriction required in the Heckman two-step sample selection model.

Now, even though cash flow turns out to be significantly positive and larger for the set of financially constrained firms as compared to those that are not, a test for the existence of financial frictions in our model is not predicated on sensitivity of R&D investment to cash flow for constrained and unconstrained firms, but through the test of the effect of reported financial constraint on R&D investment. While sensitivity of R&D investment to cash flow can indicate the existence of financing frictions, as BFP claim, it could be possible that cash flow are correlated with the R&D investment opportunity set and provide information about future investment opportunities, hence, R&D investment-cash flow sensitivity may equally occur because firms respond to demand signals that cash flow contain. Besides, $SINS$, which we include in the specification to control for future expected profitability, may not perfectly control for the firm's R&D investment opportunity, giving predictive power to cash flow. Moyen (2004) too finds that cash flow is an excellent proxy for investment opportunity, and that cash flow is an increasing function of the income shock. HW discuss mechanisms, that are related to costs of issuing new equity, bankruptcy costs, and curvature of profit functions, that drive investment-cash flow sensitivity. However, it is beyond the scope of this paper to test for exact mechanism that drives the results on R&D investment-cash flow sensitivity across constrained and unconstrained firms.

We find that firms whose share of innovative sales, $SINS$, is high are more likely to be R&D intensive. This suggests that the share of innovative sales is also indicative of demand signals for R&D activity. This finding is in line with stylized facts studied in KK that more innovative firms have higher R&D intensity. However, the difference, though positive, in the size of the coefficients of $SINS$ across constrained and unconstrained firms is not high. We also find that $SINS$ is endogenous, as is reflected in the significance of correction term for $SINS$.

Here, we want to test whether financing frictions, as summarized by F_t, adversely affects a firm's R&D investment. In Specification 2, where the correction term for $SIZE$ has been dropped, we find that the coefficient of F_t is significantly negative. Now, while the $SIZE$ of the firm turns out to be endogenous to the decision to innovate, as can be evinced from the results of the second stage regression, it seems that $SIZE$, as reflected in Specification 1 of Table VI, conditional on unobserved heterogeneity $\tilde{\alpha}_i$, is exogenous to the amount invested in R&D. This could be either because the additional correction terms – C_{11}, C_{12}, C_{01}, C_{02} – that take in account the endogeneity of the decision to innovate also accounts for the endogeneity of $SIZE$. It could also reflect the fact that R&D investment, which is a fraction of total investment, affects $SIZE$ of the firm in a predetermined way. However, what does not turn out significant is the APE of financial constraint on R&D intensity, $\Delta_F \mathrm{E}(R_{it}|\bar{\mathcal{X}})$, defined in equation (3.23).

The other variables included in the specification are $SIZE$, $MKSH$, AGE, and $DMULTI$ that takes value 1 if the number of enterprises consolidated within a firm is more than one. We find that even though large firms are more likely to engage in innovative activity, among the innovators smaller firms invest relatively more in R&D than larger firms. This finding is contrary to KK who model firm dynamics with R&D and where R&D intensity is independent of firm size. This is because KK in their model do not consider the financing aspect of R&D. The finding that smaller firms are more R&D intensive could be because, as has been argued in CQ and Gomes (2001), of the fact that smaller firms have a higher Tobin's "q" than large firms, which can even be true of R&D capital. Thus smaller firms in their bid to grow exhibit risky behavior in terms of investment in R&D. Also, for larger firms investing as much as or proportionately more in R&D than smaller firms would imply subjecting themselves to higher risk. This is because large firms, as argued in CQ, operating on a larger scale are more subject to exogenous shocks, and tying up more capital, or in proportionate to size, in a risky venture as R&D can potentially make large firms more susceptible to default. This is specially true when the price process of R&D output is correlated with the output of the existing operation of the firm. Thus, given the fact that R&D capital is highly intangible, which lacks second hand market, and with decreasing returns to R&D, investing in R&D proportionate to size or more would imply making itself more prone to default. We also find that for a given $SIZE$, a constrained firm will invest less in R&D.

Young firms are found to be more R&D intensive, and as we saw earlier, are also more likely to engage in innovative activity as compared to mature firms. We also find that for a given age, constrained firms invest less compared to unconstrained firms. In our sample we find that constrained firms with a large market share, $MKSH$, invest more in R&D, but market share does not have any explanatory power for unconstrained firms. In another set of regression, where we had removed $DMULTI$ from the specification we did find a marginally significant positive sign for market share among the unconstrained firms, but the comparison of the size and the significance of the coefficients across the two regimes remained the same. Similar to the result on innovation we find that firms that have a number of enterprises consolidated within them, $DMULTI$, are more R&D intensive.

In our analysis we find that the correction term for long-term debt and dividends are significant for financially constrained firms but not for the unconstrained ones, suggesting that financing with long-term debt and dividend payout are determined endogenously with R&D investment for constrained firms but not for the unconstrained ones. This is consistent with the results of the some of the papers, cited above, that model endogenous borrowing constraint, firm investment, and firm dynamics. We find that the control function for liquidity reserve is significant for the unconstrained firms but not for the constrained ones. In another set of regression, where we had removed $DMULTI$ from the specification we found a significant sign for the control function of liquidity reserve for the constrained firms. This finding suggests that R&D investment and cash retention along with other financial decision are endogenous. This is in line with the findings of Gamba and Triantis (2008) where they analyze optimal liquidity policies and their resulting effects on firm value. In their model the decision on investment, borrowing and cash retention/distribution represent endogenous response to costs of external financing, the level of corporate and personal tax rates that determine the effective cost of holding cash, the firm's growth potential and maturity, and the reversibility of capital.

While the significance of individual control functions correcting for endogeneity of financial state variables differ across constrained and unconstrained firms we find that $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ is significant across both the regimes, suggesting overall a strong simultaneity in R&D investment and financial choices. Besides, we find that the additional correction terms – C_{11}, C_{12}, C_{01}, C_{02} – that take in account the endogeneity of the decision to innovate and the financial constraint faced are also significant.

6. CONCLUDING REMARKS

The main objective of this paper was to empirically study how incentives to innovate interact with financing frictions, frictions that assume a special status given the risky and idiosyncratic nature of R&D and innovative activity. We focused on (I) the firms' decision to innovate and the financial constraint faced given the endogenous financial choices made by the firms. Then conditional on financial choices made, the decision to innovate, and the constraint faced we tried to determine (II) how financial constraints affect R&D investment.

To the above mentioned end, we presented an empirical strategy to estimate a fully specified model of endogenous R&D investment, endogenous financial constraint, endogenous decision to innovate, and endogenous financial choices made. The strategy entailed estimating in three steps (1) a system of structural equations pertaining to – (a) a model for the decision to innovate, where we try to explain how incentives to innovate are shaped

(b) a model for financial constraint, where we try to explain why certain firms report they are financially constrained, and (c) a model for R&D investment, where we try to assess the impact of financial constraint on R&D investment – and (2) a system of reduced form equations of financing choice and other endogenous variables. The structural part (I) of the analysis was carried out conditional on the first stage reduced form estimation, and part (II) was done conditional on the first and second stage estimates.

Our methodology combined the method of "correlated random effect" and "control function" to account for unobserved heterogeneity and endogeneity of regressors in the structural equations. We believe that the estimation technique is new to the literature and solves the much discussed endogeneity problem in empirical corporate finance.

From the estimates of the second stage, where we estimated jointly the probability of being an innovator and the probability of being financially constrained, conditional on endogenous financial choices, we could garner that debt is not the preferred means of external finance for firms engaging in R&D activity, and that a highly leveraged firm is more likely to be financially constrained. We found that large and young firms, and those enjoying a higher degree of monopoly are more likely to be innovators. Also, firms that have many enterprises consolidated within them are more likely to be innovators. We found that small and young firms and firms with lower collateralizable assets are more likely to be financially constrained. Besides, the analysis also revealed that the decision to engage in R&D activity, the various financial choices, and the financial constraint faced are all endogenously determined.

Interestingly, we found that under no financial constraint, the marginal propensity to innovate with respect to leverage is lower as compared to a situation in which firms find themselves financially constrained. Also, though the marginal propensity to innovate under no financial constraint, barely varies with firm characteristics such as maturity, size and leverage, under financial constraint the propensity to innovate with respect to leverage varies with the distribution of firm characteristics. The above implies that when a firm is not financially constrained, regardless of its characteristic, it will be unwilling to engage in innovative activity by raising debt. On the other hand under constraint, even though on average debt it not a preferred means to finance innovative activity, firms do show a propensity to engage in innovative activity by raising debt. However, this propensity is influenced both by the incentives to innovate and the capacity to raise debt; both of which vary with firm characteristics. These findings draw our attention to the fact that innovation and financing policy are not independent of firms dynamics of survival, exit, and growth.

The results of the third stage R&D switching regression imply that financial constraint do adversely affect R&D investment. We found that small, young, and firms with multiple enterprises are more R&D intensive. However, for a given size and age, the financially constrained ones invest less, which again shows how financing frictions condition firm dynamics that are brought about through R&D investment. Besides, our analysis also showed that R&D investment and financing decisions are determined simultaneously. Finally, among others, one of the aims of this paper has been to gauge the magnitude of the impact of financial constraint. However, since the measure of the magnitude is not statistically significant we can not assert this finding.

These results underscore the fact that capital-market imperfections do affect the incentives to innovate, and the interaction between financing frictions and innovation is not uniform across firm characteristics. Our results therefore, taken together, point towards the fact that financing frictions that affect innovation and R&D activity also affect firm dynamics. While these findings are consistent with some of the empirical and theoretical results that seek to explain the implication of financing frictions and firm dynamics, none, to our knowledge, has explored the implications of innovation and its interaction with financing frictions in determining firm dynamics. On the other hand while models in industrial organization do study firm and industry dynamics where R&D and the stochastic nature of innovation drive the dynamics, the financial aspect and its interaction with innovative activity is found lacking. Our results suggest that future work in this area is needed.

REFERENCES

Aboody, David, and Baruch Lev. 2000. Information asymmetry, R&D, and insider gains. *Journal of Finance* 55:2747 – 2766.

Albuquerque, Rui, and Hugo A. Hopenhayn. 2004. Optimal lending contracts and firm dynamics. *Review of Economic Studies* 71:285–315.

Almeida, Heitor, and Murillo Campello. 2007. Financial constraints, asset tangibility, and corporate investment. *Review of Financial Studies* 20:1429 – 1460.

Audretsch, David B. 1995. *Innovation and Industry Evolution.* MIT Press, Cambridge, Massachusetts.

Bayer, Christian. 2006. Investment dynamics with fixed capital adjustment cost and capital market imperfections. *Journal of Monetary Economics* 53:1909–1947.

—. 2008. On the interaction of financial frictions and fixed capital adjustment costs: Evidence from a panel of german firms. *Journal of Economic Dynamics and Control* 32:3538–3559.

Berk, Johnathan B., Richard C. Green, and Vasant Naik. 2004. Valuation and return dynamics of new ventures. *Review of Financial Studies* 17:1–35.

Biørn, Erik. 2004. Regression systems for unbalanced panel data: A stepwise maximum likelihood procedure. *Journal of Econometrics* 122:281–291.

Blundell, Richard, and James Powell. 2003. Endogeneity in nonparametric and semiparametric regression models. In *Advances in Economics and Econonometrics: Theory and Applications, Eighth World Congress, Vol. II*, edited by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge: Cambridge University Press.

Brown, James R., Steven M. Fazzari, , and Bruce C. Petersen. 2009. Financing innovation and growth: Cash flow, external equity, and the 1990's R&D boom. *Journal of Finance* 64:151–185.

Brown, James R., Gustav Martinsson, and Bruce C Petersen. 2012. Do financing constraints matter for R&D? *European Economic Review* 58:1512–1529.

Carpenter, Robert E, and Bruce C Petersen. 2002. Capital market imperfections, high-tech investment, and new equity financing. *Economic Journal* 112:F54–F72.

Cassiman, Bruno, Massimo Colombo, Paola Garrone, and Reinhilde Veugelers. 2005. The impact of M&A on the R&D process. *Research Policy* 34:195–220.

Chan, Louis K. C., Josef Lakonishok, and Theodore Sougiannis. 2001. The stock market valuation of research and development expenditures. *The Journal of Finance* 56:24312456.

Clementi, Gian L., and Hugo A. Hopenhayn. 2006. A theory of financing constraints and firm dynamics. *The Quarterly Journal of Economics* 121:229–265.

Cooley, Thomas F., and Vincenzo Quadrini. 2001. Financial markets and firm dynamics. *American Economic Review* 91:1286–1310.

Cooper, Russell W., and John C. Haltiwanger. 2006. On the nature of capital adjustment costs. *Review of Economic Studies* 73:611–633.

Fazzari, Steven M., Robert G. Hubbard, and Bruce C. Petersen. 1988. Financing constraints and corporate investment. *Brookings Papers on Economic Activity* 1:144–195.

Gale, Douglas, and Martin Hellwig. 1985. Incentive compatible debt contracts: The one period problem. *Review of Economic Studies* 52:647–663.

Gamba, Andrea, and Alexander Triantis. 2008. The value of financial flexibility. *Journal of Finance* 63:2263–2296.

Gomes, Joao F. 2001. Financing investment. *American Economic Review* 91:1263–1285.

Gomes, Joao F., Amir Yaron, and Lu Zhang. 2006. Asset pricing implications of firms' financing constraints. *Review of Financial Studies* 19:1321–1356.

Hajivassiliou, Vassilis, and Frédérique Savignac. 2011. Novel approaches to coherency conditions in LDV models with an application to interactions between financing constraints and a firm's decision and ability to innovate. LSE discussion papers.

Hall, Bronwyn, and Josh Lerner. 2010. The financing of R&D and innovation. In *Handbook of the Economics of Innovation*, edited by B. H. Hall and N. Rosenberg. Cambridge, forthcoming: Cambridge University Press.

Hennessy, Christopher A, and Toni M. Whited. 2007. How costly is external financing? evidence from a structural estimation. *Journal of Finance* 62:1705–1745.

Holmstrom, Bengt. 1989. Agency costs and innovation. *Journal of Economic Behavior and Organization* 12:305–327.

Huergo, Elena, and Jordi Jaumandreu. 2004. How does probability of innovation change with firm age? *Small Business Economics* 22:193–207.

Kaplan, Steven N., and Luigi Zingales. 1997. Do investment-cash flow sensitivities provide useful measures of financing constraints? *Quarterly Journal of Economics* 112:169–215.

Klepper, Steven, and Peter Thompson. 2006. Submarkets and the evolution of market structure. *RAND Journal of Economics* 37:861–886.

Klette, Tor J., and Samuel S. Kortum. 2004. Innovating firms and aggregate innovation. *Journal of Political Economy* 112:986–1018.

Lambrecht, Bart M., and Stewart C. Myers. 2008. Debt and managerial rents in a real-options model of the firm. *Journal of Financial Economics* 89:209–231.

Leland, Hayne E., and David H. Pyle. 1977. Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance* 32:371–387.

Moyen, Nathalie. 2004. Investment-cash flow sensitivities: Constrained versus unconstrained firms. *Journal of Finance* 69:2061–2092.

Myers, Stewart C. 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5:147–175.

Papke, Leslie E., and Jeffery M. Wooldridge. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145:121–133.

Raymond, Wladimir, Pierre Mohnen, Franz C. Palm, and Sybrand Schim van der Loeff. 2010. Persistence of innovation in dutch manufacturing: Is it spurious? *Review of Economics and Statistics* 92:495–504.

Reddick, Leigh A., and Toni M. Whited. 2009. The corporate propensity to save. *Journal of Financial Economics* 64:1729–1766.

Roberts, Michael R., and Toni M. Whited. 2010. Endogeneity in empirical corporate finance. In *Handbook of the Economics of Finance*, edited by George Constantinides, Milton Harris, and Rene Stulz, vol. 2. Elsevier, Amsterdam.

Semykina, Anastasia, and Jefferey M. Wooldridge. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* 157:375–380.

Titman, Sheridan, and Roberto Wessels. 1988. The determinants of capital structure choice. *Journal of*

Finance 43:1–19.

Whited, Toni M. 2006. External finance constraints and the intertemporal pattern of intermittent investment. *Journal of Financial Economics* 81:467–502.

Whited, Toni M., and Guojun Wu. 2005. Financial constraints risk. *Review of Financial Studies* 19:531–559.

TABLE III

Total number of enterprises, N_f, and number of enterprises surveyed within a firm, n_f

The table illustrates the number of firms, in each of the three CIS waves, for which the number of number of enterprises surveyed is equal to the number of enterprises present in the firm, $N_f = n_f$, and the number of firms, for which the number of enterprises present in the firm exceeds the number of enterprises surveyed. These figures pertain to the CIS data set prior to merging with the SF data set. Since not all the CIS firms are in the SF data set, the CIS data used for estimation after cleaning is a bit less than half the size of the original data set.

CIS2.5			CSI3			CIS3.5		
	No. of firms for which			No. of firms for which			No. of firms for which	
N_f	$N_f = n_f$	$N_f > n_f$	N_f	$N_f = n_f$	$N_f > n_f$	N_f	$N_f = n_f$	$N_f > n_f$
1	9400	0	1	6155	0	1	7096	0
2	151	1255	2	67	823	2	137	978
3	20	608	3	4	424	3	24	553
4	3	316	4	3	237	4	2	290
5	3	247	5	2	108	5		222
6		149	6		115	6		122
7		107	7		48	7		105
8		60	8		77	8		50
9	2	93	9		58	9		77
10		83	10		39	10		82
11		106	11		63	11		50
12		49	12		39	12		58
13		43	13		15	13		49
14		59	14		50	14		46
15		46	15		17	15		25
16		31	16		28	16		51
17		62	17		15	17		15
18		36	18		26	18		55
19		37	19		13	19		8
20		29	20		21	20		28
21		13	21		2	21		43
22		23	22		27	22		36
23		15	24		5	23		18
25		34	25		9	24		25
26		46	26		8	25		11
27		4	27		21	27		17
29		14	28		13	28		19
30		14	29		8	29		11
31		18	30		8	30		15
32		15	31		3	31		7
33		11	32		16	32		16
34		18	34		22	33		25
37		15	40		10	37		21
38		15	45		14	38		13
43		15	48		18	39		20
44		17	50		19	40		9
45		14	57		16	41		10
48		20	60		16	46		15
49		22				50		16
51		28				53		47
56		19				55		16
66		33						
85		41						

TABLE IV

Means of Variables for Innovators and Non-Innovators

	CIS2.5		CSI3		CIS3.5	
	Innovator	Non-Innovator	Innovator	Non-Innovator	Innovator	Non-Innovator
R&D*	0.506		0.338		0.192	
Share of Innovative Sales in Total Sales (%)	8.532		10.944		8.025	
Long-term Debt*	0.789	0.834	0.739	0.8080	1.149	0.954
Cash flow*	0.869	0.841	0.638	1.167	0.589	0.352
Dummy for Multiple Enterprises	0.369	0.019	0.478	0.008	0.539	0.019
Liquidity Reserve*	0.913	1.837	0.840	1.689	1.152	1.532
Dividends*	0.082	0.133	0.089	0.268	0.176	0.253
Market Share (%)	0.926	0.067	1.295	0.073	1.267	0.099
Size (Log of Employed)	5.038	4.007	4.808	3.304	4.980	3.759
Age	21.696	19.489	24.817	21.978	25.131	21.109
Ratio of Intangible to Total Assets (%)	4.284	2.771	5.254	2.230	7.773	2.702
Dummy for Negative Cash flow	0.069	0.110	0.079	0.109	0.119	0.135
No. of Observations	2,947	2,416	1,844	1,579	1,980	2,266

* Variables normalized by total capital assets

TABLE V

Second Stage Coefficient Estimates: Financial Constraints and Innovation

	Specification 1		**Specification 2**		**Specification 3**	
Variables of interest	**Financial Constraints**	**Innovation**	**Financial Constraints**	**Innovation**	**Financial Constraints**	**Innovation**
Share of Innovative Sales	0.201***		0.206***		0.206***	
	(0.024)		(0.021)		(0.021)	
Long Term Debt	0.781***	-0.366***	0.788***	-0.366***	0.788***	-2.292***
	(0.247)	(0.108)	(0.248)	(0.108)	(0.248)	(0.133)
Cash flow	0.313***		0.317***		0.317***	
	(0.041)		(0.041)		(0.041)	
Dummy for Negative Cash flow	0.99***		1.018***		1.018***	
	(0.116)		(0.097)		(0.097)	
Liquidity Reserve	-0.26***	0.515***	-0.298***	0.515***	-0.298***	1.524***
	(0.086)	(0.095)	(0.038)	(0.095)	(0.038)	(0.121)
Dividends	-3.624***	0.019	-3.677***	0.019	-3.677***	-0.096***
	(0.454)	(0.018)	(0.452)	(0.018)	(0.452)	(0.018)
Size	-0.49***	0.29***	-0.486***	0.29***	-0.486	0.741***
	(0.069)	(0.033)	(0.067)	(0.033)	(0.067)	(0.042)
Market Share	0.008	0.131***	0.004	0.131***	0.004	0.059***
	(0.008)	(0.021)	(0.004)	(0.021)	(0.004)	(0.021)
Age	-0.011**	-0.012***	-0.011***	-0.012***	-0.011***	-0.017***
	(0.004)	(0.002)	(0.004)	(0.002)	(0.004)	(0.002)
Ratio of Intangible Assets to Total Assets	0.041	-0.259***	0.056***	-0.259***	0.056***	0.175***
	(0.029)	(0.03)	(0.014)	(0.03)	(0.014)	(0.024)
Dummy for Multiple Enterprise Firms	0.082	3.177***		3.177***		2.041***
	(0.162)	(0.172)		(0.172)		(0.155)
Control Functions† for						
Share of Innovative Sales	-1.328***	0.549***	-1.378***	0.549***	-1.378***	
	(0.184)	(0.031)	(0.154)	(0.031)	(0.154)	
Long-term Debt	-6.209***	2.633***	-6.217***	2.633***	-6.217***	18.626***
	(2.198)	(0.892)	(2.199)	(0.892)	(2.199)	(1.06)
Dividends	17.387***	-2.105***	17.787***	-2.105***	17.787***	-4.964***
	(2.058)	(0.369)	(1.98)	(0.369)	(1.98)	(0.443)
Liquidity Reserve	7.637***	-5.833***	8.164***	-5.833***	8.164***	-15.145***
	(1.089)	(1.044)	(0.404)	(1.044)	(0.404)	(1.288)
Ratio of Intangible to Total Assets	-1.209**	5.286***	-1.517***	5.286***	-1.517***	-2.749***
	(0.59)	(0.609)	(0.257)	(0.609)	(0.257)	(0.476)
Size	-0.871***	0.775***	-0.937***	0.775***	-0.937***	2.044***
	(0.167)	(0.164)	(0.111)	(0.164)	(0.111)	(0.189)
Individual Effects $(\bar{\mathcal{Z}}_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)$	-0.729***	-0.265***	-0.688***	-0.265***	-0.688***	1.779***
	(0.187)	(0.084)	(0.16)	(0.084)	(0.16)	(0.102)
$\rho_{\tilde{\zeta}\tilde{\upsilon}}$	0.589***		0.589***		0.589***	
	(0.033)		(0.033)		(0.033)	
Total Number of Observations: 13032						

Significance levels : $*$: 10% $**$: 5% $***$: 1%

†The estimated coefficients of the Control Function for Share of Innovative Sales, Long-term Debt, Dividends, Liquidity Reserve, Ratio of Intangible to total Assets, and Size are the estimated terms in $\tilde{\Sigma}_{\upsilon\epsilon} = \{\rho_{\upsilon\epsilon_1}\sigma_\upsilon, \ldots, \rho_{\upsilon\epsilon_m}\sigma_\upsilon\}$ of equation (3.12) and $\tilde{\Sigma}_{\zeta\epsilon} = \{\rho_{\zeta\epsilon_1}\sigma_\zeta, \ldots, \rho_{\zeta\epsilon_m}\sigma_\zeta\}$ of equation (3.13).

TABLE VI

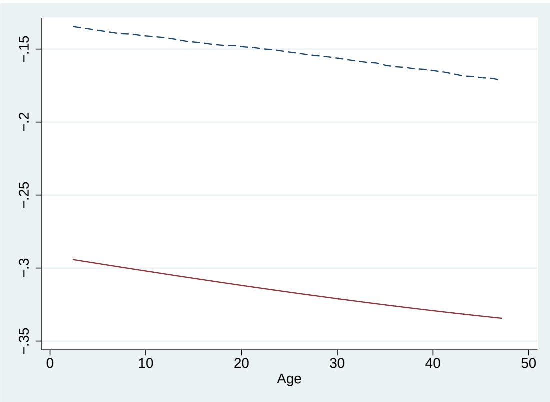
Average Partial Effects of Second Stage Estimates

	Specification 1		**Specification 2**		**Specification 3**	
	Eq. (14)	**Eq. (15)**	**Eq. (14)**	**Eq. (15)**	**Eq. (14)**	**Eq. (15)**
	Financial Constraints	**Innovation**	**Financial Constraints**	**Innovation**	**Financial Constraints**	**Innovation**
Share of Innovative Sales	0.028***		0.028***		0.028***	
	(0.004)		(0.003)		(0.003)	
Long Term debt	0.107***	-0.091***	0.108***	-0.091***	0.108***	-0.3***
	(0.034)	(0.025)	(0.034)	(0.027)	(0.034)	(0.01)
Cash flow	0.043***		0.043***		0.043***	
	(0.006)		(0.006)		(0.006)	
Dummy for Negative Cash flow	0.163***		0.166***		0.166***	
	(0.02)		(0.019)		(0.019)	
Liquidity Reserve	-0.036***	0.127***	-0.041***	0.127***	-0.041***	0.199***
	(0.013)	(0.023)	(0.005)	(0.024)	(0.005)	(0.013)
Dividends	-0.497***	0.005	-0.502***	0.005	-0.502***	-0.013***
	(0.066)	(0.005)	(0.062)	(0.005)	(0.062)	(0.002)
Size	-0.067***	0.072***	-0.066***	0.072***	-0.066***	0.097***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.005)
Market Share	0.001	0.032***	0.001	0.032***	0.001	0.008***
	(0.001)	(0.005)	(0.001)	(0.005)	(0.001)	(0.003)
Age	-0.001**	-0.003***	-0.002***	-0.003***	-0.002***	-0.002***
	(0.001)	(0)	(0.001)	(0)	(0.001)	(0)
Ratio of Intangible Assets to Total Assets	0.006	-0.064***	0.008***	-0.064***	0.008***	0.023***
	(0.004)	(0.008)	(0.002)	(0.008)	(0.002)	(0.003)
Dummy for Multiple Enterprise Firms	0.011	0.555***		0.621***		0.866***
	(0.023)	(0.097)		(0.013)		(0)

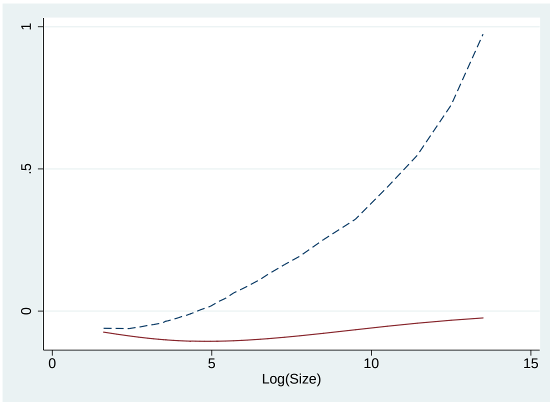
Significance levels : $*$: 10% $**$: 5% $***$: 1%

Figure 1: Plot of APE of Long-term Debt on the Probability of Innovation conditional on being Financially Constrained, $\int \frac{\partial \Pr(I=1|F=1,\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}})}{\partial DEBT} dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}}$, or *not* Financially Constrained, $\int \frac{\partial \Pr(I=1|F=0,\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}})}{\partial DEBT} dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}}$, against Age, Size, and Leverage.

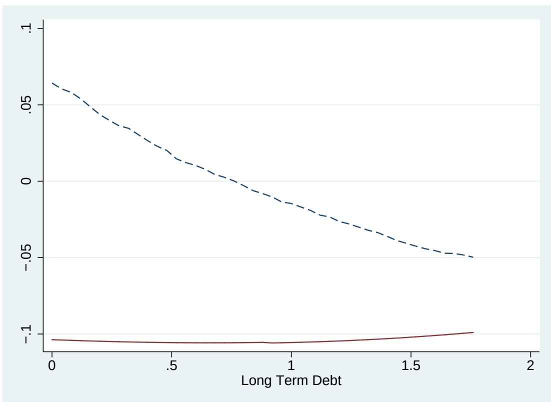
(a) Age



(b) Log of Employed



(c) Long- Term Debt



$$\int \frac{\partial \Pr(I=1|F=1,\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}})}{\partial DEBT} dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}} \text{ - - - - - - }, \int \frac{\partial \Pr(I=1|F=0,\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}})}{\partial DEBT} dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}} \text{ ——— }$$

TABLE VII

Third Stage Estimates: R&D Switching Regression Model

Variables of Interest	Specification 1	Specification 2 No Control Function for Size	**Control Functions†**	Specification 1	Specification 2 No Control Function for Size
f, Binary variable for Financial Constraint	-1.049 (0.661)	-0.84** (0.408)	**For Financially Constrained Firms**		
$f*$ Share of Innovative Sales	0.217*** (0.018)	0.219*** (0.017)	Share of Innovative Sales	-1.559*** (0.159)	-1.597*** (0.141)
$(1-f)*$Share of Innovative Sales	0.201*** (0.018)	0.205*** (0.015)	Long-trem Debt	0.525** (0.213)	0.511** (0.215)
$f*$ Cash flow	0.07* (0.041)	0.071* (0.041)	Dividends	-1.296*** (0.39)	-1.232*** (0.363)
$(1-f)*$ Cash flow	0.005 (0.003)	0.005 (0.003)	Liquidity Reserve	-0.395 (0.291)	-0.352 (0.27)
$f*$Dummy for Multiple Enterprise	0.799*** (0.245)	0.682*** (0.158)	Ratio of Intangible to Total Assets	-0.034 (0.046)	-0.036 (0.046)
$(1-f)*$ Dummy for Multiple Enterprise	0.514*** (0.189)	0.429*** (0.078)	Size	0.067 (0.106)	
$f*$Market Share	0.027* (0.015)	0.019** (0.009)	Financial Constraint $(C_{11}(.)_t)$	0.967*** (0.319)	0.83*** (0.209)
$(1-f)*$Market share	0.011 (0.012)	0.005 (0.004)	Selection $(C_{12}(.)_t)$	0.636* (0.326)	0.589* (0.306)
$f*$Size	-0.494*** (0.118)	-0.431*** (0.071)	Individual effects $(\bar{\mathcal{Z}}_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)$	-0.413* (0.236)	-0.297** (0.142)
$(1-f)*$Size	-0.364*** (0.102)	-0.318*** (0.035)	**For Financially Unconstrained Firms**		
$f*$Age	-0.012*** (0.004)	-0.012*** (0.004)	Share of Innovative Sales	-1.52*** (0.164)	-1.57*** (0.125)
$(1-f)*$Age	-0.002 (0.002)	-0.003** (0.001)	Long-trem Debt	-0.029 (0.084)	-0.034 (0.08)
			Dividends	0.022 (0.053)	0.027 (0.051)
			Liquidity Reserve	0.18*** (0.063)	0.189*** (0.058)
			Ratio of Intangible to Total Assets	-0.089*** (0.013)	-0.092*** (0.012)
			Size	0.034 (0.074)	
			Financial Constraint $(C_{01}(.)_t)$	-0.277 (0.198)	-0.186** (0.065)
			Selection $(C_{02}(.)_t)$	-0.883*** (0.324)	-0.745*** (0.114)
			Individual effects $(\bar{\mathcal{Z}}_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)$	0.346*** (0.091)	0.312*** (0.064)
Average Partial Effect of Financial Constraint	-0.241 (0.7)	-0.175 (0.393)			
Total Number of Observations: 6771					

Significance levels : $*$: 10% $**$: 5% $***$: 1%

†The estimated coefficients of the Control Function for Share of Innovative Sales, Long-term Debt, Dividends, Liquidity Reserve, Ratio of Intangible to total Assets, and Size are the terms in $\tilde{\Sigma}_{\eta_1\epsilon} = \{\rho_{\eta_1\epsilon_1}\sigma_{\eta_1}, \ldots, \rho_{\eta_1\epsilon_m}\sigma_{\eta_1}\}$ for firms that are financially constrained and $\tilde{\Sigma}_{\eta_0\epsilon} = \{\rho_{\eta_0\epsilon_1}\sigma_{\eta_0}, \ldots, \rho_{\eta_0\epsilon_m}\sigma_{\eta_0}\}$ for firms that are not financially constrained of the R&D equation (3.22).

SUPPLEMENT TO "MICROECONOMETRIC EVIDENCE OF FINANCING FRICTIONS AND INNOVATIVE ACTIVITY": SUPPLEMENTARY APPENDIX

Amaresh K Tiwari[1], Pierre Mohnen[2], Franz C Palm[3]
and Sybrand Schim van der Loeff[4]

Not meant to be included with the main text of the paper.

APPENDIX A: IDENTIFICATION OF STRUCTURAL PARAMETERS WITH EXPECTED A POSTERIORI VALUES OF INDIVIDUAL EFFECTS

We began with a set of structural equations

$$\mathbf{y}_t^* = \mathbb{X}_t'\mathbf{B} + \tilde{\alpha}\mathbf{k} + \Upsilon_t, \tag{A-1}$$

and a set of reduced m form equations for the endogenous regressors in the above equation,

$$\mathbf{x}_t = \mathbf{Z}_t'\boldsymbol{\delta} + \tilde{\alpha}\boldsymbol{\kappa} + \boldsymbol{\epsilon}_t, \tag{A-2}$$

The distributional assumptions that we made, which eventually will allow us to construct the control functions that correct for the bias due to the endogeneity of $\mathbf{x}_t$ and help us identify the structural parameters of interest are:

A1. $\Upsilon_{it}|\tilde{\alpha}_i, \mathcal{Z}_i \sim \Upsilon_{it}|\tilde{\alpha}_i$ and $\boldsymbol{\epsilon}_{it}|\tilde{\alpha}_i, \mathcal{Z}_i \sim \boldsymbol{\epsilon}_{it}$,

A2. $\Upsilon_{it}|\tilde{\alpha}_i, \boldsymbol{\epsilon}_i \sim \Upsilon_{it}|\boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i = \{\boldsymbol{\epsilon}_{i1}' \ldots \boldsymbol{\epsilon}_{iT_i}'\}'$, and

A3. The error terms Υ_{it} and $\boldsymbol{\epsilon}_{it}$ are i.i.d. and jointly distributed as

$$\begin{pmatrix} \Upsilon_{it} \\ \boldsymbol{\epsilon}_{it} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \Sigma_{\Upsilon\Upsilon} & \Sigma_{\Upsilon\epsilon} \\ \Sigma_{\epsilon\Upsilon} & \Sigma_{\epsilon\epsilon} \end{pmatrix}\right].$$

We also specified the conditional expectation and the distribution of the individual effects $\tilde{\alpha}_i$.We assumed that

$$\text{A4.} \qquad \mathrm{E}(\tilde{\alpha}_i|\mathcal{Z}_i) = \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}},$$

[1]University of Liege, A.Tiwari@ulg.ac.be
[2] Maastricht University, UNU-MERIT and CIRANO, P.Mohnen@maastrichtuniversity.nl
[3]Maastricht University and CESifo fellow, F.Palm@maastrichtuniversity.nl
[4] Maastricht University, S.Loeff@maastrichtuniversity.nl

where $\bar{\mathcal{Z}}_i$, is the mean of time-varying variables in $\mathcal{Z}_{it}$. We also assumed that

$$\text{A5.} \quad \tilde{\alpha}_i | \mathcal{Z}_i \sim \mathrm{N}\left[\mathrm{E}(\tilde{\alpha}_i|\mathcal{Z}_i), \sigma^2_\alpha\right],$$

so that the tail, $\alpha_i = \tilde{\alpha}_i - \mathrm{E}(\tilde{\alpha}_i|\mathcal{Z}_i) = \tilde{\alpha}_i - \bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}}$, is distributed normally with mean zero and variance σ^2_α, and was assumed to be independent of $\mathcal{Z}_i$[1].

These assumptions gave us equations (3.8), where

$$\begin{aligned} \Upsilon_t|\mathbf{X}, \mathcal{Z}, \tilde{\alpha} &\sim \Upsilon_t|\mathbf{X} - \mathrm{E}(\mathbf{X}|\mathcal{Z}, \tilde{\alpha}), \mathcal{Z}, \tilde{\alpha} \\ &\sim \Upsilon_t|\boldsymbol{\epsilon}, \mathcal{Z}, \tilde{\alpha} \\ &\sim \Upsilon_t|\boldsymbol{\epsilon}, \tilde{\alpha} \\ &\sim \Upsilon_t|\boldsymbol{\epsilon}, \end{aligned} \tag{A-3}$$

where the second equality in distribution follows from the fact that $\mathbf{X}_i - \mathrm{E}(\mathbf{X}_i|\mathcal{Z}_i, \tilde{\alpha}_i) = \boldsymbol{\epsilon}_i$, the third follows from A1, and the fourth from assumption A2. According to the above, the dependence of the structural error term Υ_t on $\mathbf{X}$, $\mathcal{Z}$, and $\tilde{\alpha}$ is completely characterized by the reduced form errors $\boldsymbol{\epsilon}$. The expectation of Υ_t given $\boldsymbol{\epsilon}$ in (3.9) was given by

$$E(\Upsilon_t|\boldsymbol{\epsilon}) = E(\Upsilon_t|\boldsymbol{\epsilon}_t) = \Sigma_{\Upsilon\epsilon}\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_t = \tilde{\Sigma}_{\Upsilon\epsilon}\Sigma_\epsilon\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_t = \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_t, \tag{A-4}$$

where the first equality followed from the assumption that conditional on $\boldsymbol{\epsilon}_{it}$, Υ_{it} is independent of $\boldsymbol{\epsilon}_{i_{-t}}$. This assumption has also been made in Papke and Wooldridge (2008), and Semykina and Wooldridge (2010). The $(4 \times m)$ matrices $\tilde{\Sigma}_{\Upsilon\epsilon}$ in the fourth equality is

$$\tilde{\Sigma}_{\Upsilon\epsilon} = \begin{pmatrix} \rho_{\eta_1\epsilon_1}\sigma_{\eta_1} & \dots & \rho_{\eta_1\epsilon_m}\sigma_{\eta_1} \\ \rho_{\eta_0\epsilon_1}\sigma_{\eta_0} & \dots & \rho_{\eta_0\epsilon_m}\sigma_{\eta_0} \\ \rho_{\zeta\epsilon_1}\sigma_{\zeta} & \dots & \rho_{\zeta\epsilon_m}\sigma_{\zeta} \\ \rho_{\upsilon\epsilon_1}\sigma_{\upsilon} & \dots & \rho_{\upsilon\epsilon_m}\sigma_{\upsilon} \end{pmatrix}$$

and the $(m \times m)$ matrix Σ_ϵ is $\mathrm{diag}(\sigma_{\epsilon 1}, \dots, \sigma_{\epsilon m})$, so that $\tilde{\Sigma}_{\Upsilon\epsilon}\Sigma_\epsilon = \Sigma_{\Upsilon\epsilon}$. Finally, in the last equality $\tilde{\Sigma}^{-1}_{\epsilon\epsilon} = \Sigma_\epsilon\Sigma^{-1}_{\epsilon\epsilon}$. The assumption about the conditional distribution $\tilde{\alpha}$ and equations (A-3) and (A-4) led us the relationship in (3.10):

$$\mathrm{E}(\mathbf{y}^*_t|\mathbf{X}, \mathcal{Z}, \alpha) = \mathbb{X}_t'\mathbf{B} + (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \alpha)\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_t.$$

[1]In the rest of the appendix, except when needed, we will suppress the firm subscript i.

Given the above, we can write the linear predictor of $\mathbf{y}_t^*$ in error form as

$$\mathbf{y}_t^* = \mathbb{X}_t'\mathbf{B} + (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \alpha)\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t + \tilde{\Upsilon}_t. \tag{A-5}$$

We had argued that in order to estimate the system of equations in (A-3) the standard technique of the control function approach is to replace $\boldsymbol{\epsilon}_t$ by the residuals from the first stage reduced form regression. However, the residuals $\mathbf{x}_t - \mathrm{E}(\mathbf{x}_t|\mathcal{Z}, \alpha) = \mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \alpha)\boldsymbol{\kappa}$, remain unidentified because the α's are unobserved even though the reduced form parameters, $\boldsymbol{\delta}$, $\bar{\boldsymbol{\delta}}$, and $\boldsymbol{\kappa}$, can be consistently estimated from the first stage estimation of the modified reduced form equation given in (3.4a).

However, it can still be possible to estimate the structural parameters if we can integrate out the α's. But given that α's are correlated with the endogenous regressors we have to integrate it out with respect to its conditional distribution. Let $\mathbf{f}(\alpha_i|\mathbf{X}_i, \mathcal{Z}_i)$ be the conditional distribution of time invariant individual effect α_i conditional on $\mathbf{X}_i$ and $\mathcal{Z}_i$. For any firm, i, taking expectation of the above with respect to the conditional distribution of α, $\mathbf{f}(\alpha|\mathbf{X}, \mathcal{Z})$ we obtain

$$\begin{aligned}
E(\mathbf{y}_t^*|\mathbf{X}, \mathcal{Z}) &= \int \mathrm{E}(\mathbf{y}_t^*|\mathbf{X}, \mathcal{Z}, \alpha)\mathbf{f}(\alpha|\mathbf{X}, \mathcal{Z})d\alpha \\
&= \mathbb{X}_t'\mathbf{B} + \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + \int (\mathbf{k} - \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa})\alpha\mathbf{f}(\alpha|\mathbf{X}, \mathcal{Z})d\alpha \\
&= \mathbb{X}_t'\mathbf{B} + \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + \int (\mathbf{k} - \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa})\alpha\mathbf{f}(\alpha|\mathbf{X})d\alpha \\
&= \mathbb{X}_t'\mathbf{B} + \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + (\mathbf{k} - \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa})\hat{\alpha} \\
&= \mathbb{X}_t'\mathbf{B} + (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha})\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha})\boldsymbol{\kappa}) \\
&= \mathbb{X}_t'\mathbf{B} + \hat{\tilde{\alpha}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t,
\end{aligned} \tag{A-6}$$

where the second equality follows from the fact that $\mathcal{Z}$ and α are independent. $\hat{\alpha}_i = \hat{\alpha}_i(\mathbf{X}_i, \mathcal{Z}_i, \Theta_1)$ is the expected a posteriori (EAP) value of time invariant individual effects $\boldsymbol{\alpha}_i$, and Θ_1 is the set of first stage reduced form parameters.

To obtain the EAP values, $\hat{\alpha}_i$, in (A-6), we use Bayes rule we can write $\mathbf{f}(\alpha|\mathbf{X}, \mathcal{Z})$ as

$$\mathbf{f}(\alpha|\mathbf{X}) = \frac{\mathbf{f}(\mathbf{X}|\alpha)\mathbf{g}(\alpha)}{\mathbf{h}(\mathbf{X})} = \frac{\mathbf{f}(\mathbf{X}, \mathcal{Z}|\alpha)\mathbf{g}(\alpha)}{\mathbf{h}(\mathbf{X}, \mathcal{Z})}, \tag{A-7}$$

where $\mathbf{g}$ and $\mathbf{h}$ are density functions. The above can be written as

$$\mathbf{f}(\alpha|\mathbf{X},\mathcal{Z}) = \frac{\mathbf{f}(\mathbf{X}|\mathcal{Z},\alpha)\mathbf{p}(\mathcal{Z}|\alpha)\mathbf{g}(\alpha)}{\mathbf{h}(\mathbf{X}|\mathcal{Z})\mathbf{p}(\mathcal{Z})},$$

By our assumption the, αs are independent of the exogenous variables $\mathcal{Z}$, hence $\mathbf{p}(\mathcal{Z}|\alpha) = \mathbf{p}(\mathcal{Z})$, that is,

$$\mathbf{f}(\alpha|\mathbf{X},\mathcal{Z}) = \frac{\mathbf{f}(\mathbf{X}|\mathcal{Z},\alpha)\mathbf{g}(\alpha)}{\mathbf{h}(\mathbf{X}|\mathcal{Z})} = \frac{\mathbf{f}(\mathbf{X}|\mathcal{Z},\alpha)\mathbf{g}(\alpha)}{\int \mathbf{f}(\mathbf{X}|\mathcal{Z},\alpha)\mathbf{g}(\alpha)d\alpha}, \tag{A-8}$$

Hence,

$$\begin{aligned}
\int \alpha\mathbf{f}(\alpha|\mathbf{X},\mathcal{Z})d(\alpha) &= \int \frac{\alpha\mathbf{f}(\mathbf{X}|\mathcal{Z},\alpha)\mathbf{g}(\alpha)d\alpha}{\int \mathbf{f}(\mathbf{X}|\mathcal{Z},\alpha)\mathbf{g}(\alpha)d\alpha} \\
&= \frac{\int \alpha \prod_{t=1}^{T} \mathbf{f}(\mathbf{x}_t|\mathcal{Z},\alpha)\mathbf{g}(\alpha)d\alpha}{\int \prod_{t=1}^{T} \mathbf{f}(\mathbf{x}_t|\mathcal{Z},\alpha)\mathbf{g}(\alpha)d\alpha} \\
&= \hat{\alpha}(\mathbf{X},\mathcal{Z},\Theta_1)
\end{aligned} \tag{A-9}$$

where the second equality follow from the fact that conditional on $\mathcal{Z}$ and α, each of the $\mathbf{x}_t$, $\mathbf{x}_t \in \{\mathbf{x}_1,\ldots,\mathbf{x}_T\}$ are independently normally distributed with mean $\mathbf{Z}_t'\boldsymbol{\delta} + (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \alpha)\boldsymbol{\kappa}$ and standard deviation $\Sigma_{\epsilon\epsilon}$. $\mathbf{g}(\alpha)$ by our assumption is normally distributed with mean zero and variance σ_α^2 and $\mathfrak{a} = \frac{\alpha}{\sigma_\alpha}$ follows a standard normal distribution. The functional form of $\hat{\alpha}(\mathbf{X},\mathcal{Z},\Theta_1)$ is given by:

$$\hat{\alpha}(\mathbf{X},\mathcal{Z},\Theta_1) = \frac{\int \sigma_\alpha\mathfrak{a} \prod_{t=1}^{T} \exp(-\frac{1}{2}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha\mathfrak{a})\boldsymbol{\kappa})^T\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha\mathfrak{a})\boldsymbol{\kappa}))\phi(\mathfrak{a})d\mathfrak{a}}{\int \prod_{t=1}^{T} \exp(-\frac{1}{2}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha\mathfrak{a})\boldsymbol{\kappa})^T\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha\mathfrak{a})\boldsymbol{\kappa}))\phi(\mathfrak{a})d\mathfrak{a}}. \tag{A-10}$$

The right hand side of (A-10) is the expected a posteriori (EAP) value of α. $\hat{\hat{\alpha}}(\mathbf{x},\mathcal{Z},\hat{\Theta}_1)$ is the estimated expected a posteriori value of α, which can be estimated by employing numerical integration techniques, such as Gauss-Hermite quadratures, with respect to α at the estimated Θ_1 from the first stage. Also, it can be shown that

Lemma 1 *$\hat{\hat{\alpha}}_i(\boldsymbol{X}_i,\mathcal{Z}_i,\hat{\Theta}_1)$ converges a.s. to $\hat{\alpha}_i(\boldsymbol{X}_i,\mathcal{Z}_i,\Theta_1)$, where $\hat{\Theta}_1$ is the consistent first stage estimates.*

Proof of Lemma 1 *Given in* Section A.1.

Lemma 1 implies that

$$\mathbb{X}_t'\mathbf{B} + (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\hat{\alpha}})\mathbf{k} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\hat{\boldsymbol{\epsilon}}}_t, \stackrel{a.s}{\to} E(\mathbf{y}_t^*|\mathbf{X},\mathcal{Z}) = \int E(\mathbf{y}_t^*|\mathbf{X},\mathcal{Z},\alpha)\mathbf{f}(\alpha|\mathbf{X},\mathcal{Z})d(\alpha),$$

where $\hat{\hat{\boldsymbol{\epsilon}}}_t = \mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\hat{\alpha}})\boldsymbol{\kappa}$. Therefore, if the reduced form population parameters, Θ_1, are known, the above implies that we could write the linear predictor of $\mathbf{y}_{it}^*$, given $\mathbf{X}_i$ and $\mathcal{Z}_i$ in error form as

$$\mathbf{y}_t^* = \mathbb{X}_t'\mathbf{B} + \hat{\hat{\alpha}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \tilde{\Upsilon}_t, \tag{A-11}$$

where $\hat{\hat{\alpha}} = \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha}$ and conditional of $\mathbf{X}$ and $\mathcal{Z}$, $\tilde{\Upsilon}_t$ is i.i.d. with mean 0. For linear models, say if all the variables in $\mathbf{y}_t^*$ were continuous and observed, with estimates of $\hat{\hat{\alpha}}$, and the estimates of $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t$ the parameters of interest, $\mathbf{B}$, can be consistently estimated by running a seemingly unrelated regression (SUR) or a panel version of SUR to gain efficiency. We note here that for any k, $k \in \{1,\ldots,n\}$, $n = 4$ in our model, $\tilde{\Sigma}_{\Upsilon_k\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t$ take the form

$$\rho_{\Upsilon_k\epsilon_1}\sigma_{\Upsilon_k}f_1(\Sigma_{\epsilon\epsilon},\hat{\epsilon}_{1t},\ldots,\hat{\epsilon}_{mt}) + \ldots + \rho_{\Upsilon_k\epsilon_m}\sigma_{\Upsilon_k}f_m(\Sigma_{\epsilon\epsilon},\hat{\epsilon}_{1t},\ldots,\hat{\epsilon}_{mt})$$

where each of the f's above are linear in $\hat{\boldsymbol{\epsilon}}_t$. The estimates $\rho_{\Upsilon_k\epsilon_l}\sigma_{\Upsilon_k}$, $l \in \{1,\ldots,m\}$, provides us with a test of exogeneity of the regressor x_l with respect to Υ_k.

However, when response outcomes are discrete and we have to deal with nonlinear models additional assumptions than those made above are required. Let us consider F_t^* of $\mathbf{y}_t^*$ where F_t^* is the latent variable underlying F_t, the binary variable that takes value 1 when the firm is financially constrained and 0 otherwise.

$$F_t = 1\{F_t^* > 0\} = 1\{\mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\tilde{\alpha} + \zeta_t > 0\} = H(\mathcal{X}_t^F,\alpha,\zeta_t), \tag{A-12}$$

For a firm i, what we are interested is the Average Structural Function (ASF),

$$\mathrm{E}(F_t|\mathcal{X}_t^F) = G(\mathcal{X}_t^F) = \int H(\mathcal{X}_t^F,\tilde{\alpha},\zeta_t)dF_{\tilde{\alpha},\zeta}, \tag{A-13}$$

and the Average Partial Effect (APE) of changing a variable, say w, in time period t from w_t to $w_t + \Delta_w$,

$$\frac{\Delta\mathrm{E}(F_t|\mathcal{X}_t^F)}{\Delta_w} = \frac{\Delta G(\mathcal{X}_t^F)}{\Delta_w} = \frac{\int\Big(H(\mathcal{X}_{t-w}^F,(w_t+\Delta_w),\tilde{\alpha},\zeta_t) - H(\mathcal{X}_t^F,\tilde{\alpha},\zeta_t)\Big)dF_{\tilde{\alpha},\zeta}}{\Delta_w}, \tag{A-14}$$

where the average is taken over the marginal distribution of the error terms $\tilde{\alpha}$ and ζ. However, the above could only be possible if the endogeneity of $\mathcal{X}_t^F$ were absent, that is, if the regressors $\mathcal{X}_t^F$ could be manipulated independently of the errors, $\tilde{\alpha}$ and ζ_t. To obtain the ASF, $G(\mathcal{X}_t^F)$, consider $\mathrm{E}(F_t|\mathcal{X}_t^F, \mathbf{X}, \mathcal{Z}) = \mathrm{E}(F_t|\mathbf{X}, \mathcal{Z})$. For a firm i, we have

$$\begin{aligned}
\mathrm{E}(F_t|\mathbf{X}, \mathcal{Z}) &= \mathrm{E}(H(\mathcal{X}_t^F, \tilde{\alpha}, \zeta_t)|\mathbf{X}, \mathcal{Z}) \\
&= \mathrm{E}(\mathrm{E}(H(\mathcal{X}_t^F, \tilde{\alpha}, \zeta_t)|\mathbf{X}, \mathcal{Z}, \tilde{\alpha})|\mathbf{X}, \mathcal{Z}) \\
&= \mathrm{E}(\mathrm{E}(H(\mathcal{X}_t^F, \tilde{\alpha}, \zeta_t)|\tilde{\alpha}, \boldsymbol{\epsilon})|\mathbf{X}, \mathcal{Z}) \\
&= \mathrm{E}(\mathrm{E}(H(\mathcal{X}_t^F, \tilde{\alpha}, \zeta_t)|\boldsymbol{\epsilon})|\mathbf{X}, \mathcal{Z}) \\
&= \mathrm{E}(\mathrm{E}(H(\mathcal{X}_t^F, \tilde{\alpha}, \zeta_t)|\boldsymbol{\epsilon}_t)|\mathbf{X}, \mathcal{Z}) \\
&= \mathrm{E}(H^*(\mathcal{X}_t^F, \tilde{\alpha}, \boldsymbol{\epsilon}_t)|\mathbf{X}, \mathcal{Z}) = \int H^*(\mathcal{X}_t, \tilde{\alpha}, \boldsymbol{\epsilon}_t) dF_{\tilde{\alpha}|\mathbf{X},\mathcal{Z}} \\
&= H^*(\mathcal{X}_t^F, \hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \mathrm{E}(F_t|\mathbf{X}, \mathcal{Z}, \hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \mathrm{E}(F_t|\mathcal{X}_t^F, \hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t). \qquad \text{(A-15)}
\end{aligned}$$

The second equality above is obtained by the Law of Iterated Expectation, the fourth follows from the fact that $\boldsymbol{\epsilon}_i = \mathbf{X}_i - \mathrm{E}(\mathbf{X}_i|\mathcal{Z}_i, \tilde{\alpha}_i)$, where $\boldsymbol{\epsilon}_i = \{\boldsymbol{\epsilon}_{i1}', \ldots, \boldsymbol{\epsilon}_{iT}'\}'$. The third follows from equation (A-3), according to which the dependence of $\tilde{\alpha}_i$ and ζ_{it} on the vector of regressors $\mathbf{X}_i$, $\mathcal{Z}_i$, and $\tilde{\alpha}_i$ is completely characterized by the reduced form error vectors $\boldsymbol{\epsilon}_i$ and $\tilde{\alpha}_i$. The fourth equality follows from $dF_{\tilde{\alpha},\zeta_t|\tilde{\alpha},\boldsymbol{\epsilon}} = dF_{\zeta_t|\tilde{\alpha},\boldsymbol{\epsilon}} = dF_{\zeta_t|\boldsymbol{\epsilon}}$[2]. The fifth equality follows from the assumption that conditional on $\boldsymbol{\epsilon}_{it}$, ζ_{it} is independent of $\boldsymbol{\epsilon}_{i-t}$.

In the fifth equality the intermediate regression function, $H^*(\mathcal{X}_t, \tilde{\alpha}, \boldsymbol{\epsilon}_t)$, is the conditional CDF of ζ_t given $\boldsymbol{\epsilon}_t$ evaluated at $\mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\tilde{\alpha}$. That is

$$H^*(\mathcal{X}_t, \tilde{\alpha}, \boldsymbol{\epsilon}_t) = F_{\zeta_t|\boldsymbol{\epsilon}_t}(\mathcal{X}_t'\boldsymbol{\varphi} + \lambda\tilde{\alpha}|\boldsymbol{\epsilon}_t).$$

Had we observed $\tilde{\alpha}$ and $\boldsymbol{\epsilon}_t$ we could have made some suitable assumption about the conditional distribution of ζ_t and obtained $H^*(\mathcal{X}_t, \tilde{\alpha}, \boldsymbol{\epsilon}_t)$, but we do not observe $\tilde{\alpha}$ and $\boldsymbol{\epsilon}_t$. We, however, have shown that

$$\mathrm{E}(\lambda\tilde{\alpha}|\mathbf{X}, \mathcal{Z}) = \lambda\hat{\tilde{\alpha}} \text{ and } \mathrm{E}(\zeta_t|\mathbf{X}, \mathcal{Z}) = \mathrm{E}(\mathrm{E}(\zeta_t|\tilde{\alpha}, \mathbf{X}, \mathcal{Z})|\mathbf{X}, \mathcal{Z}) = \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t.$$

To obtain the regression function, $H^*(\mathcal{X}_t, \hat{\tilde{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)$, the conditional CDF of ζ_t given $\mathbf{X}$ and

[2]The last equality here follows from assumption A2.

$\mathcal{Z}$, we, like Chamberlain (1984), assume that

$$\zeta_t|\mathbf{X},\mathcal{Z} \sim \mathrm{N}\left[\mathrm{E}(\zeta_t|\mathbf{X},\mathcal{Z}),\sigma^2_{\bar{\zeta}}\right] \text{ and } \lambda\tilde{\alpha}|\mathbf{X},\mathcal{Z} \sim \mathrm{N}\left[\mathrm{E}(\lambda\tilde{\alpha}|\mathbf{X},\mathcal{Z}),\sigma^2_{\alpha_\lambda}\right] \tag{A-16}$$

and that the tail, $\bar{\zeta}_t = \zeta_t - \mathrm{E}(\zeta_t|\mathbf{X},\mathcal{Z}) = \zeta_t - \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t$, and the tail, $\alpha_\lambda = \lambda\tilde{\alpha} - \mathrm{E}(\lambda\tilde{\alpha}|\mathbf{X},\mathcal{Z}) = \lambda\tilde{\alpha} - \lambda\hat{\tilde{\alpha}}$, are distributed normally with mean 0 and some variance.

Having assumed the conditional distribution of $\lambda\tilde{\alpha}$ and ζ_t, we can write the linear projection of F_t^* in error form as

$$F_t^* = \mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \alpha_\lambda + \bar{\zeta}_t, \tag{A-17}$$

With the assumptions in (A-16) and the fact that in probit models the parameters are identified only up to a scale[3], the probability of $F_t = 1$, given $\mathbf{X}$ and $\mathcal{Z}$, is given by

$$H^*(\mathcal{X}_t^F,\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t) = \int \Pr(F_t = 1|\mathcal{X}_t^F,\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t,\alpha_\lambda)dF_{\alpha_\lambda} = \Phi\Big(\{\mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t\}\tilde{\sigma}_\zeta^{-1}\Big),$$

where the $\tilde{\sigma}_\zeta$ is the variance of $\tilde{\zeta}_t = \alpha_\lambda + \bar{\zeta}_t$. Thus, we see that once we have the estimates of $\hat{\tilde{\alpha}}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$, we can simply pool the data and run a ordinary probit to get the structural estimates of the Financial constraint equation.

Having obtain $H^*(\mathcal{X}_t^F,\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t)$, the measure ASF, $G(\mathcal{X}_t^F)$, can be obtained by averaging over $\hat{\tilde{\alpha}}$ and $\hat{\boldsymbol{\epsilon}}_t$.

$$G(\mathcal{X}_t^F) = \Pr(F_t = 1|\mathcal{X}_t^F) = \int H^*(\mathcal{X}_t^F,\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t)dF_{\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t} = \int \Phi(\mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\tilde{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t)dF_{\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t} \tag{A-18}$$

To see that the above indeed gives the ASF, consider the following:

$$\begin{aligned}
&\int\left[\int \Pr(F_t = 1|\mathcal{X}_t^F,\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t,\alpha_\lambda)dF_{\alpha_\lambda}\right]dF_{\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t} = \int H^*(\mathcal{X}_t^F,\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t)dF_{\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t} \\
&= \int\left[\int H^*(\mathcal{X}_t^F,\tilde{\alpha},\boldsymbol{\epsilon}_t)dF_{\tilde{\alpha},\boldsymbol{\epsilon}|\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t}\right]dF_{\hat{\tilde{\alpha}},\hat{\boldsymbol{\epsilon}}_t} = \int H^*(\mathcal{X}_t^F,\tilde{\alpha},\boldsymbol{\epsilon}_t)dF_{\tilde{\alpha},\boldsymbol{\epsilon}_t} \\
&= \int\left[\int H(\mathcal{X}_t^F,\tilde{\alpha},\zeta_t)dF_{\tilde{\alpha},\zeta_t|\tilde{\alpha},\boldsymbol{\epsilon}_t}\right]dF_{\tilde{\alpha},\boldsymbol{\epsilon}_t} = \int H(\mathcal{X}_t^F,\tilde{\alpha},\zeta_t)dF_{\tilde{\alpha},\zeta_t} = G(\mathcal{X}_t^F).
\end{aligned}$$

[3]In the rest of the appendix, with a slight abuse of notations, we will denote the scaled parameters by their original notation.

The sample analog of ASF, $G(\mathcal{X}_{it}^F)$, for any fixed $\mathcal{X}_{it}^F = \bar{\mathcal{X}}^F$ can be computed as

$$\hat{G}(\bar{\mathcal{X}}^F) = \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \Phi(\bar{\mathcal{X}}^{F\prime}\hat{\boldsymbol{\varphi}} + \hat{\lambda}\hat{\hat{\alpha}}_i + \hat{\tilde{\Sigma}}_{\zeta\epsilon}\hat{\tilde{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\hat{\boldsymbol{\epsilon}}}_{it}), \tag{A-19}$$

where $\hat{\hat{\alpha}}_i$ and $\hat{\hat{\boldsymbol{\epsilon}}}_{it}$ are the estimated values of $\hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$.

The APE, $\frac{\Delta G(\mathcal{X}_t^F)}{\Delta_w}$ in (A-14), of changing a variable, say w_t, from w_t to $w_t + \Delta_w$ can be obtained by taking the difference of ASF at w_t and $w_t + \Delta_w$ and dividing the difference by Δ_w. In our case, since the integrand is a smooth function of its arguments, in the limit when Δ_w tends to zero we can change the order of differentiation and integration in (A-14) to get

$$\frac{\partial G(\mathcal{X}_t^F)}{\partial w} = \frac{\partial \Pr(F_t = 1|\mathcal{X}_t^F)}{\partial w} = \int \varphi_w \phi(\bar{\mathcal{X}}^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\alpha} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t) dF_{\hat{\boldsymbol{\epsilon}}_t,\hat{\alpha}}, \tag{A-20}$$

where ϕ is the density function of a standard normal and φ_w is the coefficient of w. If w is dummy variable taking values 0 and 1, then the APE of change of w_{it} from 0 to 1 on the probability of $y_{it} = 1$, given other covariates, is given by

$$\int \left[\Phi(\bar{\mathcal{X}}_{-w}^{F\prime}\boldsymbol{\varphi}_{-w} + \varphi_w + \lambda\hat{\alpha} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t) - \Phi(\bar{\mathcal{X}}_{-w}^{F\prime}\boldsymbol{\varphi}_{-w} + \lambda\hat{\boldsymbol{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t)\right] dF_{\hat{\boldsymbol{\epsilon}}_t,\hat{\boldsymbol{\alpha}}}. \tag{A-21}$$

The sample analog of the APE's in equation (A-20) and (A-21) can be computed in exactly the same way as was done for the ASF in (A-19).

A similar equation as (A-17) also holds for the Innovation and R&D equations. That is $\tilde{\Upsilon}_t$ in equations (3.12) to (3.14) in the main text is $\tilde{\Upsilon}_t = \{\tilde{\upsilon}_t, \tilde{\zeta}_t, \tilde{\eta}_{1t}, \tilde{\eta}_{0t}\}' = \{\alpha_\theta + \bar{\upsilon}_t, \alpha_\lambda + \bar{\zeta}_t, \alpha_{\mu_1} + \bar{\eta}_{1t}, \alpha_{\mu_0} + \bar{\eta}_{0t}\}'$, where $\alpha_\theta + \bar{\upsilon}_t$, $\alpha_{\mu_1} + \bar{\eta}_{1t}$, and $\alpha_{\mu_0} + \bar{\eta}_{0t}$ are defined in the same way as $\alpha_\lambda + \bar{\zeta}_t$ is defined in (A-17).

A.1. *Proof of Lemma 1*

Proof: Let Θ_1^* be true value of first stage reduced form parameters. Now, for a firm i

$$\hat{\alpha}(\mathbf{X}, \mathcal{Z}, \Theta_1) = \frac{\int \alpha \exp(-\frac{1}{2}r(\Theta_1, \alpha))\phi(\alpha)d\alpha}{\int \exp(-\frac{1}{2}r(\Theta_1, \alpha))\phi(\alpha)d\alpha},$$

where $r(\Theta_1, \alpha) = \sum_{t=1}^T (\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}_t'\bar{\boldsymbol{\delta}} + \alpha)\boldsymbol{\kappa})'\Sigma_{\epsilon\epsilon}^{-1}(\mathbf{x}_t - \mathbf{Z}_t'\boldsymbol{\delta} - (\bar{\mathcal{Z}}_t'\bar{\boldsymbol{\delta}} + \alpha)\boldsymbol{\kappa})$.

First consider the expression in the numerator $\int \alpha \exp(-\frac{1}{2}r(\Theta_1, \alpha))\phi(\alpha)d\alpha$. Now, $|\alpha|$, $|.|$ being the absolute value of its argument, is continuous in α and $|\alpha| \geq \alpha \exp(-\frac{1}{2}r(\Theta_1, \alpha))$ $\forall \Theta_1 \in \mathbf{\Theta_1}$. We also know that $\hat{\Theta}_1 \xrightarrow{a.s.} \Theta_1^*$, and since $\alpha \exp(-\frac{1}{2}r(\Theta_1, \alpha))$, is continuous in Θ_1 and α, $\alpha \exp(-\frac{1}{2}r(\hat{\Theta}_1, \alpha)) \xrightarrow{a.s.} \alpha \exp(-\frac{1}{2}r(\Theta_1^*, \alpha))$ for any given α. Thus by an application of Lebesque Dominated Convergence Theorem we can conclude that

$$\int \alpha \exp(-\frac{1}{2}r(\hat{\Theta}_1, \alpha))\phi(\alpha)d\alpha \xrightarrow{a.s.} \int \alpha \exp(-\frac{1}{2}r(\Theta_1^*, \alpha))\phi(\alpha)d\alpha.$$

Also, since $1 \geq \exp(-\frac{1}{2}r(\Theta_1, \alpha))$, again by an application of Lebesque Dominated Convergence Theorem we can conclude that

$$\int \exp(-\frac{1}{2}r(\hat{\Theta}_1, \alpha))\phi(\alpha)d\alpha \xrightarrow{a.s.} \int \exp(-\frac{1}{2}r(\Theta_1^*, \alpha))\phi(\alpha)d\alpha.$$

Given that both the numerator and the denominator in (A-9) defined at $\hat{\Theta}_1$ converge almost surly to the same defined at Θ_1^*, it can now be easily shown that

$$\hat{\hat{\alpha}}(\mathbf{X}, \mathcal{Z}, \hat{\Theta}_1) \xrightarrow{a.s.} \hat{\alpha}(\mathbf{X}, \mathcal{Z}, \Theta_1^*).$$

APPENDIX B: MAXIMUM LIKELIHOOD ESTIMATION OF THE REDUCED FORM EQUATIONS

Let N be the total number of firms. The firms are observed in at least one and at most P periods. Let N_p denote the number of firms observed in p periods, that is $N = \sum_{p=1}^{P} N_p$. Let $\mathcal{N}$ be the total number of observations, i.e., $\mathcal{N} = \sum_{p=1}^{P} N_p p$. Assume that the firms are ordered in P groups such that the N_1 firms observed once come first, the N_2 firms observed twice come second, etc. Let $M_p = \sum_{k=1}^{p} N_k$ be the cumulated number of firms observed up to p times, so that the index sets of the firms observed p times can be written as $I_{(p)} = (M_{p-1} + 1, \ldots, M_p)(p = 1, \ldots, P; M_0 = 0)$. We may, formally, consider I_1 as a cross section and $I_p(p = 2, \ldots, P)$ as a balanced panel with p observations of each firm.

The system of m reduced form equations in equation (3.4a) is given by

$$\mathbf{x}_{it} = \mathbf{Z}_{it}'\boldsymbol{\delta} + \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa} + \alpha_i\boldsymbol{\kappa} + \boldsymbol{\epsilon}_{it} = \mathbf{Z}_{it}'\boldsymbol{\delta} + \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa} + \boldsymbol{u}_{it}, \tag{B-1}$$

where $\mathbf{x}_{it} = (x_{1it}, \ldots, x_{mit})'$ and $\mathbf{Z}_{it} = \text{diag}(\mathbf{z}_{1it}, \ldots, \mathbf{z}_{mit})$ is the matrix of exogenous variables appearing in each of the m reduced form equation in (B-1). $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_m')'$,

$\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_m)'$, and $\boldsymbol{\epsilon}_{it} = (\epsilon_{1it}, \dots, \epsilon_{mit})'$. σ_α^2 is the variance of α_i, which is normally distributed with mean 0.We employ a step-wise maximum likelihood method developed by Biørn (2004) to obtain consistent estimates of parameters, $\boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}$, and σ_α^2. Given the distribution of α_i, $\boldsymbol{\kappa}\alpha_i$ is normally distributed with mean zero and variance Σ_α, given by:

$$\Sigma_\alpha = \sigma_\alpha^2 \Sigma_\kappa = \sigma_\alpha^2 \begin{pmatrix} \kappa_1^2 & & & \\ \kappa_1\kappa_2 & \kappa_2^2 & & \\ \vdots & \vdots & & \\ \kappa_1\kappa_m & \kappa_2\kappa_m & \dots & \kappa_m^2 \end{pmatrix}.$$

$\boldsymbol{\epsilon}_{it}$ is normally distributed with mean zero and variance $\Sigma_{\epsilon\epsilon}$. We assume that α_i and ϵ_{it} are mutually uncorrelated and given that $\mathbf{Z}'_{it}$ is exogenous, α_i and ϵ_{it} are uncorrelated with $\mathbf{Z}'_{it}$. Let $\boldsymbol{x}_{i(p)} = \{\boldsymbol{x}'_{i1}, \dots \boldsymbol{x}'_{ip}\}'$, $\boldsymbol{Z}_{i(p)} = \{\boldsymbol{Z}'_{i1}, \dots \boldsymbol{Z}'_{ip}\}'$ and $\boldsymbol{\epsilon}_{i(p)} = \{\boldsymbol{\epsilon}'_{i1}, \dots \boldsymbol{\epsilon}'_{ip}\}'$ and write the model as

$$\mathbf{x}_{i(p)} = \mathbf{Z}'_{i(p)}\boldsymbol{\delta} + (e_p \otimes \bar{\mathcal{Z}}'_i \bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + (e_p \otimes \alpha_i \boldsymbol{\kappa}) + \boldsymbol{\epsilon}_{i(p)} = \mathbf{Z}'_{i(p)}\boldsymbol{\delta} + (e_p \otimes \bar{\mathcal{Z}}'_i \bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + \boldsymbol{u}_{i(p)}, \tag{B-2}$$

$$\mathrm{E}(\boldsymbol{u}_{i(p)}\boldsymbol{u}'_{i(p)}) = I_p \otimes \Sigma_{\epsilon\epsilon} + E_p \otimes \Sigma_\alpha = K_p \otimes \Sigma_{\epsilon\epsilon} + J_p \otimes \Sigma_{(p)} = \Omega_{u(p)} \tag{B-3}$$

where

$$\Sigma_{(p)} = \Sigma_{\epsilon\epsilon} + p\Sigma_\alpha, \ p = 1, \dots, P \tag{B-4}$$

and I_p is the p dimensional identity matrix, e_p is the $(p \times 1)$ vector of ones, $E_p = e_p e'_p$, $J_p = (1/p)E_p$, and $K_p = I_p - J_p$. The latter two matrices are symmetric and idempotent and have orthogonal columns, which facilitates inversion of $\Omega_{u(p)}$.

B.1. *GMM estimation*

Before addressing the maximum likelihood problem, we consider the GMM problem for $\tilde{\boldsymbol{\delta}} = \{\boldsymbol{\delta}', \bar{\boldsymbol{\delta}}'\}'$ when $\boldsymbol{\kappa}$, σ_α (hence Σ_α), and $\Sigma_{\epsilon\epsilon}$ are known. Define $Q_{i(p)} = \boldsymbol{u}'_{i(p)}\Omega^{-1}_{u(p)}\boldsymbol{u}_{i(p)}$, then

GMM estimation is the problem of minimizing $Q = \sum_{p=1}^{P} \sum_{i \in I_{(p)}} Q_{i(p)}$ with respect to $\tilde{\boldsymbol{\delta}}$. Since $\Omega_{u(p)}^{-1} = K_p \otimes \Sigma_{\epsilon\epsilon}^{-1} + J_p \otimes (\Sigma_{\epsilon\epsilon} + p\Sigma_{\alpha})^{-1}$, we can rewrite Q as

$$Q = \sum_{p=1}^{P} \sum_{i \in I_{(p)}} Q_{i(p)}(\boldsymbol{\delta}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2) = \sum_{p=1}^{P} \sum_{i \in I_{(p)}} \boldsymbol{u}_{i(p)}' [K_p \otimes \Sigma_{\epsilon\epsilon}^{-1} + J_p \otimes (\Sigma_{\epsilon\epsilon} + p\Sigma_{\alpha})^{-1}] \boldsymbol{u}_{i(p)}, \tag{B-5}$$

with $\boldsymbol{u}_{i(p)} = \mathbf{x}_{i(p)} - \mathbf{Z}_{i(p)}' \boldsymbol{\delta} - (e_p \otimes \bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} \boldsymbol{\kappa})$. Had we not imposed the restriction that $\bar{\boldsymbol{\delta}}$ be the same for each of the m equations we could have estimated $\boldsymbol{\delta}$ and $\bar{\boldsymbol{\delta}}$ by employing GLS estimation as in Biørn.

B.2. *Maximum Likelihood Estimation*

We now consider ML estimation of $\Theta_1 = \{\tilde{\boldsymbol{\delta}}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2\}$. Assuming normality of α_i and the disturbances $\boldsymbol{\epsilon}_{it}$, i.e., $\alpha_i \boldsymbol{\kappa} \sim IIN(0, \sigma_{\alpha}^2 \Sigma_{\kappa})$ and $\epsilon_{it} \sim IIN(0, \Sigma_{\epsilon\epsilon})$, then $\boldsymbol{u}_{i(p)} = (e_p \otimes \alpha_i \boldsymbol{\kappa}) + \boldsymbol{\epsilon}_{i(p)} \sim IIN(0_{mp,1}, \Omega_{u(p)})$. The log-likelihood function of all $\mathbf{x}$'s conditional on all $\mathbf{Z}$'s for a firm in group p and for all firms then become, respectively,

$$\mathcal{L}_{i(p)1}(\Theta_1) = \frac{-mp}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega_{u(p)}| - \frac{1}{2} Q_{i(p)}(\tilde{\boldsymbol{\delta}}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2) \tag{B-6}$$

$$\mathcal{L}_1(\Theta_1) = \sum_{p=1}^{P} \sum_{i \in I_{(p)}} \mathcal{L}_{i(p)1} = \frac{-m\mathcal{N}}{2} \ln(2\pi) - \frac{1}{2} \sum_{p=1}^{P} N_p \ln |\Omega_{u(p)}| - \frac{1}{2} \sum_{p=1}^{P} \sum_{i \in I_{(p)}} Q_{i(p)}(\tilde{\boldsymbol{\delta}}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2), \tag{B-7}$$

where $|\Omega_{u(p)}| = |\Sigma_{(p)}||\Sigma_{\epsilon\epsilon}|^{p-1}$.

We split the problem into: (A) *Maximization of* $\mathcal{L}$ *with respect to* $\tilde{\boldsymbol{\delta}}$ *for given* $(\Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2)$ and (B) *Maximization of* $\mathcal{L}_1(\Theta_1)$ *with respect to* $(\Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2)$ *for given* $\tilde{\boldsymbol{\delta}}$. *Subproblem* (A) is identical with the GMM problem, since maximization of $\mathcal{L}_1(\Theta_1)$ with respect to $\tilde{\boldsymbol{\delta}}$ for given $(\Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2)$ is equivalent to minimization of $\sum_{p=1}^{P} \sum_{i \in I_{(p)}} Q_{i(p)}(\tilde{\boldsymbol{\delta}}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_{\alpha}^2)$.

The first order conditions with respect to $\Sigma_{\epsilon\epsilon}$, $\boldsymbol{\kappa}$, and σ_{α}^2, which we derive in Appendix E does not have a closed form solution. To obtain estimates of $\Sigma_{\epsilon\epsilon}$, $\boldsymbol{\kappa}$, and σ_{α}^2, we numerically maximize $\mathcal{L}_1(\Theta_1)$ with respect to $\Sigma_{\epsilon\epsilon}$, $\boldsymbol{\kappa}$, and σ_{α}^2 for a given $\tilde{\boldsymbol{\delta}}$ and use the first order conditions as vector of gradients in the maximization routine.

The complete stepwise algorithm for solving jointly subproblems (A) and (B) then consists in switching between minimizing (B-5) with respect to $\tilde{\boldsymbol{\delta}}$ and (B-7) with respect to $\Sigma_{\epsilon\epsilon}$, $\boldsymbol{\kappa}$, and σ^2_α and iterating until convergence. Biørn and the reference there in have monotonicity properties of such a sequential procedure which ensure that its solution converges to the ML estimator even if the likelihood function is not globally concave.

APPENDIX C: DERIVATION OF THE CORRECTION TERMS FOR THE THIRD STAGE SWITCHING REGRESSION MODEL

To avoid complicating the notations, we denote the idiosyncratic error components – $\tilde{\upsilon}$, $\tilde{\zeta}$, $\tilde{\eta}_1$, and $\tilde{\eta}_0$ – in equations (3.12) to (3.14), that were defined in appendix A, respectively as υ, ζ, η_1, and η_0. We know that the conditional expectation of η, where η is either η_1 or η_0, given ζ and υ, $\mathrm{E}[\eta|\zeta,\upsilon]$, is given by

$$\mathrm{E}[\eta|\zeta,\upsilon] = \mu_\eta + \frac{\sigma_\eta(\rho_{\eta\zeta} - \rho_{\eta\upsilon}\rho_{\zeta\upsilon})(\zeta - \mu_\zeta)}{\sigma_\zeta(1-\rho^2_{\zeta\upsilon})} + \frac{\sigma_\eta(\rho_{\eta\upsilon} - \rho_{\eta\zeta}\rho_{\zeta\upsilon})(\upsilon - \mu_\upsilon)}{\sigma_\upsilon(1-\rho^2_{\zeta\upsilon})}.$$

Since, $\mu_\eta = \mu_\zeta = \mu_\upsilon = 0$ we have,

$$\mathrm{E}[\eta|\zeta,\upsilon] = \frac{\sigma_\eta(\rho_{\eta\zeta} - \rho_{\eta\upsilon}\rho_{\zeta\upsilon})(\zeta)}{\sigma_\zeta(1-\rho^2_{\zeta\upsilon})} + \frac{\sigma_\eta(\rho_{\eta\upsilon} - \rho_{\eta\zeta}\rho_{\zeta\upsilon})(\upsilon)}{\sigma_\upsilon(1-\rho^2_{\zeta\upsilon})}.$$

Define, $\bar{\zeta} = \frac{\zeta}{\sigma_\zeta}$ and $\bar{\upsilon} = \frac{\upsilon}{\sigma_\upsilon}$, then

$$\mathrm{E}[\eta|\zeta,\upsilon] = \frac{\sigma_\eta(\rho_{\eta\zeta} - \rho_{\eta\upsilon}\rho_{\zeta\upsilon})\bar{\zeta}}{(1-\rho^2_{\zeta\upsilon})} + \frac{\sigma_\eta(\rho_{\eta\upsilon} - \rho_{\eta\zeta}\rho_{\zeta\upsilon})\bar{\upsilon}}{(1-\rho^2_{\zeta\upsilon})},$$

which can be written as

$$\mathrm{E}[\eta|\zeta,\upsilon] = \frac{\sigma_\eta\rho_{\eta\zeta}}{(1-\rho^2_{\zeta\upsilon})}(\bar{\zeta} - \rho_{\zeta\upsilon}\bar{\upsilon}) + \frac{\sigma_\eta\rho_{\eta\upsilon}}{(1-\rho^2_{\zeta\upsilon})}(\bar{\upsilon} - \rho_{\zeta\upsilon}\bar{\zeta}). \tag{C-1}$$

Hence,

$$\mathrm{E}[\eta|\zeta > -a, \upsilon > -b] = \mathrm{E}[\eta|\bar{\zeta} > \frac{-a}{\sigma_\zeta}, \bar{\upsilon} > \frac{-b}{\sigma_\upsilon}] = \frac{\int_{\frac{-b}{\sigma_\upsilon}}^\infty \int_{\frac{-a}{\sigma_\zeta}}^\infty \mathrm{E}[\eta|\bar{\zeta},\bar{\upsilon}]\phi_2(\bar{\zeta},\bar{\upsilon},\rho_{\zeta\upsilon})d\bar{\zeta}d\bar{\upsilon}}{\Phi_2\left(\frac{a}{\sigma_\zeta},\frac{b}{\sigma_\upsilon},\rho_{\zeta\upsilon}\right)}$$

$$\begin{aligned} &= \frac{1}{\Phi_2\left(\frac{a}{\sigma_\zeta},\frac{b}{\sigma_\upsilon},\rho_{\zeta\upsilon}\right)} \frac{\sigma_\eta\rho_{\eta\zeta}}{(1-\rho^2_{\zeta\upsilon})} \int_{\frac{-b}{\sigma_\upsilon}}^\infty \int_{\frac{-a}{\sigma_\zeta}}^\infty (\bar{\zeta} - \rho_{\zeta\upsilon}\bar{\upsilon})\phi_2(\bar{\zeta},\bar{\upsilon},\rho_{\zeta\upsilon})d\bar{\zeta}d\bar{\upsilon} \\ &+ \frac{1}{\Phi_2\left(\frac{a}{\sigma_\zeta},\frac{b}{\sigma_\upsilon},\rho_{\zeta\upsilon}\right)} \frac{\sigma_\eta\rho_{\eta\upsilon}}{(1-\rho^2_{\zeta\upsilon})} \int_{\frac{-b}{\sigma_\upsilon}}^\infty \int_{\frac{-a}{\sigma_\zeta}}^\infty (\bar{\upsilon} - \rho_{\zeta\upsilon}\bar{\zeta})\phi_2(\bar{\zeta},\bar{\upsilon},\rho_{\zeta\upsilon})d\bar{\zeta}d\bar{\upsilon}, \end{aligned} \tag{C-2}$$

where, ϕ_2 and Φ_2 denote respectively the density and cumulative density function function of a standard bivariate normal. Now, consider the expression $\int_{\frac{-b}{\sigma_\upsilon}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}(\bar{\zeta}-\rho_{\zeta\upsilon}\bar{\upsilon})\phi_2(\bar{\zeta},\bar{\upsilon},\rho_{\zeta\upsilon})d\bar{\zeta}d\bar{\upsilon}$, of the RHS in (C-2). Given that $\phi_2(\bar{\zeta},\bar{\upsilon},\rho_{\zeta\upsilon}) = \phi(\bar{\zeta})\frac{1}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\phi\left(\frac{\bar{\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)$, the concerned expression can be written as

$$\int_{\frac{-b}{\sigma_\upsilon}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}(\bar{\zeta}-\rho_{\zeta\upsilon}\bar{\upsilon})\phi(\bar{\zeta})\frac{1}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\phi\left(\frac{\bar{\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)d\bar{\zeta}d\bar{\upsilon} =$$

$$\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\zeta}\phi(\bar{\zeta})\left(1-\Phi\left(\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)\right)d\bar{\zeta} - \rho_{\zeta\upsilon}\int_{\frac{-b}{\sigma_\upsilon}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\upsilon}\phi(\bar{\zeta})\frac{1}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\phi\left(\frac{\bar{\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)d\bar{\zeta}d\bar{\upsilon}. \tag{C-3}$$

Now, let $y = \frac{\bar{\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}$, then $dy = \frac{d\bar{\upsilon}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}$. Having defined y, the right hand side of (C-3) can now be written as

$$\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\zeta}\phi(\bar{\zeta})\left(1-\Phi\left(\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)\right)d\bar{\zeta} - \rho_{\zeta\upsilon}\int_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}(y\sqrt{(1-\rho_{\zeta\upsilon}^2)}+\rho_{\zeta\upsilon}\bar{\zeta})\phi(\bar{\zeta})\phi(y)d\bar{\zeta}dy$$

$$= \int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\zeta}\phi(\bar{\zeta})\left(1-\Phi\left(\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)\right)d\bar{\zeta}$$

$$-\rho_{\zeta\upsilon}\int_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}y\sqrt{(1-\rho_{\zeta\upsilon}^2)}\phi(\bar{\zeta})\phi(y)d\bar{\zeta}dy - \rho_{\zeta\upsilon}^2\int_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\zeta}\phi(\bar{\zeta})\phi(y)d\bar{\zeta}dy \tag{C-4}$$

$$= (1-\rho_{\zeta\upsilon}^2)\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\zeta}\phi(\bar{\zeta})\left(1-\Phi\left(\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)\right)d\bar{\zeta}$$
$$-\rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}y\phi(\bar{\zeta})\phi(y)d\bar{\zeta}dy$$

$$= (1-\rho_{\zeta\upsilon}^2)\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\zeta}\phi(\bar{\zeta})\Phi\left(\frac{\frac{b}{\sigma_\upsilon}+\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)d\bar{\zeta} - \rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}y\phi(\bar{\zeta})\phi(y)d\bar{\zeta}dy. \tag{C-5}$$

Now, note that $\bar{\zeta}\phi(\bar{\zeta})d\bar{\zeta} = -d\phi(\bar{\zeta})$ and $\phi(\bar{\zeta}) = \phi(-\bar{\zeta})$, hence using integration by parts, the first part of the last equation of (C-5) can now be written as

$$\begin{aligned}(1-\rho_{\zeta\upsilon}^2)\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\bar{\zeta}\phi(\bar{\zeta})\Phi\left(\frac{\frac{b}{\sigma_\upsilon}+\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)d\bar{\zeta} &= (1-\rho_{\zeta\upsilon}^2)\int_{\frac{-a}{\sigma_\zeta}}^{\infty}-d\phi(\bar{\zeta})\Phi\left(\frac{\frac{b}{\sigma_\upsilon}+\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)\\
= -(1-\rho_{\zeta\upsilon}^2)\phi(\bar{\zeta})\Phi\left(\frac{\frac{b}{\sigma_\upsilon}+\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)\Bigg|_{\frac{-a}{\sigma_\zeta}}^{\infty} &+ \rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\phi(\bar{\zeta})\phi\left(\frac{\frac{b}{\sigma_\upsilon}+\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)d\bar{\zeta}\\
= (1-\rho_{\zeta\upsilon}^2)\phi(\frac{a}{\sigma_\zeta})\Phi\left(\frac{\frac{b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\frac{a}{\sigma_\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right) &+ \rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\phi(\bar{\zeta})\phi\left(\frac{\frac{b}{\sigma_\upsilon}+\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)d\bar{\zeta}.\end{aligned} \tag{C-6}$$

The second expression of the last line in equation (C-5) can be written as

$$\begin{aligned}-\rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}y\phi(\bar{\zeta})\phi(y)d\bar{\zeta}dy &= \rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\int_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}d\phi(y)\phi(\bar{\zeta})d\bar{\zeta}\\
= \rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\phi(y)\Bigg|_{\frac{\frac{-b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}}^{\infty}\phi(\bar{\zeta})d\bar{\zeta} &= -\rho_{\zeta\upsilon}\sqrt{(1-\rho_{\zeta\upsilon}^2)}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}\phi\left(\frac{\frac{b}{\sigma_\upsilon}+\rho_{\zeta\upsilon}\bar{\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)\phi(\bar{\zeta})d\bar{\zeta}.\end{aligned} \tag{C-7}$$

Plugging the results obtained in (C-6) and (C-7) into (C-4), we obtain

$$\int_{\frac{-b}{\sigma_\upsilon}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}(\bar{\zeta}-\rho_{\zeta\upsilon}\bar{\upsilon})\phi_2(\bar{\zeta},\bar{\upsilon},\rho_{\zeta\upsilon})d\bar{\zeta}d\bar{\upsilon} = (1-\rho_{\zeta\upsilon}^2)\phi(\frac{a}{\sigma_\zeta})\Phi\left(\frac{\frac{b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\frac{a}{\sigma_\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right).$$

Similarly, it can be shown that

$$\int_{\frac{-b}{\sigma_\upsilon}}^{\infty}\int_{\frac{-a}{\sigma_\zeta}}^{\infty}(\bar{\upsilon}-\rho_{\zeta\upsilon}\bar{\zeta})\phi_2(\bar{\zeta},\bar{\upsilon},\rho_{\zeta\upsilon})d\bar{\zeta}d\bar{\upsilon} = (1-\rho_{\zeta\upsilon}^2)\phi(\frac{b}{\sigma_\upsilon})\Phi\left(\frac{\frac{a}{\sigma_\zeta}-\rho_{\zeta\upsilon}\frac{b}{\sigma_\upsilon}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right).$$

Hence,

$$\mathrm{E}[\eta|\bar{\zeta} > \frac{-a}{\sigma_\zeta}, \bar{\upsilon} > \frac{-b}{\sigma_\upsilon}] = \frac{\sigma_\eta\rho_{\eta\zeta}\phi(\frac{a}{\sigma_\zeta})}{\Phi_2\left(\frac{a}{\sigma_\zeta},\frac{b}{\sigma_\upsilon},\rho_{\zeta\upsilon}\right)}\Phi\left(\frac{\frac{b}{\sigma_\upsilon}-\rho_{\zeta\upsilon}\frac{a}{\sigma_\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right) + \frac{\sigma_\eta\rho_{\eta\upsilon}\phi(\frac{b}{\sigma_\upsilon})}{\Phi_2\left(\frac{a}{\sigma_\zeta},\frac{b}{\sigma_\upsilon},\rho_{\zeta\upsilon}\right)}\Phi\left(\frac{\frac{a}{\sigma_\zeta}-\rho_{\zeta\upsilon}\frac{b}{\sigma_\upsilon}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right). \tag{C-8}$$

Now, consider

$$\mathrm{E}[\eta|\zeta \leq -a, \upsilon > -b] = \mathrm{E}[\eta|\bar{\zeta} \leq \frac{-a}{\sigma_\zeta}, \bar{\upsilon} > \frac{-b}{\sigma_\upsilon}] = \frac{\int_{\frac{-b}{\sigma_\upsilon}}^{\infty} \int_{-\infty}^{\frac{-a}{\sigma_\zeta}} \mathrm{E}[\eta|\bar{\zeta}, \bar{\upsilon}]\phi_2(\bar{\zeta}, \bar{\upsilon}, \rho_{\zeta\upsilon}) d\bar{\zeta} d\bar{\upsilon}}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_\upsilon}, -\rho_{\zeta\upsilon}\right)}$$

$$= \frac{1}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_\upsilon}, -\rho_{\zeta\upsilon}\right)} \frac{\sigma_\eta \rho_{\eta\zeta}}{(1-\rho_{\zeta\upsilon}^2)} \int_{\frac{-b}{\sigma_\upsilon}}^{\infty} \int_{-\infty}^{\frac{-a}{\sigma_\zeta}} (\bar{\zeta} - \rho_{\zeta\upsilon}\bar{\upsilon}) \phi_2(\bar{\zeta}, \bar{\upsilon}, \rho_{\zeta\upsilon}) d\bar{\zeta} d\bar{\upsilon}$$

$$+ \frac{1}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_\upsilon}, -\rho_{\zeta\upsilon}\right)} \frac{\sigma_\eta \rho_{\eta\upsilon}}{(1-\rho_{\zeta\upsilon}^2)} \int_{\frac{-b}{\sigma_\upsilon}}^{\infty} \int_{-\infty}^{\frac{-a}{\sigma_\zeta}} (\bar{\upsilon} - \rho_{\zeta\upsilon}\bar{\zeta}) \phi_2(\bar{\zeta}, \bar{\upsilon}, \rho_{\zeta\upsilon}) d\bar{\zeta} d\bar{\upsilon}. \tag{C-9}$$

By a method analogous to that used in deriving (C-8), it can be shown that

$$\mathrm{E}[\eta|\bar{\zeta} \leq \frac{-a}{\sigma_\zeta}, \bar{\upsilon} > \frac{-b}{\sigma_\upsilon}] = \frac{-\sigma_\eta \rho_{\eta\zeta} \phi(\frac{a}{\sigma_\zeta})}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_\upsilon}, -\rho_{\zeta\upsilon}\right)} \Phi\left(\frac{\frac{b}{\sigma_\upsilon} - \rho_{\zeta\upsilon}\frac{a}{\sigma_\zeta}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right)$$

$$+ \frac{\sigma_\eta \rho_{\eta\upsilon} \phi(\frac{b}{\sigma_\upsilon})}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_\upsilon}, -\rho_{\zeta\upsilon}\right)} \Phi\left(\frac{\frac{-a}{\sigma_\zeta} + \rho_{\zeta\upsilon}\frac{b}{\sigma_\upsilon}}{\sqrt{(1-\rho_{\zeta\upsilon}^2)}}\right). \tag{C-10}$$

APPENDIX D: ASYMPTOTIC COVARIANCE MATRIX OF THE SECOND AND THIRD STAGE ESTIMATES

In this section we give the asymptotic covariance matrix of the coefficients of the second stage and third stage R&D switching regression model. Newey (1984) has shown that sequential estimators can be interpreted as members of a class of Method of Moments (MM) estimators and that this interpretation facilitates derivation of asymptotic covariance matrices for multi-step estimators. Let $\Theta = \{\Theta_1', \Theta_2', \Theta_3'\}'$, where Θ_1, Θ_2, and Θ_3 are respectively the parameters to be estimated in the first, second and third step estimation of the sequential estimator. Following Newey (1984) we write the first, second, and third step estimation as an MM estimation based on the following population moment conditions:

$$E(\mathcal{L}_{i(p)1\Theta_1}) = E\frac{\partial \ln L_{i(p)1}(\Theta_1)}{\partial \Theta_1} = 0 \tag{D-1}$$

$$E(\mathcal{L}_{i(p)2\Theta_2}) = E\frac{\partial \ln L_{i(p)2}(\Theta_1, \Theta_2)}{\partial \Theta_2} = 0 \tag{D-2}$$

and

$$E(\mathcal{L}_{i(p)3\Theta_3}) = E[\sum_{t=1}^{p} I_{it}\mathbb{X}_{it}^{R}(R_{it} - \mathbb{X}_{it}^{R\prime}\Theta_3)] = 0 \tag{D-3}$$

where $L_{i(p)1}(\Theta_1)$ is the likelihood function for firm i belonging to the group p, $p \in \{1, \ldots, P\}$, for the first step system of reduced form equations. The notation p was introduced in Appendix B. p is the number of time period a firm is observed in an unbalanced panel; the minimum being 1 and maximum P. Hence $\sum_{i=1}^{N}\sum_{t=1}^{T_i} = \sum_{p=1}^{P}\sum_{i\in I_{(p)}}\sum_{t=1}^{p}$, where $I_{(p)}$ has been defined in Appendix B. $L_{i(p)2}(\Theta_1, \Theta_2)$ is the likelihood function for the second step estimation in which the joint probability of a firm being an innovator and the firm being financially constrained is estimated. Equation (D-3) is the first order condition for minimizing the sum of squared error for the pooled OLS regression of $\mathbb{X}_{it}^{R}$ on R_{it} for those firms, that have been selected, $I_{it} = 1$, where

$$\begin{aligned} R_{it} &= F_{it}R_{it} \\ \mathbb{X}_{it}^{R} &= F_{it}\{F_{it}, \mathcal{X}_{it}^{R\prime}, \hat{\hat{\alpha}}_i(\Theta_1), (\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}(\Theta_1))', C_{11it}(\Theta_1, \Theta_2), C_{12it}(\Theta_1, \Theta_2)\}' \end{aligned}$$

if $F_{it} = 1$, else

$$\begin{aligned} R_{it} &= (1 - F_{it})R_{it} \\ \mathbb{X}_{it}^{R} &= (1 - F_{it})\{F_{it}, \mathcal{X}_{it}^{R\prime}, \hat{\alpha}_i(\Theta_1), (\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}(\Theta_1))', C_{01it}(\Theta_1, \Theta_2), C_{02it}(\Theta_1, \Theta_2)\}' \end{aligned}$$

if $F_{it} = 0$.

The estimates for Θ_1, Θ_2, and Θ_3 are obtained by solving the sample analog of the above population moment conditions. The sample analog of moment conditions for the first step estimation is given by

$$\frac{1}{N}\mathcal{L}_{1\Theta_1}(\hat{\Theta}_1) = \frac{1}{N}\sum_{p=1}^{P}\sum_{i\in I_{(p)}}\frac{\partial \ln L_{i(p)1}(\hat{\Theta}_1)}{\partial \Theta_1} \tag{D-4}$$

where $\mathcal{L}_{i(p)1} = \ln L_{i(p)1}(\Theta_1)$ is given by equation (B-6) in Appendix B. $\Theta_1 = \{\boldsymbol{\delta}', \bar{\boldsymbol{\delta}}'$ $\text{vech}(\Sigma_{\epsilon\epsilon})', \boldsymbol{\kappa}', \sigma_\alpha^2\}'$ and N is the total number of firms. The first order moment conditions for solving $\hat{\Theta}_1$ are derived in Subsection D.1.

Since in the second stage we pool all data to estimate the parameters of the financial constraint and innovation equation, the sample analog of population moment condition for the second step estimation is given by

$$\frac{1}{N}\mathcal{L}_{2\Theta_2}(\hat{\Theta}_1,\hat{\Theta}_2)=\frac{1}{N}\sum_{p=1}^{P}\sum_{i\in I_{(p)}}\frac{\partial\mathcal{L}_{i(p)2}(\hat{\Theta}_1,\hat{\Theta}_2)}{\partial\Theta_2}=\frac{1}{N}\sum_{p=1}^{P}\sum_{i\in I_{(p)}}\sum_{t=1}^{p}\frac{\partial\mathcal{L}_{it2}(\hat{\Theta}_1,\hat{\Theta}_2)}{\partial\Theta_2} \tag{D-5}$$

where $\mathcal{L}_{it2}(\Theta_1,\Theta_2)$ is given by equations (3.15) and (3.17) in the main text and $\Theta_2 = \{\underline{\varphi}', \underline{\gamma}', \rho_{\tilde{\zeta}\tilde{v}}\}'$ was defined in Appendix D. Finally, the sample analog of the population for the third step estimation is given by

$$\frac{1}{N}\mathcal{L}_{3\Theta_3}(\hat{\Theta}_1,\hat{\Theta}_2,\hat{\Theta}_3)=\frac{1}{N}\sum_{p=1}^{P}\sum_{i\in I_{(p)}}\sum_{t=1}^{p}I_{it}\mathbb{X}_{it}^{R}(R_{it}-\mathbb{X}_{it}^{R\prime}\hat{\Theta}_3) \tag{D-6}$$

In Appendix A, we had shown that with $\hat{\tilde{\alpha}}_i(\mathbf{X}_i,\mathcal{Z}_i,\Theta_1)$ substituted for $\tilde{\alpha}_i$ still leads to the identification of Θ_2 and Θ_3. Let Θ_1^*, Θ_2^*, and Θ_3^* respectively be the true values of Θ_1, Θ_2 and Θ_3. Under the assumptions we make, maximizing $\mathcal{L}_{i(p)2}(\hat{\Theta}_1,\Theta_2)$ is asymptotically equivalent to maximizing $\mathcal{L}_{i(p)2}(\Theta_1^*,\Theta_2)$, where $\hat{\Theta}_1$ is a consistent first step estimate of Θ_1. Hence $\hat{\Theta}_2$ obtained by solving $\frac{1}{N}\mathcal{L}_{2\Theta_2}(\hat{\Theta}_1,\hat{\Theta}_2)=0$ is a consistent estimate of Θ_2. By the same logic $\hat{\Theta}_3$ obtained by solving $\frac{1}{N}\mathcal{L}_{3\Theta_3}(\hat{\Theta}_1,\hat{\Theta}_2,\hat{\Theta}_3)=0$ in the third step gives consistent estimate of the third stage parameters. Newey gives a general formulation of the asymptotic distribution of the subsequent step estimates for a sequential step sequential estimator.

To derive the asymptotic distribution of the second and third step estimates $\hat{\Theta}_2$ and $\hat{\Theta}_3$ respectively, consider the stacked up sample moment conditions

$$\frac{1}{N}\begin{bmatrix}\mathcal{L}_{1\Theta_1}(\hat{\Theta}_1)\\ \mathcal{L}_{2\Theta_2}(\hat{\Theta}_1,\hat{\Theta}_2)\\ \mathcal{L}_{3\Theta_3}(\hat{\Theta}_1,\hat{\Theta}_2,\hat{\Theta}_3)\end{bmatrix}=0. \tag{D-7}$$

A series of Taylor's expansion of $\mathcal{L}_{1\Theta_1}(\hat{\Theta}_1)$, $\mathcal{L}_{2\Theta_2}(\hat{\Theta}_1,\hat{\Theta}_2)$ and $\mathcal{L}_{\Theta_3}(\hat{\Theta}_1,\hat{\Theta}_2,\hat{\Theta}_3)$ around Θ^* gives

$$\frac{1}{N}\begin{bmatrix}\mathcal{L}_{1\Theta_1\Theta_1} & 0 & 0\\ \mathcal{L}_{2\Theta_2\Theta_1} & \mathcal{L}_{2\Theta_2\Theta_2} & 0\\ \mathcal{L}_{3\Theta_3\Theta_1} & \mathcal{L}_{3\Theta_3\Theta_2} & \mathcal{L}_{3\Theta_3\Theta_3}\end{bmatrix}\begin{bmatrix}\sqrt{N}(\hat{\Theta}_1-\Theta_1^*)\\ \sqrt{N}(\hat{\Theta}_2-\Theta_2^*)\\ \sqrt{N}(\hat{\Theta}_3-\Theta_3^*)\end{bmatrix}=-\frac{1}{\sqrt{N}}\begin{bmatrix}L_{1\Theta_1}\\ \mathcal{L}_{2\Theta_2}\\ \mathcal{L}_{3\Theta_3}\end{bmatrix}$$

In matrix notation the above can be written as

$$B_{\Theta\Theta_N}\sqrt{N}(\hat{\Theta}-\Theta) = -\frac{1}{\sqrt{N}}\Lambda_{\Theta_N}, \tag{D-8}$$

where Λ_{Θ_N} is evaluated at Θ^* and $B_{\Theta\Theta_N}$ is evaluated at points somewhere between $\hat{\Theta}$ and Θ^*. Under the standard regularity conditions for Generalized Method of Moments (GMM), (see Newey, 1984), $B_{\Theta\Theta_N}$ converges in probability to the lower block triangular matrix $B_* = \lim EB_{\Theta\Theta_N}$. B_* is given by

$$B_* = \begin{bmatrix} \mathbb{L}_{1\Theta_1\Theta_1} & 0 & 0 \\ \mathbb{L}_{2\Theta_2\Theta_1} & \mathbb{L}_{2\Theta_2\Theta_2} & 0 \\ \mathbb{L}_{3\Theta_3\Theta_1} & \mathbb{L}_{3\Theta_3\Theta_2} & \mathbb{L}_{3\Theta_3\Theta_3} \end{bmatrix},$$

where a typical element, say, $\mathbb{L}_{2\Theta_2\Theta_1} = \mathrm{E}(\mathcal{L}_{i(p)2\Theta_2\Theta_1})$. $\frac{1}{\sqrt{N}}\Lambda_N$ in (D-8) converges in distribution to an asymptotically normal random variable with mean zero and a covariance matrix $A_* = \lim E\frac{1}{N}\Lambda_N\Lambda_N'$, where A_* is given by

$$A_* = \begin{bmatrix} V_{L_1L_1} & V_{L_1L_2} & V_{L_1L_3} \\ V_{L_2L_1} & V_{L_2L_2} & V_{L_2L_3} \\ V_{L_3L_1} & V_{L_3L_2} & V_{L_3L_3} \end{bmatrix},$$

where a typical element of A_*, say $V_{L_1L_2}$ is given by $V_{L_1L_2} = E[\mathcal{L}_{i(p)1\Theta_1}(\Theta_1)\mathcal{L}_{i(p)2\Theta_2}(\Theta_1,\Theta_2)']$. Under the regularity conditions $\sqrt{N}(\hat{\Theta}-\Theta^*)$ is asymptotically normal with zero mean and covariance matrix[4] given by $B_*^{-1}A_*B_*^{-1\prime}$.

$$\sqrt{N}(\hat{\Theta}-\Theta^*) \overset{a}{\sim} \mathrm{N}[(0),(B_*^{-1}A_*B_*^{-1\prime})] \tag{D-9}$$

The moment conditions for every firm, at the estimates of Θ_1, Θ_2, and Θ_3, of the three stages can be employed to obtain the sample analog of every element in A_*. For example, to get an estimate of $V_{L_1L_2}$ we have to estimate $\frac{1}{N}\sum_{p=1}^{P}\sum_{i\in I_{(p)}}[\mathcal{L}_{i(p)1\Theta_1}(\hat{\Theta}_1)\mathcal{L}_{i(p)2\Theta_2}(\hat{\Theta}_1,\hat{\Theta}_2)']$.

Consider now the elements of B_*. Since in the first and the second stage we employ

[4] The covariance matrices V_{2F}^* in equation (E-5), V_2^* in equation (E-7), and V_s^* in equation (E-12) can obtained by selecting the appropriate submatrix of $\frac{1}{N}B_*^{-1}A_*B_*^{-1\prime}$.

MLE, at Θ_1^* and Θ_2^* to which $\hat{\Theta}_1$ and $\hat{\Theta}_2$ converge, we have

$$\mathbb{L}_{1\Theta_1\Theta_1} = \mathrm{E}\left[\frac{\partial \mathcal{L}_{i(p)1}(\Theta_1)}{\partial \Theta_1 \Theta_1'}\right] = -\mathrm{E}\left[\frac{\partial \mathcal{L}_{i(p)1}(\Theta_1)}{\partial \Theta_1}\frac{\partial \mathcal{L}_{i(p)1}(\Theta_1)}{\partial \Theta_1'}\right] = -V_{L_1L_1} \text{ and}$$
$$\mathbb{L}_{2\Theta_2\Theta_2} = \mathrm{E}\left[\frac{\partial \mathcal{L}_{i(p)2}(\Theta_2)}{\partial \Theta_2 \Theta_2'}\right] = -\mathrm{E}\left[\frac{\partial \mathcal{L}_{i(p)2}(\Theta_2)}{\partial \Theta_2}\frac{\partial \mathcal{L}_{i(p)2}(\Theta_2)}{\partial \Theta_1'}\right] = -V_{L_2L_2}.$$

We can employ the derivative $\mathcal{L}_{i(p)1}(\Theta_1)$ of with respect to Θ_1 and of $\mathcal{L}_{i(p)2}(\Theta_1, \Theta_2)$ with respect to Θ_2 to compute $\mathcal{L}_{i(p)1\Theta_1\Theta_1}$ and $\mathcal{L}_{i(p)2\Theta_2\Theta_2}$ for all firms, which can then be used to compute the sample analog of $\mathbb{L}_{1\Theta_1\Theta_1}$ and $\mathbb{L}_{2\Theta_2\Theta_2}$. This leaves us with the problem of constructing sample analogs of $\mathbb{L}_{2\Theta_2\Theta_1}$, $\mathbb{L}_{3\Theta_3\Theta_1}$, $\mathbb{L}_{3\Theta_3\Theta_2}$, and $\mathbb{L}_{3\Theta_3\Theta_3}$. While it is straightforward to compute sample analog of $\mathbb{L}_{3\Theta_3\Theta_3}$, computation of sample analogs of $\mathbb{L}_{2\Theta_2\Theta_1}$, $\mathbb{L}_{3\Theta_3\Theta_1}$, and $\mathbb{L}_{3\Theta_3\Theta_2}$ can be challenging. In the next subsections we derive the derivative of $\mathcal{L}_{i(p)2\Theta_2}(\Theta_1, \Theta_2)$ and $\mathcal{L}_{i(p)3\Theta_3}(\Theta_1, \Theta_2, \Theta_2)$ with respect to Θ_1 and the derivative of $\mathcal{L}_{i(p)3\Theta_3}(\Theta_1, \Theta_2, \Theta_2)$ with respect to Θ_2. But first we begin by deriving the first order conditions of the log likelihood function of the first stage.

D.1. *Derivation of the First Order Conditions for First Stage Reduced Form Likelihood Function*

To derive the first order conditions it is convenient to arrange the disturbances, $\boldsymbol{u}_{it}$, given in (B-1), for a firm i, $i \in I_p$, in the $(m \times p)$ matrix $\tilde{E}_{i(p)} = [\boldsymbol{u}_{i1}, \ldots, \boldsymbol{u}_{ip}]$, write $\boldsymbol{u}_{i(p)} = \mathrm{vec}(E_{i(p)})$, where 'vec()' is the vectorization operator, and define

$$W_{ui(p)} = \tilde{E}_{i(p)} K_p \tilde{E}_{i(p)}' \text{ and } B_{ui(p)} = \tilde{E}_{i(p)} J_p \tilde{E}_{i(p)}', \tag{D-10}$$

where J_p and K_p defined earlier in Appendix B are $J_p = (1/p)E_p$, and $K_p = I_p - J_p$, where I_p is the p dimensional identity matrix, e_p is the $(p \times 1)$ vector of ones, $E_p = e_p e_p'$.

Below we show that

$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \boldsymbol{\delta}} = 2\boldsymbol{Z}_{i(p)}\Omega_{u(p)}^{-1}\boldsymbol{u}_{i(p)},$$
$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \bar{\boldsymbol{\delta}}} = -2\bar{\mathcal{Z}}_i\boldsymbol{\kappa}'\left[\Sigma_{(p)}^{-1}\tilde{E}_{i(p)}J_p + \Sigma_{\epsilon\epsilon}^{-1}\tilde{E}_{i(p)}K_p\right]e_p,$$

$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \mathrm{vech}(\Sigma_{\epsilon\epsilon})} = -\frac{1}{2}\mathrm{vech}\left(\Sigma_{(p)}^{-1} + (p-1)\Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}W_{ui(p)}\Sigma_{\epsilon\epsilon}^{-1}\right),$$

$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \boldsymbol{\kappa}} = -p\sigma_\alpha^2[\Sigma_{(p)}^{-1} - \Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1}]\boldsymbol{\kappa} + \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\left[\Sigma_{(p)}^{-1}\tilde{E}_{i(p)}J_p + \Sigma_{\epsilon\epsilon}^{-1}\tilde{E}_{i(p)}K_p\right]e_p,$$

and

$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \sigma_\alpha^2} = -\frac{1}{2}p[\text{vec}(\Sigma_{(p)}^{-1})' - \text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})']\text{vec}(\Sigma_\kappa), \tag{D-11}$$

where 'vech()' operator is column-wise vectorization of the lower triangle of the symmetric matrix $\Sigma_{\epsilon\epsilon}$[5].

To derive the above we utilize the following matrix results:

1. $|J_p \otimes C + K_p \otimes D| = |C||D|^{p-1}$, since J_p and K_p have ranks 1 and $p-1$,
2. $\frac{\partial \ln |A|}{\partial A} = (A')^{-1}$,
3. $\text{tr}(ABCD) = \text{tr}(CDAB) = \text{vec}(A')'(D' \otimes B)\text{vec}(C) = \text{vec}(C')'(B' \otimes D)\text{vec}(A)$,
4. $\frac{\partial \text{tr}(CB^{-1})}{\partial B} = -(B^{-1}CB^{-1})'$,
5. $\frac{\partial xx'}{\partial x} = x \otimes I_n + I_n \otimes x$, where x is a $(n \times 1)$ matrix and I_n is a n dimensional identity matrix

   and
6. $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$.

The log-likelihood for a firm i belonging to group p is given by

$$\mathcal{L}_{i(p)} = \frac{-mp}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{u(p)}| - \frac{1}{2}Q_{i(p)}(\boldsymbol{\delta}, \bar{\boldsymbol{\delta}}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_\alpha^2).$$

Then

$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \Sigma_{\epsilon\epsilon}} = -\frac{1}{2}\frac{\partial \ln|\Omega_{u(p)}|}{\partial \Sigma_{\epsilon\epsilon}} - \frac{1}{2}\frac{\partial Q_{i(p)}(\boldsymbol{\delta}, \bar{\boldsymbol{\delta}}, \Sigma_{\epsilon\epsilon}, \boldsymbol{\kappa}, \sigma_\alpha^2)}{\partial \Sigma_{\epsilon\epsilon}}.$$

Now from (a) we have $|\Omega_{u(p)}| = |K_p \otimes \Sigma_{\epsilon\epsilon} + J_p \otimes \Sigma_{(p)}| = |\Sigma_{\epsilon\epsilon}|^{p-1}|\Sigma_{(p)}|$ and from (b) we have

$$\frac{\partial \ln|\Omega_{u(p)}|}{\partial \Sigma_{\epsilon\epsilon}} = \frac{\partial \ln|\Sigma_{(p)}|}{\partial \Sigma_{\epsilon\epsilon}} + (p-1)\frac{\partial \ln|\Sigma_{\epsilon\epsilon}|}{\partial \Sigma_{\epsilon\epsilon}} = \Sigma_{(p)}^{-1} + (p-1)\Sigma_{\epsilon\epsilon}^{-1} \tag{D-12}$$

For any given $\boldsymbol{\delta}$ and $\bar{\boldsymbol{\delta}}$ we have

$$\begin{aligned} Q_{i(p)}() &= \boldsymbol{u}_{i(p)}'[K_p \otimes \Sigma_{\epsilon\epsilon}^{-1}]\boldsymbol{u}_{i(p)} + \boldsymbol{u}_{i(p)}'[J_p \otimes \Sigma_{(p)}^{-1}]\boldsymbol{u}_{i(p)} \\ &= \text{vec}(E_{i(p)})'[K_p \otimes \Sigma_{\epsilon\epsilon}^{-1}]\text{vec}(E_{i(p)}) + \text{vec}(E_{i(p)})'[J_p \otimes \Sigma_{(p)}^{-1}]\text{vec}(E_{i(p)}) \end{aligned}$$

[5]Because $\Sigma_{\epsilon\epsilon}$ is symmetric we only need to optimize with respect to $\frac{m(m+1)}{2}$ elements of the lower triangle of the $\Sigma_{\epsilon\epsilon}$.

From (c) we get

$$Q_{i(p)}() = \mathrm{tr}[E_{i(p)}\Sigma_{(p)}^{-1}E'_{i(p)}J_p] + \mathrm{tr}[E_{i(p)}\Sigma_{\epsilon\epsilon}^{-1}E'_{i(p)}K_p] = \mathrm{tr}[E_{i(p)}J_pE'_{i(p)}\Sigma_{(p)}^{-1}] + \mathrm{tr}[E_{i(p)}K_pE'_{i(p)}\Sigma_{\epsilon\epsilon}^{-1}].$$

Using (D-10) we obtain

$$Q_{i(p)}() = \mathrm{tr}[B_{ui(p)}\Sigma_{(p)}^{-1}] + \mathrm{tr}[W_{ui(p)}\Sigma_{\epsilon\epsilon}^{-1}],$$

and from (d) we get

$$\frac{\partial Q_{i(p)}()}{\partial \Sigma_{\epsilon\epsilon}} = -[\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1} + \Sigma_{\epsilon\epsilon}^{-1}W_{ui(p)}\Sigma_{\epsilon\epsilon}^{-1}]. \tag{D-13}$$

Combining (D-12) and (D-13) we obtain

$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \mathrm{vech}(\Sigma_{\epsilon\epsilon})} = -\frac{1}{2}\mathrm{vech}\left(\Sigma_{(p)}^{-1} + (p-1)\Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}W_{ui(p)}\Sigma_{\epsilon\epsilon}^{-1}\right). \tag{D-14}$$

To find expressions for the first order condition with respect to $\boldsymbol{\kappa}$ and σ_α^2, consider the total differential $\mathrm{d}(\ln|\Omega_{u(p)}|)$ and $\mathrm{d}(Q_{i(p)}())$ for given $\Sigma_{\epsilon\epsilon}$, $\boldsymbol{\delta}$, and $\bar{\boldsymbol{\delta}}$.

$$\begin{aligned}\mathrm{d}(\ln|\Omega_{u(p)}|) &= \mathrm{d}(\ln(|\Sigma_{\epsilon\epsilon}|^{p-1}|\Sigma_{(p)}|)) = \mathrm{d}(\ln(|\Sigma_{(p)}|)) = \mathrm{vec}[\Sigma_{(p)}^{-1}]'\mathrm{vec}[\mathrm{d}(\Sigma_{(p)})] \\ &= \mathrm{vec}[\Sigma_{(p)}^{-1}]'\mathrm{vec}[p\mathrm{d}(\sigma_\alpha^2)\Sigma_\kappa + p\sigma_\alpha^2\mathrm{d}(\Sigma_\kappa)] \\ &= \mathrm{vec}[\Sigma_{(p)}^{-1}]'\mathrm{vec}[p\mathrm{d}(\sigma_\alpha^2)\Sigma_\kappa + p\sigma_\alpha^2(\boldsymbol{\kappa}\otimes I_m + I_m\otimes\boldsymbol{\kappa})\mathrm{d}(\boldsymbol{\kappa})],\end{aligned} \tag{D-15}$$

where the third equality follows from employing (b). Since $\Sigma_{\epsilon\epsilon}$ is given, $\mathrm{d}\Sigma_{(p)} = d(\Sigma_{\epsilon\epsilon} + p\sigma_\alpha^2\Sigma_\kappa) = p\mathrm{d}(\sigma_\alpha^2\Sigma_\kappa) = p\mathrm{d}(\sigma_\alpha^2)\Sigma_\kappa + p\sigma_\alpha^2\mathrm{d}(\Sigma_\kappa)$, hence the fourth equality. Also, since $\Sigma_\kappa = \boldsymbol{\kappa}\boldsymbol{\kappa}'$, the last equality follows using (e).

The total differential, $\mathrm{d}(Q_{i(p)}())$, is given by

$$
\begin{aligned}
\mathrm{d}(Q_{i(p)}()) =& \mathrm{d}(\mathrm{tr}[B_{ui(p)}\Sigma_{(p)}^{-1}] + \mathrm{tr}[W_{ui(p)}\Sigma_{\epsilon\epsilon}^{-1}]) \\
=& -\mathrm{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})'\mathrm{vec}(p\mathrm{d}(\sigma_\alpha^2)\Sigma_\kappa + p\sigma_\alpha^2\mathrm{d}(\Sigma_\kappa)) \\
& + \mathrm{vec}(\Sigma_{(p)}^{-1})'\mathrm{vec}(\mathrm{d}(B_{ui(p)})) + \mathrm{vec}(\Sigma_{\epsilon\epsilon}^{-1})'\mathrm{vec}(\mathrm{d}(W_{ui(p)})) \\
=& -\mathrm{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})'\mathrm{vec}\Big(p\mathrm{d}(\sigma_\alpha^2)\Sigma_\kappa + p\sigma_\alpha^2(\boldsymbol{\kappa}\otimes I_m + I_m\otimes\boldsymbol{\kappa})\mathrm{d}\boldsymbol{\kappa}\Big) \\
& + \mathrm{vec}(\Sigma_{(p)}^{-1})'\Big[(I_m\otimes\tilde{E}_{i(p)}J_p)\mathrm{vec}(\mathrm{d}(\tilde{E}'_{i(p)})) + (\tilde{E}_{i(p)}J'_p\otimes I_m)\mathrm{vec}(\mathrm{d}(\tilde{E}_{i(p)}))\Big] \\
& + \mathrm{vec}(\Sigma_{\epsilon\epsilon}^{-1})'\Big[(I_m\otimes\tilde{E}_{i(p)}K_p)\mathrm{vec}(\mathrm{d}(\tilde{E}'_{i(p)})) + (\tilde{E}_{i(p)}K'_p\otimes I_m)\mathrm{vec}(\mathrm{d}(\tilde{E}_{i(p)}))\Big] \\
=& -\mathrm{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})'\mathrm{vec}\Big(p\mathrm{d}(\sigma_\alpha^2)\Sigma_\kappa + p\sigma_\alpha^2(\boldsymbol{\kappa}\otimes I_m + I_m\otimes\boldsymbol{\kappa})\mathrm{d}\boldsymbol{\kappa}\Big) \\
& - \mathrm{vec}(\Sigma_{(p)}^{-1})'\Big[(I_m\otimes\tilde{E}_{i(p)}J_p)(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa}\otimes e_p) + (\tilde{E}_{i(p)}J'_p\otimes I_m)(e_p\otimes\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa})\Big] \\
& - \mathrm{vec}(\Sigma_{\epsilon\epsilon}^{-1})'\Big[(I_m\otimes\tilde{E}_{i(p)}K_p)(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa}\otimes e_p) + (\tilde{E}_{i(p)}K'_p\otimes I_m)(e_p\otimes\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa})\Big] \\
=& -p\mathrm{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})'\mathrm{vec}(\Sigma_\kappa)\mathrm{d}(\sigma_\alpha^2) - p\sigma_\alpha^2\mathrm{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})'(\boldsymbol{\kappa}\otimes I_m + I_m\otimes\boldsymbol{\kappa})\mathrm{d}\boldsymbol{\kappa} \\
& - \Big[\mathrm{vec}(J'_p\tilde{E}'_{i(p)}\Sigma_{(p)}^{-1})'(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa}\otimes e_p) + \mathrm{vec}(\Sigma_{(p)}^{-1}\tilde{E}_{i(p)}J'_p)'(e_p\otimes\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa})\Big] \\
& - \Big[\mathrm{vec}(K'_p\tilde{E}'_{i(p)}\Sigma_{\epsilon\epsilon}^{-1})'(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa}\otimes e_p) + \mathrm{vec}(\Sigma_{\epsilon\epsilon}^{-1}\tilde{E}_{i(p)}K'_p)'(e_p\otimes\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa})\Big] \\
=& -p\mathrm{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})'\mathrm{vec}(\Sigma_\kappa)\mathrm{d}(\sigma_\alpha^2) - 2p\sigma_\alpha^2\mathrm{vec}([\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1}]\boldsymbol{\kappa})'\mathrm{d}\boldsymbol{\kappa} \\
& - 2e'_p\Big[J'_p\tilde{E}'_{i(p)}\Sigma_{(p)}^{-1} + K'_p\tilde{E}'_{i(p)}\Sigma_{\epsilon\epsilon}^{-1}\Big]\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\mathrm{d}\boldsymbol{\kappa},
\end{aligned} \tag{D-16}
$$

where the second equality follows from employing (d), and the fact that $\Sigma_{\epsilon\epsilon}$ being given, $\mathrm{d}\Sigma_{(p)} = d(\Sigma_{\epsilon\epsilon} + p\sigma_\alpha^2\Sigma_\kappa) = p\mathrm{d}(\sigma_\alpha^2\Sigma_\kappa)$. Since $\Sigma_\kappa = \boldsymbol{\kappa}\boldsymbol{\kappa}'$, the third equality follows using (e) and taking the differential of $B_{ui(p)}$ and $W_{ui(p)}$. The fourth equality follows from taking the differential of $\tilde{E}_{i(p)}$, and finally in the fifth and sixth we have used matrix algebra to rearrange terms. Given (D-16) we can conclude that

$$
\frac{\partial Q_{i(p)}()}{\partial\boldsymbol{\kappa}} = -2p\sigma_\alpha^2[\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1}]\boldsymbol{\kappa} - 2\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\Big[\Sigma_{(p)}^{-1}\tilde{E}_{i(p)}J_p + \Sigma_{\epsilon\epsilon}^{-1}\tilde{E}_{i(p)}K_p\Big]e_p. \tag{D-17}
$$

Similarly, from (D-15) we obtain

$$
\frac{\partial\ln|\Omega_{u(p)}|}{\partial\boldsymbol{\kappa}} = 2p\sigma_\alpha^2[\Sigma_{(p)}^{-1}]\boldsymbol{\kappa}. \tag{D-18}
$$

Combining (D-17) and (D-18) we get the expression in (D-11) for $\frac{\partial \mathcal{L}_{i(p)}}{\partial \boldsymbol{\kappa}}$.

Again, from (D-15) and (D-16) respectively we obtain $\frac{\partial \ln|\Omega_{u(p)}|}{\partial \sigma_{\alpha}^2} = p\text{vec}[\Sigma_{(p)}^{-1}]'\text{vec}(\Sigma_{\kappa})$ and $\frac{\partial Q_{i(p)}()}{\partial \sigma_{\alpha}^2} = -p\text{vec}[\Sigma_{(p)}^{-1} B_{ui(p)} \Sigma_{(p)}^{-1}]'\text{vec}(\Sigma_{\kappa})$, which when combined yields the expression for $\frac{\partial \mathcal{L}_{i(p)}}{\partial \sigma_{\alpha}^2}$ in (D-11). By a similar derivation as in (D-16), we can conclude that

$$\frac{\partial \mathcal{L}_{i(p)}}{\partial \bar{\boldsymbol{\delta}}} = -2\bar{\mathcal{Z}}_i \boldsymbol{\kappa}' \left[\Sigma_{(p)}^{-1} \tilde{E}_{i(p)} J_p + \Sigma_{\epsilon\epsilon}^{-1} \tilde{E}_{i(p)} K_p \right] e_p. \tag{D-19}$$

D.2. *Derivative of* $\mathcal{L}_{i(p)2\Theta_2}$ *with respect to* Θ_1

Let us begin by deriving the derivative of score functions, $\mathcal{L}_{i(p)2\Theta_2}$, of second stage likelihood with respect to Θ_1. Since the second step is essentially a combination of probit and bivariate probit, we have to take the derivative of the score functions of the probit and bivariate probit with respect to Θ_1. Now, we know that Θ_1 enters the second stage of the sequential estimator through $\bar{\mathbf{z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i(\Theta_1)$ and $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}(\Theta_1)$, and that $\mathcal{L}_{i(p)2\Theta_2} = \sum_{t=1}^{p} \mathcal{L}_{it2\Theta_2}$. Hence in order to compute the derivative of $\mathcal{L}_{i(p)2\Theta_2}$ with respect to Θ_1 we have to compute $\frac{\partial \mathcal{L}_{it2\Theta_2}(\Theta_1,\Theta_2)}{\partial \Theta_1'}$. To do so let us first separate the coefficients of the second stage into coefficients of the Financial Constraint equation, Θ_{2F}, coefficients of the Innovation equation Θ_{2I} and $\rho_{\tilde{\zeta}\tilde{v}}$, the correlation between the idiosyncratic components of the Financial Constraint and the Innovation equation. In matrix form we can write

$$\mathcal{L}_{i(p)2\Theta_2\Theta_1} = \frac{\partial \mathcal{L}_{i(p)2\Theta_2}}{\partial \Theta_1'} = \sum_{t=1}^{p} \frac{\partial \mathcal{L}_{it2\Theta_2}}{\partial \Theta_1'} = \sum_{t=1}^{p} \begin{bmatrix} \frac{\partial \mathcal{L}_{it2\Theta_{2F}}}{\partial \Theta_1'} \\ \frac{\partial \mathcal{L}_{it2\Theta_{2I}}}{\partial \Theta_1'} \\ \frac{\partial \mathcal{L}_{it2\rho_{\tilde{\zeta}\tilde{v}}}}{\partial \Theta_1'} \end{bmatrix},$$

where the score functions, $\mathcal{L}_{it2\Theta_{2F}}$, $\mathcal{L}_{it2\Theta_{2I}}$, and $\mathcal{L}_{it2\Theta_{2\rho_{\tilde{\zeta}\tilde{v}}}}$, above are the score functions of the log likelihood function for bivariate probit when *it* belongs to CIS3 and CIS3.5, and are given by

$$\mathcal{L}_{it2\Theta_{2F}}(\Theta_1,\Theta_2) = \frac{q_{itF}g_{itF}}{\Phi_2}\mathbb{X}_{it}^F, \quad \mathcal{L}_{it2\Theta_{2I}}(\Theta_1,\Theta_2) = \frac{q_{itI}g_{itI}}{\Phi_2}\mathbb{X}_{it}^I, \text{ and } \mathcal{L}_{it2\rho_{\tilde{\zeta}\tilde{v}}}(\Theta_1,\Theta_2) = \frac{q_{itI}q_{itI}\phi_2}{\Phi_2} \tag{D-20}$$

where $\mathbb{X}^F_{it} = \{\mathcal{X}^{F\prime}_{it}, \bar{\mathcal{Z}}^{\prime}_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i, (\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it})^{\prime}\}^{\prime}$, $\mathbb{X}^I_{it} = \{\mathcal{X}^{I\prime}_{it}, \bar{\mathcal{Z}}^{\prime}_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i, (\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it})^{\prime}\}^{\prime}$, $q_{itF} = 2F_{it} - 1$, $q_{itI} = 2I_{it} - 1$. g_{itF} and g_{itI} in (D-20) are defined as

$$g_{itF} = \phi(\varphi_{it})\Phi(\frac{\gamma_{it} - \rho^*_{\tilde{\zeta}\tilde{v}}\varphi_{it}}{\sqrt{(1-\rho^{*2}_{\tilde{\zeta}\tilde{v}})}}), \text{ and } g_{itI} = \phi(\gamma_{it})\Phi(\frac{\varphi_{it} - \rho^*_{\tilde{\zeta}\tilde{v}}\gamma_{it}}{\sqrt{(1-\rho^{*2}_{\tilde{\zeta}\tilde{v}})}}),$$

where $\rho^*_{\tilde{\zeta}\tilde{v}} = q_{itF}q_{itI}\rho_{\tilde{\zeta}\tilde{v}}$, $\varphi_{it} = \mathbb{X}^{F\prime}_{it}\Theta_{2F}$, and $\gamma_{it} = \mathbb{X}^{I\prime}_{it}\Theta_{2I}$. However, for CIS2.5 we do not observe F_{it} when $I_{it} = 0$. So, while the score functions remain the same as in (D-20) when $I_{it} = 1$, the functions are

$$\mathcal{L}_{it2\Theta_{2F}}(\Theta_1, \Theta_2) = \mathbf{0}_{\Theta_{2F}}, \quad \mathcal{L}_{it2\Theta_{2I}}(\Theta_1, \Theta_2) = -\frac{\phi(-\mathbb{X}^{I\prime}_{it}\Theta_{2I})}{\Phi(-\mathbb{X}^{I\prime}_{it}\Theta_{2I})}\mathbb{X}^I_{it}, \text{ and } \mathcal{L}_{it2\rho_{\tilde{\zeta}\tilde{v}}}(\Theta_1, \Theta_2) = 0 \tag{D-21}$$

when $I_{it} = 0$, where $\mathbf{0}_{\Theta_{2F}}$ is a vector of zeros.

To ease notations we now suppress firm and time subscript except when necessary. Given the above, we have

$$\begin{aligned}
\frac{\partial \mathcal{L}_{2\Theta_{2j}}(\Theta_1, \Theta_2)}{\partial \Theta_1^{\prime}} &= q_j\left\{\frac{\partial}{\partial \Theta_1^{\prime}}\left(\frac{g_j}{\Phi_2}\right)\mathbb{X}^j + \frac{g_j}{\Phi_2}\frac{\partial \mathbb{X}^j}{\partial \Theta_1^{\prime}}\right\} \\
&= q_j\left\{\left(\frac{\partial(g_j/\Phi_2)}{\partial \varphi_{it}}\frac{\partial \varphi_{it}}{\partial \Theta_1^{\prime}} + \frac{\partial(g_j/\Phi_2)}{\partial \gamma_{it}}\frac{\partial \gamma_{it}}{\partial \Theta_1^{\prime}}\right)\mathbb{X}^j + \frac{g_j}{\Phi_2}\frac{\partial \mathbb{X}^j}{\partial \Theta_1^{\prime}}\right\} \\
&= q_j\left\{\left(\frac{\partial(g_j/\Phi_2)}{\partial \varphi_{it}}\frac{\partial \mathbb{X}^{F\prime}_{it}}{\partial \Theta_1^{\prime}}\Theta_{2F} + \frac{\partial(g_j/\Phi_2)}{\partial \gamma_{it}}\frac{\partial \mathbb{X}^{I\prime}_{it}}{\partial \Theta_1^{\prime}}\Theta_{2I}\right)\mathbb{X}^j + \frac{g_j}{\Phi_2}\frac{\partial \mathbb{X}^j}{\partial \Theta_1^{\prime}}\right\}.
\end{aligned} \tag{D-22}$$

where $j \in \{F, I\}$ and

$$\frac{\partial \mathcal{L}_{2\rho_{\tilde{\zeta}\tilde{v}}}(\Theta_1, \Theta_2)}{\partial \Theta_1^{\prime}} = q_F q_I\left\{\frac{\partial}{\partial \Theta_1^{\prime}}\left(\frac{\phi_2}{\Phi_2}\right)\right\} = q_F q_I\left\{\frac{\partial(\phi_2/\Phi_2)}{\partial \varphi_{it}}\frac{\partial \mathbb{X}^{F\prime}_{it}}{\partial \Theta_1^{\prime}}\Theta_{2F} + \frac{\partial(\phi_2/\Phi_2)}{\partial \gamma_{it}}\frac{\partial \mathbb{X}^{I\prime}_{it}}{\partial \Theta_1^{\prime}}\Theta_{2I}\right\} \tag{D-23}$$

when the firm year observation, it, is such that it belongs to CIS3 and CIS3.5, and CIS2.5 when $I_{it} = 1$. When $I_{it} = 0$, for CIS2.5 we have $\frac{\partial \mathcal{L}_{2\Theta_{2F}}(\Theta_1, \Theta_2)}{\partial \Theta_1^{\prime}} = \mathbf{0}_{\Theta_{2F}}$, $\frac{\partial \mathcal{L}_{2\rho_{\tilde{\zeta}\tilde{v}}}(\Theta_1, \Theta_2)}{\partial \Theta_1^{\prime}} = 0$, and

$$\begin{aligned}
\frac{\partial \mathcal{L}_{2\Theta_{2I}}(\Theta_1, \Theta_2)}{\partial \Theta_1^{\prime}} &= -\left\{\frac{\partial}{\partial \Theta_1^{\prime}}\left(\frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})}\right)\mathbb{X}^I_{it} + \frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})}\frac{\partial \mathbb{X}^I_{it}}{\partial \Theta_1^{\prime}}\right\} \\
&= -\left\{\frac{\partial}{\partial \gamma_{it}}\left(\frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})}\right)\frac{\partial \mathbb{X}^{I\prime}_{it}}{\partial \Theta_1^{\prime}}\Theta_{2I}\mathbb{X}^I_{it} + \frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})}\frac{\partial \mathbb{X}^I_{it}}{\partial \Theta_1^{\prime}}\right\}
\end{aligned} \tag{D-24}$$

To obtain expressions for (D-22), (D-23), and (D-24) we need the derivative of $\frac{g_j}{\Phi_2}$, $j \in \{F, I\}$, with respect to φ_{it} and γ_{it}, the derivative of $\frac{\phi_2}{\Phi_2}$ with respect to φ_{it} and γ_{it}, and the derivative of $\frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})}$ with respect to γ_{it}. While these can be easily obtained and can be found in Greene (2002), what is challenging to obtain is the derivative of $\mathbb{X}_{it}^F$ and $\mathbb{X}_{it}^I$ with respect to Θ_1.

$$\frac{\partial \mathbb{X}_{it}^j}{\partial \Theta_1'} = \begin{bmatrix} \frac{\partial \mathcal{X}_{it}^j}{\partial \boldsymbol{\delta}'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \bar{\boldsymbol{\delta}}'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \boldsymbol{\kappa}'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \sigma_\alpha^2} \\ \frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}+\hat{\alpha}_i)}{\partial \boldsymbol{\delta}'} & \frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}+\hat{\alpha}_i)}{\partial \bar{\boldsymbol{\delta}}'} & \frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}+\hat{\alpha}_i)}{\partial \boldsymbol{\kappa}'} & \frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}+\hat{\alpha}_i)}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} & \frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}+\hat{\alpha}_i)}{\partial \sigma_\alpha^2} \\ \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\delta}'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \bar{\boldsymbol{\delta}}'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\kappa}'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \sigma_\alpha^2} \end{bmatrix},$$

where $j \in \{F, I\}$. While $\frac{\partial \mathcal{X}_{it}^j}{\partial \Theta_1'} = \mathbf{0}$, below we show that

$$\begin{aligned}
\frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \boldsymbol{\delta}'} &= -\frac{1}{U_{dr}^2}\sum_{t=1}^{p}\Big[U_{nr}^2 - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\mathbf{Z}_{it}' \\
\frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \bar{\boldsymbol{\delta}}'} &= \bar{\mathcal{Z}}_i' - \frac{p}{U_{dr}^2}\Big[U_{nr}^2 - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\bar{\mathcal{Z}}_i' \\
\frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \boldsymbol{\kappa}'} &= -\frac{1}{U_{dr}^2}\sum_{t=1}^{p}\bigg\{\Big[U_{nr}^2 - U_{dr}F_{dr}\Big](\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}' - \mathbf{r}_{it}')\Sigma_{\epsilon\epsilon}^{-1} + \Big[U_{nr}F_{dr} - U_{dr}F_{nr}\Big]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\bigg\} \\
\frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} &= \frac{1}{2U_{dr}^2}\sum_{t=1}^{p}\Big[(U_{nr}^2 - U_{dr}F_{dr})\text{vec}(\boldsymbol{\kappa}\mathbf{r}_{it}' + \mathbf{r}_{it}\boldsymbol{\kappa}')' \\
&\quad + (U_{dr}F_{nr} - U_{nr}F_{dr})\text{vec}(\Sigma_\kappa)'\Big](\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})'L_m' \\
\frac{\partial (\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \sigma_\alpha^2} &= \frac{1}{2\sigma_\alpha^4 U_{dr}^2}(U_{dr}F_{nr} - U_{nr}F_{dr}) \qquad \text{(D-25)}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\delta}'} &= -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\mathbf{Z}_{it}' + \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{U_{dr}^2}\sum_{t=1}^{p}\Big[U_{nr}^2 - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\mathbf{Z}_{it}' \\
\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \bar{\boldsymbol{\delta}}'} &= -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\bar{\mathcal{Z}}_i' + \frac{p\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{U_{dr}^2}\Big[U_{nr}^2 - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\bar{\mathcal{Z}}_i' \\
\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\kappa}'} &= -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} - \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\alpha}_i + \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{U_{dr}^2}\sum_{t=1}^{p}\Big\{\Big[U_{nr}^2 - U_{dr}F_{dr}\Big](\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}' - \mathbf{r}_{it}')\Sigma_{\epsilon\epsilon}^{-1} \\
&\qquad + \Big[U_{nr}F_{dr} - U_{dr}F_{nr}\Big]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\Big\} \\
\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \mathrm{vech}(\Sigma_{\epsilon\epsilon})'} &= \Big[(\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\otimes I_m)\mathrm{vec}((\mathrm{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})' - (\boldsymbol{\epsilon}_{it}\otimes\Sigma_{\epsilon}')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'\Big]L_m' \\
&\quad - \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{2(U_{dr})^2}\sum_{t=1}^{p}\Big[(U_{nr}^2 - U_{dr}F_{dr})\mathrm{vec}(\boldsymbol{\kappa}\mathbf{r}_{it}' + \mathbf{r}_{it}\boldsymbol{\kappa}')' \\
&\qquad + (U_{dr}F_{nr} - U_{nr}F_{dr})\mathrm{vec}(\Sigma_{\kappa})'\Big](\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m' \\
\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \sigma_\alpha^2} &= -\frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{2\sigma_\alpha^4 U_{dr}^2}(U_{dr}F_{nr} - U_{nr}F_{dr}), \qquad \text{(D-26)}
\end{aligned}
$$

where

$$
U_{nr} = \int \alpha \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha, \quad F_{nr} = \int \alpha^3 \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha,
$$

$$
U_{dr} = \int \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha, \quad F_{dr} = \int \alpha^2 \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha,
$$

and L_m in the set of equations in (D-25) and (D-26) is the elimination matrix and $\mathbf{r}_{it} = \mathbf{x}_{it} - \mathbf{Z}_{it}\boldsymbol{\delta} - \boldsymbol{\kappa}\bar{\mathcal{Z}}_i\bar{\boldsymbol{\delta}}$. U_{nr}, U_{dr}, F_{nr}, and F_{dr} needed to estimate the covariance matrix of the structural parameters are obtained using Gauss Hermit quadrature rules.

D.2.1. *Derivation of the derivative of* $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ *and* $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}$ *and with respect to* Θ_1

Let us first consider the derivative of $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \boldsymbol{\alpha}_i$ and $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}$ with respect to $\boldsymbol{\delta}'$. We have

$$
\begin{aligned}
\frac{\partial(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \boldsymbol{\delta}'} &= \frac{\partial \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}}{\partial \boldsymbol{\delta}'} + \frac{\partial \hat{\alpha}_i}{\partial \boldsymbol{\delta}'} = 0 + \frac{\partial}{\partial \boldsymbol{\delta}'}\left[\frac{\int \alpha \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha}{\int \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha}\right] \\
&= 0 - \frac{1}{(\int \exp(.)\phi(\alpha)d\alpha)^2}\sum_{t=1}^{p}\Big[\int \alpha \exp(.)\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\mathbf{Z}_{it}'\phi(\alpha)d\alpha \int \exp(.)\phi(\alpha)d\alpha \\
&\quad - \int \alpha \exp(.)\phi(\alpha)d\alpha \int \exp(.)\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\mathbf{Z}_{it}'\phi(\alpha)d\alpha\Big], \qquad \text{(D-27)}
\end{aligned}
$$

To derive the above result in (D-27) we used the fact that

$$\frac{\partial(\boldsymbol{\epsilon}'_{it}\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_{it})}{\partial\boldsymbol{\delta}'} = 2\boldsymbol{\epsilon}'_{it}\Sigma^{-1}_{\epsilon\epsilon}\frac{\partial(\boldsymbol{\epsilon}_{it})}{\partial\boldsymbol{\delta}'} = -2\boldsymbol{\epsilon}'_{it}\Sigma^{-1}_{\epsilon\epsilon}\mathbf{Z}'_{it}.$$

Taking into account the fact that $\boldsymbol{\epsilon}_{it} = \mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - (\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}} + \alpha_i)\boldsymbol{\kappa}$, after some rearrangements it can be shown that

$$\frac{\partial\hat{\alpha}_i}{\partial\boldsymbol{\delta}'} = -\frac{1}{U^2_{dr}}\sum_{t=1}^{T}\Big[U^2_{nr} - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma^{-1}_{\epsilon\epsilon}\mathbf{Z}'_{it},$$

where

$$\begin{aligned} U_{nr} &= \int \alpha \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}'_{it}\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha, \quad F_{nr} = \int \alpha^3 \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}'_{it}\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha \\ U_{dr} &= \int \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}'_{it}\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha, \quad F_{dr} = \int \alpha^2 \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}'_{it}\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha. \end{aligned} \tag{D-28}$$

Hence we have

$$\frac{\partial(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial\boldsymbol{\delta}'} = -\frac{1}{U^2_{dr}}\sum_{t=1}^{p}\Big[U^2_{nr} - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma^{-1}_{\epsilon\epsilon}\mathbf{Z}'_{it}, \tag{D-29}$$

and

$$\begin{aligned} \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\boldsymbol{\delta}'} &= \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\boldsymbol{\kappa})}{\partial\boldsymbol{\delta}'} - \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\alpha}_i\boldsymbol{\kappa}}{\partial\boldsymbol{\delta}'} \\ &= -\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\mathbf{Z}'_{it} + \frac{\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\kappa}}{U^2_{dr}}\sum_{t=1}^{p}\Big[U^2_{nr} - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma^{-1}_{\epsilon\epsilon}\mathbf{Z}'_{it}. \end{aligned} \tag{D-30}$$

From (D-29) and (D-30) we can see that while $\frac{\partial(\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}}+\hat{\boldsymbol{\alpha}}_i)}{\partial\boldsymbol{\delta}'}$ for a firm i remains the same for all time periods, $\frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\boldsymbol{\delta}'}$ varies with time. Similarly it can be shown that

$$\frac{\partial(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial\bar{\boldsymbol{\delta}}'} = \bar{\mathcal{Z}}'_i - \frac{1}{U^2_{dr}}\sum_{t=1}^{p}\Big[U^2_{nr} - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\kappa}\bar{\mathcal{Z}}'_i = \bar{\mathcal{Z}}'_i - \frac{p}{U^2_{dr}}\Big[U^2_{nr} - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\kappa}\bar{\mathcal{Z}}'_i, \tag{D-31}$$

and

$$\begin{aligned} \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\boldsymbol{\delta}'} &= \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}}\boldsymbol{\kappa})}{\partial\boldsymbol{\delta}'} - \frac{\partial\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\hat{\alpha}_i\boldsymbol{\kappa}}{\partial\boldsymbol{\delta}'} \\ &= -\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\kappa}\bar{\mathcal{Z}}'_i + \frac{p\tilde{\Sigma}^{-1}_{\epsilon\epsilon}\boldsymbol{\kappa}}{U^2_{dr}}\Big[U^2_{nr} - U_{dr}F_{dr}\Big]\boldsymbol{\kappa}'\Sigma^{-1}_{\epsilon\epsilon}\boldsymbol{\kappa}\bar{\mathcal{Z}}'_i. \end{aligned} \tag{D-32}$$

Let us now consider the derivative of $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ with respect to $\boldsymbol{\kappa}$. We have

$$\begin{aligned}\frac{\partial(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \boldsymbol{\kappa}'} &= \frac{\partial}{\partial \boldsymbol{\kappa}'}\left[\frac{\int \alpha \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha}{\int \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha}\right] \\ &= -\frac{1}{(\int \exp(.)\phi(\alpha)d\alpha)^2}\sum_{t=1}^{p}\Bigg[\int \alpha \exp(.)\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}(\bar{\mathcal{Z}}_i'\boldsymbol{\delta} + \alpha_i)\phi(\alpha)d\alpha \int \exp(.)\phi(\alpha)d\alpha \\ &\quad - \int \alpha \exp(.)\phi(\alpha)d\alpha \int \exp(.)\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}(\bar{\mathcal{Z}}_i'\boldsymbol{\delta} + \alpha_i)'\phi(\alpha)d\alpha\Bigg], \qquad \text{(D-33)}\end{aligned}$$

which after simplification can be written as

$$\frac{\partial(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \boldsymbol{\kappa}'} = -\frac{1}{U_{dr}^2}\sum_{t=1}^{p}\left\{\left[U_{nr}^2 - U_{dr}F_{dr}\right](\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}' - \mathbf{r}_{it}')\Sigma_{\epsilon\epsilon}^{-1} + \left[U_{nr}F_{dr} - U_{dr}F_{nr}\right]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\right\}, \qquad \text{(D-34)}$$

where $\mathbf{r}_{it} = \mathbf{x}_{it} - \mathbf{Z}_{it}'\boldsymbol{\delta} - \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}$ and U_{nr}, U_{dr}, F_{nr}, and F_{dr} are given in (D-28). Also, it can be shown that

$$\begin{aligned}\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\kappa}'} &= \frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbf{Z}_{it}'\boldsymbol{\delta} - \bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}'} - \frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\alpha}_i\boldsymbol{\kappa}}{\partial \boldsymbol{\kappa}'} \\ &= -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} - \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\alpha}_i + \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{U_{dr}^2}\sum_{t=1}^{p}\Bigg\{\left[U_{nr}^2 - U_{dr}F_{dr}\right](\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}' - \mathbf{r}_{it}')\Sigma_{\epsilon\epsilon}^{-1} \\ &\quad + \left[U_{nr}F_{dr} - U_{dr}F_{nr}\right]\boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\Bigg\}. \qquad \text{(D-35)}\end{aligned}$$

Now consider the derivative of $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ with respect to $\text{vech}(\Sigma_{\epsilon\epsilon})$. We have

$$\begin{aligned}\frac{\partial(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} &= \frac{\partial\hat{\alpha}_i}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{\partial}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'}\left[\frac{\int \alpha \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha}{\int \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)d\alpha}\right] \\ &= -\frac{1}{2}\left[\frac{\int \alpha\psi(\alpha)\frac{\partial \sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'}d\alpha \int \psi(\alpha)d\alpha - \int \alpha\psi(\alpha)d\alpha \int \psi(\alpha)\frac{\partial \sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'}d\alpha}{(\int \psi(\alpha)d\alpha)^2}\right],\end{aligned}$$

where $\psi(\alpha) = \exp(-\frac{1}{2}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it})\phi(\alpha)$. With $\frac{\partial \sum_{t=1}^{p}\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = \sum_{t=1}^{p}\text{vec}(-(\Sigma_{\epsilon\epsilon}^{-1})'\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'(\Sigma_{\epsilon\epsilon}^{-1})')'L_m'$

the above can be written as

$$\begin{aligned}\frac{\partial\hat{\alpha}_i}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} =&\frac{1}{2(\int\psi(\alpha)d\alpha)^2}\sum_{t=1}^{p}\Big[\int\alpha\psi(\alpha)\text{vec}((\Sigma_{\epsilon\epsilon}^{-1})'\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'(\Sigma_{\epsilon\epsilon}^{-1})')'L_m'd\alpha\int\psi(\alpha)d\alpha\\
&-\int\psi(\alpha)\text{vec}((\Sigma_{\epsilon\epsilon}^{-1})'\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'(\Sigma_{\epsilon\epsilon}^{-1})')'L_m'd\alpha\int\alpha\psi(\alpha)d\alpha\Big]\\
=&\frac{1}{2U_{dr}^2}\sum_{t=1}^{T}\Big[\int\alpha\psi(\alpha)\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'd\alpha U_{dr}\\
&-U_{nr}\int\psi(\alpha)\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'd\alpha\Big]\\
=&\frac{1}{2U_{dr}^2}\sum_{t=1}^{p}\Big[\int(U_{dr}\alpha\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')'-U_{nr}\text{vec}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}')')\psi(\boldsymbol{\alpha})d\alpha\Big](\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m',\end{aligned}\tag{D-36}$$

where L_m is an elimination matrix. To simply further, write $\boldsymbol{\epsilon}_{it}$ as $\boldsymbol{\epsilon}_{it}=\mathbf{x}_{it}-\mathbf{Z}_{it}\boldsymbol{\delta}-\boldsymbol{\kappa}\bar{\mathcal{Z}}_i\bar{\boldsymbol{\delta}}-\boldsymbol{\kappa}\alpha=\mathbf{r}_{it}-\boldsymbol{\kappa}\alpha$, where $\mathbf{r}_{it}=\mathbf{x}_{it}-\mathbf{Z}_{it}\boldsymbol{\delta}-\boldsymbol{\kappa}\bar{\mathcal{Z}}_i\bar{\boldsymbol{\delta}}$. Then $\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}'=\mathbf{r}_{it}\mathbf{r}_{it}'-\boldsymbol{\kappa}\mathbf{r}_{it}'\alpha-\mathbf{r}_{it}\boldsymbol{\kappa}'\alpha+\boldsymbol{\kappa}\boldsymbol{\kappa}'\alpha^2$, then (D-36) after some simplification can be written as

$$\begin{aligned}\frac{\partial\hat{\alpha}_i}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'}=&\frac{1}{2U_{dr}^2}\sum_{t=1}^{p}\Big[(U_{nr}^2-U_{dr}F_{dr})\text{vec}(\boldsymbol{\kappa}\mathbf{r}_{it}'+\mathbf{r}_{it}\boldsymbol{\kappa}')'\\
&+(U_{dr}F_{nr}-U_{nr}F_{dr})\text{vec}(\Sigma_\kappa)'\Big](\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'.\end{aligned}\tag{D-37}$$

where U_{nr}, U_{dr}, F_{nr}, and F_{dr} have been defined in (D-28) and $\Sigma_\kappa=\boldsymbol{\kappa}\boldsymbol{\kappa}'$. Let us now consider the derivative $\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}=\frac{\partial(\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it})}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}=\frac{\partial(\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\mathbf{r}_{it})}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}-\frac{\partial(\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\hat{\alpha}_i)}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}$. The total differential of $\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\hat{\alpha}_i$ is given by:

$$d(\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\hat{\alpha}_i)=d(\Sigma_\epsilon)\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\hat{\alpha}_i+\Sigma_\epsilon d(\Sigma_{\epsilon\epsilon}^{-1})\boldsymbol{\kappa}\hat{\alpha}_i+\Sigma_\epsilon\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}d(\hat{\alpha}_i).\tag{D-38}$$

Now, as defined earlier, $\Sigma_\epsilon=(\text{dg}(\Sigma_{\epsilon\epsilon}))^{1/2}$, hence

$$\begin{aligned}\frac{\partial(\Sigma_\epsilon)\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\hat{\alpha}_i}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}&=\frac{1}{2}(\boldsymbol{\kappa}'\hat{\alpha}_i\Sigma_{\epsilon\epsilon}^{-1}\otimes I_m)\text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})\frac{\partial\text{vec}(\Sigma_{\epsilon\epsilon})}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}\\
&=\frac{1}{2}(\boldsymbol{\kappa}'\hat{\alpha}_i\Sigma_{\epsilon\epsilon}^{-1}\otimes I_m)\text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})'L_m'.\end{aligned}\tag{D-39}$$

Now, consider the second term of the differential given in (D-38). It can be shown that

$$\frac{\Sigma_\epsilon\partial(\Sigma_{\epsilon\epsilon}^{-1})\boldsymbol{\kappa}\hat{\alpha}_i}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}=-(\boldsymbol{\kappa}\hat{\alpha}_i\otimes\Sigma_\epsilon')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})\frac{\partial\text{vec}(\Sigma_{\epsilon\epsilon})}{\partial\text{vech}(\Sigma_{\epsilon\epsilon})'}=-(\boldsymbol{\kappa}\hat{\alpha}_i\otimes\Sigma_\epsilon')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})L_m'.\tag{D-40}$$

Now consider the third term in the total differential in (D-38). From (D-37) we can conclude that

$$\frac{\Sigma_\epsilon \Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\partial(\hat{\alpha}_i)}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{\Sigma_\epsilon \Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{2U_{dr}^2}\sum_{t=1}^{p}\Big[(U_{nr}^2 - U_{dr}F_{dr})\text{vec}(\boldsymbol{\kappa}\mathbf{r}_{it}' + \mathbf{r}_{it}\boldsymbol{\kappa}')' \\ + (U_{dr}F_{nr} - U_{nr}F_{dr})\text{vec}(\Sigma_\kappa)'\Big](\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'. \tag{D-41}$$

Combining (D-39), (D-40), and (D-41) we obtain

$$\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = \Big[(\boldsymbol{\epsilon}_{it}'\Sigma_{\epsilon\epsilon}^{-1}\otimes I_m)\text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})' - (\boldsymbol{\epsilon}_{it}\otimes\Sigma_\epsilon')'(\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'\Big]L_m' \\ - \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{2(U_{dr})^2}\sum_{t=1}^{p}\Big[(U_{nr}^2 - U_{dr}F_{dr})\text{vec}(\boldsymbol{\kappa}\mathbf{r}_{it}' + \mathbf{r}_{it}\boldsymbol{\kappa}')' + (U_{dr}F_{nr} - U_{nr}F_{dr})\text{vec}(\Sigma_\kappa)'\Big](\Sigma_{\epsilon\epsilon}^{-1}\otimes\Sigma_{\epsilon\epsilon}^{-1})'L_m'. \tag{D-42}$$

Finally, let consider the derivative of $\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ with respect to σ_α^2. We have

$$\frac{\partial(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial\sigma_\alpha^2} = \frac{\partial\hat{\alpha}_i}{\partial\sigma_\alpha^2} = \frac{\partial}{\partial\sigma_\alpha^2}\left[\frac{\int\alpha\exp(.)\phi(\alpha)d\alpha}{\int\exp(.)\phi(\alpha)d\alpha}\right] = \\ = \frac{[\int\alpha\exp(.)\frac{\partial\phi(\alpha)}{\partial\sigma_\alpha^2}d\alpha][\int\exp(.)\phi(\alpha)d\alpha] - [\int\alpha\exp(.)\phi(\alpha)d\alpha][\int\exp(.)\frac{\partial\phi(\alpha)}{\partial\sigma_\alpha^2}d\alpha]}{[\int\exp(.)\phi(\alpha)d\alpha]^2}.$$

Given that $\frac{\partial\phi(\alpha)}{\partial\sigma_\alpha^2} = -\frac{1}{2\sigma_\alpha^2}\phi(\alpha) + \frac{\alpha^2}{2\sigma_\alpha^4}\phi(\alpha)$, the above after simplification reduces to

$$\frac{\partial(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial\sigma_\alpha^2} = \frac{1}{2\sigma_\alpha^4 U_{dr}^2}(U_{dr}F_{nr} - U_{nr}F_{dr}), \tag{D-43}$$

and we can write $\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}}{\sigma_\alpha^2}$ as

$$\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = -\frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\partial(\hat{\alpha}_i)}{\partial\sigma_\alpha^2} = -\frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{2\sigma_\alpha^4 U_{dr}^2}(U_{dr}F_{nr} - U_{nr}F_{dr}). \tag{D-44}$$

D.3. *Derivative of $\mathcal{L}_{i(p)3\Theta_3}$ with respect to Θ_1 and Θ_2*

As stated earlier in order to construct error corrected standard errors of the structural parameters we also need sample analogs of $\mathbb{L}_{3\Theta_3\Theta_1}$, $\mathbb{L}_{3\Theta_3\Theta_2}$, and $\mathbb{L}_{3\Theta_3\Theta_3}$ to construct B_* in (D-9). While it is straightforward to compute sample analog of $\mathbb{L}_{3\Theta_3\Theta_3}$, computation of sample analogs of $\mathbb{L}_{3\Theta_3\Theta_1}$ and $\mathbb{L}_{3\Theta_3\Theta_2}$ needs some work. Here we derive the derivative of $\mathcal{L}_{i(p)3\Theta_3}(\Theta_1, \Theta_2, \Theta_2)$ with respect to Θ_1 and Θ_2. Now, we know that

$$\frac{\partial \mathcal{L}_{i(p)3\Theta_3}}{\partial \Theta_j'} = \sum_{t=1}^{p} \frac{\partial \mathcal{L}_{it3\Theta_3}}{\partial \Theta_j'} = \sum_{t=1}^{p} \frac{\partial}{\partial \Theta_j'} I_{it}[\mathbb{X}_{it}^R(\Theta_1, \hat{\Theta}_2)(R_{it} - \mathbb{X}_{it}^R(.)'\Theta_3)]$$
$$= \sum_{t=1}^{p} I_{it}\left[\frac{\mathbb{X}_{it}^R(.)}{\partial \Theta_j'}(R_{it} - \mathbb{X}_{it}^R(.)'\hat{\Theta}_3) + \mathbb{X}_{it}^R(\hat{\Theta}_1, \hat{\Theta}_2)\frac{\mathbb{X}_{it}^R(.)'}{\partial \Theta_j'}\hat{\Theta}_3\right] \quad j \in \{1, 2\}, \tag{D-45}$$

where

$$\mathbb{X}_{it}^R(\hat{\Theta}_1, \hat{\Theta}_2) = \begin{bmatrix} \mathcal{X}_{it}^R \\ F_{it}(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \\ (1 - F_{it})(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \\ F_{it}\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it} \\ (1 - F_{it})\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it} \\ F_{it}C_{11}(\Theta_1, \Theta_2)_{it} \\ (1 - F_{it})C_{01}(\Theta_1, \Theta_2)_{it} \\ F_{it}C_{12}(\Theta_1, \Theta_2)_{it} \\ (1 - F_{it})C_{02}(\Theta_1, \Theta_2)_{it} \end{bmatrix}.$$

And $\mathcal{X}_{it}^R = \{\mathcal{X}_{1it}^{R\prime}, \mathcal{X}_{0it}^{R\prime}\}'$ where $\mathcal{X}_{1it}^R$ and $\mathcal{X}_{0it}^R$ have been defined in equation (3.5) in the main text.

We know that $\frac{\mathcal{X}_{it}^R}{\Theta_1'} = \frac{\mathcal{X}_{it}^R}{\Theta_2'} = \mathbf{0}$, that $\frac{\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i}{\Theta_2'} = \frac{\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\Theta_2'} = \mathbf{0}$ and $\frac{\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i}{\Theta_1'}$ and $\frac{\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\Theta_1'}$ have been derived above. Here we derive the derivatives of the remaining correction terms, C_{11}, C_{12}, C_{01}, and C_{02} with respect to Θ_1 and Θ_2. We have

$$\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \Theta_1'} = \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \varphi_{it}} \frac{\partial \mathbb{X}_{it}^{F\prime}\Theta_{2F}}{\partial \Theta_1'} + \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \gamma_{it}} \frac{\partial \mathbb{X}_{it}^{I\prime}\Theta_{2I}}{\partial \Theta_1'}, \qquad j \in \{0, 1\}, k \in \{1, 2\}. \tag{D-46}$$

Given the functional form of $C_{jk}(\Theta_1, \Theta_2)$ in equations (3.20) and (3.21), its derivative with respect to φ_{it} and γ_{it} can be easily obtained. The partial derivatives $\frac{\partial \mathbb{X}_{it}^{F\prime}}{\partial \Theta_1'}$ and $\frac{\partial \mathbb{X}_{it}^{I\prime}}{\partial \Theta_1'}$ have been worked out above. Now consider the derivative of $C_{jk}(\Theta_1, \Theta_2)$ with respect to

$\Theta_2 = \{\Theta'_{2F}, \Theta'_{2I}, \rho_{\tilde{\zeta}\tilde{\upsilon}}\}'$.

$$\frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \Theta'_2} = \begin{bmatrix} \frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \varphi_{it}} \frac{\partial \mathbb{X}^{F\prime}_{it}\Theta_{2F}}{\partial \Theta_{2F}} \\ \frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \gamma_{it}} \frac{\partial \mathbb{X}^{I\prime}_{it}\Theta_{2I}}{\partial \Theta_{2I}} \\ \frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \rho_{\tilde{\zeta}\tilde{\upsilon}}} \end{bmatrix}' = \begin{bmatrix} \frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \varphi_{it}} \mathbb{X}^{F}_{it} \\ \frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \gamma_{it}} \mathbb{X}^{I}_{it} \\ \frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \rho_{\tilde{\zeta}\tilde{\upsilon}}} \end{bmatrix}' \qquad \text{(D-47)}$$

Again, given the functional form of $C_{jk}(\Theta_1,\Theta_2)$, $\frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \varphi_{it}}$ and $\frac{\partial C_{jk}(\Theta_1,\Theta_2)_{it}}{\partial \gamma_{it}}$ can be easily computed. We note that, depending on the particular combination of j and k, the derivatives stated above involve taking derivatives of $\Pr(F_{it} = 1, I_{it} = 1)$ and $\Pr(F_{it} = 0, I_{it} = 1)$ with respect to φ_{it}, γ_{it} and $\rho_{\tilde{\zeta}\tilde{\upsilon}}$, and these are stated in Greene (2002).

APPENDIX E: ESTIMATION OF AVERAGE PARTIAL EFFECTS

In this section we discuss estimation of Average Partial Effects (APE) and testing hypothesis about the APEs for the structural equations.

E.1. *Average Partial Effects for the Second Stage*

E.1.1. *Estimation*

In the second stage, as discussed earlier, we jointly estimate the parameters of Innovation and Financial Constraint equations,

$$\begin{aligned} I_t &= 1\{I_t^* > 0\} = 1\{\mathcal{X}_t^{I\prime}\boldsymbol{\gamma} + \theta\hat{\hat{\alpha}} + \tilde{\Sigma}_{\upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \tilde{\upsilon}_t > 0\} \\ F_t &= 1\{F_t^* > 0\} = 1\{\mathcal{X}_t^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\hat{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \tilde{\zeta}_t > 0\}, \end{aligned}$$

given in equations (3.12) and (3.13) in the main text above. In our discussion of the identification of structural parameters of interest and the APE for nonlinear model in Appendix A, we had shown how to estimate the APE of covariates for the unconditional probability of being financially constrained or being an innovator.

We may also be interested in the APE of a variable on the conditional probability of an event, or compare the APE of a variable on the probability of an event conditional on two mutually exclusive events. For example, we may be interested in the marginal effect of w, say long-term debt to asset ratio, on the probability of a firm being an innovator, $I_t = 1$, conditional on it being financially constrained, $F_t = 1$, as compared to the APE of w, on the probability of $I_t = 1$, conditional on $F_t = 0$. We know that for a firm i in time period t

$$\Pr(I_t = 1|F_t = 1, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \frac{\Pr(I_t = 1, F_t = 1|\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)}{\Pr(F_t = 1|\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)} = \frac{\Phi_2(\varphi_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t), \gamma_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t), \rho_{\tilde{\zeta}\tilde{\upsilon}})}{\Phi(\varphi_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t))},$$

$$\Pr(I_t = 1|F_t = 0, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \frac{\Pr(I_t = 1, F_t = 0|\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)}{\Pr(F_t = 0|\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)} = \frac{\Phi_2(\varphi_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t), -\gamma_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t), -\rho_{\tilde{\zeta}\tilde{\upsilon}})}{1 - \Phi(\varphi_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t))},$$

where Φ_2 is the cumulative distribution function of a standard bivariate normal and

$$\varphi_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \boldsymbol{\mathcal{X}}_t^{F\prime}\boldsymbol{\varphi} + \lambda\hat{\hat{\alpha}} + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t, \text{ and } \gamma_t(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \boldsymbol{\mathcal{X}}_t^{I\prime}\boldsymbol{\gamma} + \theta\hat{\hat{\alpha}} + \tilde{\Sigma}_{\upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t.$$

Hence, for a firm i we have

$$\frac{\partial \Pr(I_t = 1|F_t = 1)}{\partial w} = \int \frac{\partial}{\partial w}\left(\frac{\Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{\upsilon}})}{\Phi(\varphi_t)}\right) dF_{\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}}. \tag{E-1}$$

If w belongs to both the specifications, φ_t and γ_t, then the above involves taking derivative of CDF of a standard bivariate normal with respect to φ_t and γ_t. It can be shown that

$$\frac{\partial}{\partial w}\left(\frac{\Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{\upsilon}})}{\Phi(\varphi_t)}\right) = \frac{1}{\Phi(\varphi_t)}\left[g_I\gamma_w + \left(g_F - \Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{\upsilon}})\frac{\phi(\varphi_t)}{\Phi(\varphi_t)}\right)\varphi_w\right], \tag{E-2}$$

where

$$g_F = \phi(\varphi_t)\Phi\left(\frac{\gamma_t - \rho_{\tilde{\zeta}\tilde{\upsilon}}\varphi_t}{\sqrt{1 - \rho^2_{\tilde{\zeta}\tilde{\upsilon}}}}\right) \text{ and } g_I = \phi(\gamma_t)\Phi\left(\frac{\varphi_t - \rho_{\tilde{\zeta}\tilde{\upsilon}}\gamma_t}{\sqrt{1 - \rho^2_{\tilde{\zeta}\tilde{\upsilon}}}}\right). \tag{E-3}$$

The derivatives of the other conditional probabilities with respect to φ_t and γ_t can be found in Greene (2002). Once the integrand in (E-1) is estimated at $\boldsymbol{\mathcal{X}}_t^F = \bar{\boldsymbol{\mathcal{X}}}^F$ and $\boldsymbol{\mathcal{X}}_t^I = \bar{\boldsymbol{\mathcal{X}}}^I$, given the estimates $\hat{\hat{\alpha}}_i$ and $\hat{\hat{\boldsymbol{\epsilon}}}_{it}$, the APE of w on the conditional probabilities are estimated by taking an average over all firm-year observations.

E.1.2. *Hypothesis Testing*

To test various hypothesis in order to draw inferences about the APE's we need to compute the standard errors of their estimates. From (A-20) in Appendix A we know that estimated APE of w on the unconditional probability of being, say, financially constrained for firm i in time period t is given by

$$\frac{\partial\widehat{\Pr}(F_t = 1)}{\partial w} = \frac{1}{\sum_{i=1}^N T_i}\sum_{i=1}^N\sum_{t=1}^{T_i}\hat{\varphi}_w\phi(\bar{\mathbb{X}}_{it}^{F\prime}\underline{\hat{\boldsymbol{\varphi}}}),$$

where $\bar{\mathbb{X}}_{it}^{F} = \{\bar{\mathcal{X}}^{F\prime}, \hat{\hat{\alpha}}_i, (\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it})'\}'$ and $\underline{\hat{\boldsymbol{\varphi}}} = \{\hat{\boldsymbol{\varphi}}', \hat{\lambda}, \hat{\hat{\Sigma}}_{\zeta\epsilon}'\}'$. Since each of the $\hat{\varphi}_w \phi(\bar{\mathbb{X}}_{it}^{F\prime}\underline{\hat{\boldsymbol{\varphi}}})$ is a function of $\underline{\hat{\boldsymbol{\varphi}}}$ the variance of $\frac{\partial \widehat{\Pr}(F_t=1)}{\partial w}$ will be a function of the variance of the estimate of $\underline{\boldsymbol{\varphi}}$. Now, we know that by the linear approximation approach (delta method), the asymptotic covariance matrix of $\frac{\partial \widehat{\Pr}(F_t=1)}{\partial w}$ is given by

$$\text{Asy. Var}[\frac{\partial \widehat{\Pr}(F_t=1)}{\partial w}] = \left[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\partial \hat{\varphi}_w \phi(\bar{\mathbb{X}}_{it}^{F\prime}\underline{\hat{\boldsymbol{\varphi}}})}{\partial \underline{\hat{\boldsymbol{\varphi}}}'}\right] V_{2F}^{*}\left[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\partial \hat{\varphi}_w \phi(\bar{\mathbb{X}}_{it}^{F\prime}\underline{\hat{\boldsymbol{\varphi}}})}{\partial \underline{\hat{\boldsymbol{\varphi}}}'}\right]', \tag{E-4}$$

where V_{2F}^{*} is the second stage error adjusted covariance matrix, shown in appendix D, of $\underline{\hat{\boldsymbol{\varphi}}}$. In the RHS of (E-4)

$$\frac{\partial \hat{\varphi}_w \phi(\bar{\mathbb{X}}_{it}^{F\prime}\underline{\hat{\boldsymbol{\varphi}}})}{\partial \underline{\hat{\boldsymbol{\varphi}}}'} = \phi(\bar{\mathbb{X}}_{it}^{F\prime}\underline{\hat{\boldsymbol{\varphi}}})[e_w - (\underline{\hat{\boldsymbol{\varphi}}}'\bar{\mathbb{X}}_{it}^{F})\hat{\varphi}_w \bar{\mathbb{X}}_{it}^{F\prime}], \tag{E-5}$$

where and e_w is a row vector having the dimension of $\underline{\boldsymbol{\varphi}}'$ and with 1 at the position of φ_w in $\underline{\boldsymbol{\varphi}}$ and zeros elsewhere.

If w is a dummy variable then from (A-21) we know that the estimated APE of w on the probability of being financially constrained in time period t, given $\mathcal{X}_t^F = \bar{\mathcal{X}}^F$ is given by

$$\begin{aligned}\Delta_w \Pr(F_t=1) &= \frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\Phi(\bar{\mathcal{X}}_{-w}^{F}, w=1, \hat{\hat{\alpha}}_i, \hat{\boldsymbol{\epsilon}}_{it}) - \Phi(\bar{\mathcal{X}}_{-w}^{F}, w=0, \hat{\hat{\alpha}}_i, \hat{\boldsymbol{\epsilon}}_{it}) \\ &= \frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\Delta_w \Phi_{it}(.).\end{aligned}$$

To obtain the variance of the above, again by the delta method we have

$$\text{Asy. Var}\Delta_w \Pr(F_t=1) = \left[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\partial \Delta\Phi_{it}(.)}{\partial \underline{\hat{\boldsymbol{\varphi}}}}\right]' V_{2f}^{*}\left[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\partial \Delta\Phi_{it}(.)}{\partial \underline{\hat{\boldsymbol{\varphi}}}}\right], \tag{E-6}$$

where

$$\frac{\partial \Delta\Phi_{it}(.)}{\partial \underline{\hat{\boldsymbol{\varphi}}}} = \frac{\partial \hat{\Phi}_{it}(., w=1)}{\partial \underline{\hat{\boldsymbol{\varphi}}}} - \frac{\partial \Phi_{it}(., w=0)}{\partial \underline{\hat{\boldsymbol{\varphi}}}} = \phi_{it}(., w=1)\begin{bmatrix}\bar{\mathbb{X}}_{it-w}^{F} \\ 1\end{bmatrix} - \phi_{it}(., w=0)\begin{bmatrix}\bar{\mathbb{X}}_{it-w}^{F} \\ 0\end{bmatrix}.$$

Substituting the above in (E-6) gives the asymptotic variance of the APE of the dummy variable w.

Delta method can also be applied for to obtain the asymptotic variance of the APE's of the continuous or dummy variable on the conditional probability of say being an innovator given the firm is financially constrained or not financially constrained. Let $\bar{\mathbb{X}}_{2it} = \{\bar{\mathbb{X}}_{it}^{F\prime}, \bar{\mathbb{X}}_{it}^{I\prime}\}'$ and $\Theta_2 = \{\underline{\boldsymbol{\varphi}}', \underline{\boldsymbol{\gamma}}', \rho_{\tilde{\zeta}\tilde{v}}\}'$, where $\bar{\mathbb{X}}_{it}^{I\prime} = \{\bar{\mathcal{X}}^{I\prime}, \hat{\hat{\alpha}}_i, (\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\hat{\boldsymbol{\epsilon}}}_{it})'\}'$ and $\underline{\boldsymbol{\gamma}} = \{\boldsymbol{\gamma}', \theta, \tilde{\Sigma}_{v\epsilon}'\}'$, and denote the right hand side of (E-2) as $\Lambda_{(I=1|F=1),w}(\bar{\mathbb{X}}_{2it}, \Theta_2)$. Then the APE of w on the conditional probability of being an innovator given that the firm is financially constrained is given by

$$\frac{\partial \widehat{\Pr}(I_t = 1|F_t = 1)}{\partial w} = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \Lambda_{(I=1|F=1),w}(\bar{\mathbb{X}}_{2it}, \hat{\Theta}_2)$$

By the delta method we know that the asymptotic variance of $\frac{\partial \widehat{\Pr}(I_t=1|F_t=1)}{\partial w}$ is given by

$$\left[\frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial \Lambda_{(I=1|F=1),w}(\bar{\mathbb{X}}_{2it}, \hat{\Theta}_2)}{\partial \Theta_2'}\right] V_2^* \left[\frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial \Lambda_{(I=1|F=1),w}(\bar{\mathbb{X}}_{2it}, \hat{\Theta}_2)}{\partial \Theta_2'}\right]', \tag{E-7}$$

where V_2^* is second stage error corrected covariance matrix of $\hat{\Theta}_2$. The derivative of $\Lambda_{(s=1|f=1),w}(\bar{\mathbb{X}}_{2it}, \hat{\Theta}_2)$ with respect to the second stage parameters, Θ_2, can easily obtained, even though the algebra is a bit messy.

E.2. *Average Partial Effects for the Third Stage*

One of the purposes of this exercise is to measure the effect of financial constraints, $F_t = 1$, on R&D expenditure. For a firm i in time period t, given $\mathcal{X}_t = \bar{\mathcal{X}}$, where $\mathcal{X}_t$ is the union of elements appearing in $\mathcal{X}_t^R$, $\mathcal{X}_t^F$, and $\mathcal{X}_t^I$, the APE of financial constraint on R&D intensity is computed as the difference in the expected R&D expenditure between the two regimes, financially constrained and non-financially constrained, averaged over $\hat{\hat{\alpha}}$ and $\hat{\boldsymbol{\epsilon}}$. The conditional, conditional on being an innovator ($s_{it} = 1$), APE of financial constraint on R&D expenditure is given by

$$\begin{aligned} \Delta_F \mathrm{E}(R_t|\bar{\mathcal{X}}) = &\int \mathrm{E}(R_{1t}|\bar{\mathcal{X}}, F_t = 1, I_t = 1, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}) dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}} \\ &- \int \mathrm{E}(R_{0t}|\bar{\mathcal{X}}, F_t = 0, I_t = 1, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}) dF_{\hat{\hat{\alpha}},\hat{\boldsymbol{\epsilon}}}. \end{aligned} \tag{E-8}$$

From the discussion of the third stage estimation we know that for a firm i

$$\begin{aligned}&\mathrm{E}(R_{1t}|\bar{\mathcal{X}}, F_t = 1, I_t = 1, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \\ &\beta_f + \bar{\mathcal{X}}^{R\prime}\boldsymbol{\beta}_1 + \mu_1\hat{\hat{\alpha}} + \tilde{\Sigma}_{\eta_1\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \sigma_{\tilde{\eta}_1}\rho_{\tilde{\eta}_1\tilde{\zeta}}C_{11}(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) + \sigma_{\tilde{\eta}_1}\rho_{\tilde{\eta}_1\tilde{\upsilon}}C_{12}(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)\end{aligned} \tag{E-9}$$

if $F_t^* > 0$, and

$$\begin{aligned}&\mathrm{E}(R_{0t}|\bar{\mathcal{X}}, F_t = 0, I_t = 1, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) = \\ &\bar{\mathcal{X}}_t^{R\prime}\boldsymbol{\beta}_0 + \mu_0\hat{\hat{\alpha}} + \Sigma_{\eta_0\epsilon}\Sigma_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \sigma_{\tilde{\eta}_0}\rho_{\tilde{\eta}_0\tilde{\zeta}}C_{01}(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t) + \sigma_{\tilde{\eta}_0}\rho_{\tilde{\eta}_0\tilde{\upsilon}}C_{02}(\hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)\end{aligned} \tag{E-10}$$

if $F_t^* \leq 0$, and where the correction terms – $C_{11}(\bar{\mathcal{X}}^I, \bar{\mathcal{X}}^F, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)$, $C_{12}(\bar{\mathcal{X}}^F, \bar{\mathcal{X}}^F, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)$, $C_{01}(\bar{\mathcal{X}}^I, \bar{\mathcal{X}}^F, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)$, and $C_{02}(\bar{\mathcal{X}}^I, \bar{\mathcal{X}}^F, \hat{\hat{\alpha}}, \hat{\boldsymbol{\epsilon}}_t)$ – are defined at the given $\mathcal{X}_t^I = \bar{\mathcal{X}}^I$ and $\mathcal{X}_t^F = \bar{\mathcal{X}}^F$. Given the above, an estimate of the APE of financial constraint on R&D intensity, can be obtained by taking the average of the difference in (E-9) and (E-10) over all firm-year observations for which $I_t = 1$.

The unconditional APE's of all other variables in the specification are simply the coefficient estimates of the two regimes of the switching regression model.

E.2.1. *Hypothesis Testing*

Since the APE of being financially constrained in the third stage switching regression model is a function of the correction terms constructed from the estimates of the seconds stage, the variance of the APE will be a function of the variances of the correction terms. Since the correction terms are in turn functions of the estimated coefficients in the second stage, the variance of the estimated APE be a function of the variance of the estimated second stage coefficients.

To see this, consider the the conditional APE of the financial constraint on the R&D expenditure, which is given by

$$\begin{aligned}\Delta_F\hat{\mathrm{E}}(R_t|\bar{\mathcal{X}}) =& \frac{1}{\sum_{i=1}^N T_i}\sum_{i=1}^N\sum_{t=1}^{T_i}\Bigg[I_t\Bigg(\hat{\beta}_f + \bar{\mathcal{X}}^{R\prime}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0) + (\hat{\mu}_1 - \hat{\mu}_0)\hat{\hat{\hat{\alpha}}} + (\hat{\hat{\Sigma}}_{\eta_1\epsilon_k} - \hat{\hat{\Sigma}}_{\eta_0\epsilon_k})\hat{\hat{\Sigma}}_{\epsilon\epsilon}^{-1}\hat{\hat{\boldsymbol{\epsilon}}}_t \\ &+ \widehat{\sigma_{\tilde{\eta}_1}\rho}_{\tilde{\eta}_1\tilde{\zeta}}C_{11}(\hat{\hat{\hat{\alpha}}}, \hat{\hat{\boldsymbol{\epsilon}}}_t) + \widehat{\sigma_{\tilde{\eta}_1}\rho}_{\tilde{\eta}_1\tilde{\upsilon}}C_{12}(\hat{\hat{\hat{\alpha}}}, \hat{\hat{\boldsymbol{\epsilon}}}_t) - \widehat{\sigma_{\tilde{\eta}_0}\rho}_{\tilde{\eta}_0\tilde{\zeta}}C_{01}(\hat{\hat{\hat{\alpha}}}, \hat{\hat{\boldsymbol{\epsilon}}}_t) - \widehat{\sigma_{\tilde{\eta}_0}\rho}_{\tilde{\eta}_0\tilde{\upsilon}}C_{02}(\hat{\hat{\hat{\alpha}}}, \hat{\hat{\boldsymbol{\epsilon}}}_t)\Bigg)\Bigg]\end{aligned} \tag{E-11}$$

Let us denote the structural coefficients of our model as $\Theta_s = \{\Theta_2', \Theta_3'\}'$ where Θ_2' and Θ_3' are the vector of structural coefficients estimated in the third stage respectively. Again, by the application of the delta method we know that

$$\text{Asy. Var}[\Delta_F \hat{\text{E}}(R_t|\bar{\mathcal{X}})] = \left[\frac{\partial \Delta_F \hat{\text{E}}(R_t|\bar{\mathcal{X}})}{\partial \Theta_s}\right]' V_s^* \left[\frac{\partial \Delta_F \hat{\text{E}}(R_t|\bar{\mathcal{X}})}{\partial \Theta_s}\right], \tag{E-12}$$

where V_s^*, the error corrected asymptotic covariance matrix of $\hat{\Theta}_s$, has been derived in appendix D. Since only the correction terms are functions of the second stage parameters Θ_2, the above involves taking the derivative of the correction terms with respect to the second stage parameters Θ_2.

REFERENCES

Biørn, Erik. 2004. Regression systems for unbalanced panel data: A stepwise maximum likelihood procedure. *Journal of Econometrics* 122:281–291.

Blundell, Richard, and James Powell. 2003. Endogeneity in nonparametric and semiparametric regression models. In *Advances in Economics and Econonometrics: Theory and Applications, Eighth World Congress, Vol. II*, edited by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge: Cambridge University Press.

Chamberlain, Gary. 1984. Panel data. In *Handbook of Econometrics*, edited by Z. Griliches and M. D. Intriligator, vol. 2. Elsevier.

Greene, William H. 2002. *Econometric Analysis.* 4^{th} ed. Prentice Hall, Upper Saddle River.

Lutkepohl, Helmut. 1996. *Handbook of Matrices.* Chichester: Wiley.

Magnus, Jan R., and Heinz Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Chichester: Wiley.

Newey, Whitney K. 1984. A method of moment interpretation of sequential estimators. *Economics Letters* 14:201–206.

Papke, Leslie E., and Jeffery M. Wooldridge. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145:121–133.

Semykina, Anastasia, and Jefferey M. Wooldridge. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* 157:375–380.

The UNU-MERIT WORKING Paper Series

2013-01 *Effects of innovation on employment in Latin America* by Gustavo Crespi and Ezequiel Tacsir

2013-02 *Revisiting the porter hypothesis: An empirical analysis of green innovation for the Netherlands* George van Leeuwen and Pierre Mohnen

2013-03 *Impact of external knowledge acquisition strategies on innovation - A comparative study based on Dutch and Swiss panel data* by Spyros Arvanitis, Boris Lokshin, Pierre Mohnen and Martin Wörter

2013-04 *Interactive knowledge exchanges under complex social relations: A simulation modelRobin* by Cowan and Anant Kamath

2013-05 *Innovation systems framework: still useful in the new global context?* by Michiko Iizuka

2013-06 *The importance of intrinsic and extrinsic motivation for measuring IQ* by Lex Borghans, Huub Meijers and Bas ter Weel

2013-07 *Firms' innovation capability-building paths and the nature of changes in learning mechanisms: Multiple case-study evidence from an emerging economy* by Paulo N. Figueiredo , Marcela Cohen and Saulo Gomes

2013-08 *A set of time series data labour market stocks and flows for the Netherlands 1980 to 2010* by Manuel Müllers, Joan Muysken and Erik de Regt

2013-09 *Designing an optimal 'tech fix' path to global climate stability: R&D in a multi-phase climate policy framework* by Adriaan van Zon and Paul A. David

2013-10 *Complementarity between internal knowledge creation and external knowledge sourcing in developing countries* by Jun Hou and Pierre Mohnen

2013-11 *Summarizing large spatial datasets: Spatial principal components and spatial canonical correlation* by Samyukta Bhupathiraju, Bart Verspagen and Thomas Ziesemer

2013-12 *Regional systems of innovation in the Arab region* by Samia Satti Osman Mohamed Nour

2013-13 *Development and social justice: Education, training and health in Sudan* by Samia Satti Osman Mohamed Nour

2013-14 *The economic importance and impacts of intellectual property rights (IPRs) in Sudan* by Samia Satti Osman Mohamed Nour

2013-15 *Overview of knowledge economy in the Arab region* by Samia Satti Osman Mohamed Nour

2013-16 *The importance (impacts) of knowledge at the macro-micro levels in the Arab Gulf countries* by Samia Satti Osman Mohamed Nour

2013-17 *Political determinants and impact analysis of using a cable system as a complement to an urban transport system* by Diego Escobar-García, Francisco García-Orozco and Carlos Cadena-Gaitán

2013-18 *Women entrepreneurs in the informal economy: Is formalization the only solution for business sustainability?* By Shyama V. Ramani, Ajay Thutupalli, Tamas Medovarszki, Sutapa Chattopadhyay, Veena Ravichandran

2013-19 *Heterogeneity in innovation strategies, evolving consumer preferences and market structure: An evolutionary multi-agent based modelling approach* by Salih Çevikarslan

2013-20 *Optimal patent length and patent breadth in an R&D driven market with evolving consumer preferences: An evolutionary multi-agent based modelling approach* by Salih Çevikarslan

2013-21 *Innovation and productivity: An update* by Pierre Mohnen and Bronwyn H. Hall

2013-22 *Fathers' use of parental leave. What do we know?* by Nevena Zhelyazkova

2013-23 *Eliciting Illegal migration rates through list randomization* by David McKenzie and Melissa Siegel

2013-24 *How do ICT firms in Turkey manage innovation? Diversity in expertise versus diversity in markets* by Semih Akçomak, Erdal Akdeve and Derya Fındık

2013-25 *Dynamic models of R&D, innovation and productivity: Panel data evidence for Dutch and French manufacturing* by Wladimir Raymond, Jacques Mairesse, Pierre Mohnen and Franz Palm

2013-26 *Centre-based versus home-based childcare* by Robert Bauchmüller

2013-27 *Microeconometric evidence of financing frictions and innovative activity* by Amaresh K Tiwari, Pierre Mohnen, Franz C Palm and Sybrand Schim van der Loeff