

2012s-24

Microeconometric Evidence of Financing Frictions and Innovative Activity

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

Série Scientifique
Scientific Series

Montréal
Septembre 2012

© 2012 Amaresh K. Tiwari, Pierre Mohnen, Franz C. Palm, Sybrand Schim van der Loeff. Tous droits réservés.
All rights reserved. Reproduction partielle permise avec citation du document source, incluant la notice ©.
Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.



Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère du Développement économique et régional et de la Recherche, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaire majeur

Ministère du Développement économique,
de l'Innovation et de l'Exportation

Partenaires corporatifs

Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque Laurentienne du Canada
Banque Nationale du Canada
Banque Royale du Canada
Banque Scotia
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Fédération des caisses Desjardins du Québec
Financière Sun Life, Québec
Gaz Métro
Hydro-Québec
Industrie Canada
Investissements PSP
Ministère des Finances du Québec
Power Corporation du Canada
Rio Tinto Alcan
State Street Global Advisors
Transat A.T.
Ville de Montréal

Partenaires universitaires

École Polytechnique de Montréal
HEC Montréal
McGill University
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

ISSN 1198-8177

Partenaire financier

Développement
économique, Innovation
et Exportation
Québec 

Microeconometric Evidence of Financing Frictions and Innovative Activity*

*Amaresh K. Tiwari[†], Pierre Mohnen[‡], Franz C. Palm[§],
Sybrand Schim van der Loeff^{**}*

Résumé / Abstract

Nous utilisons une base de données de panel néerlandaise assez originale pour examiner comment les décisions d'innovation et de financement varient selon les caractéristiques des entreprises. Nous examinons en particulier dans quelle mesure il y a une faille de marché due aux besoins de financement de l'innovation. En résumé, nous aboutissons aux résultats suivants. Premièrement, les entreprises soumises à des contraintes financières font leurs choix de financement et d'innovation en fonction de leur âge, de leur taille et de leur degré d'endettement. Sans contraintes de financement, les entreprises sont moins portées à innover en s'endettant, quelles que soient leurs caractéristiques. Deuxièmement, ont tendance à être contraintes financièrement les entreprises jeunes, petites, avec un rapport dettes/fonds propres élevé et peu d'avoirs collatéraux. Troisièmement, les entreprises jeunes, grandes et avec un faible rapport dettes/fonds propres ont plus de chances d'être innovantes. Quatrièmement, les contraintes de financement réduisent l'intensité de R-D. Cinquièmement, ce sont les entreprises petites et jeunes qui sont plus intenses en R-D. Pour estimer notre modèle, nous développons un nouvel estimateur qui combine les méthodes des effets aléatoires corrélés et des fonctions de contrôle pour tenir compte de l'endogénéité des régresseurs dans un modèle structurel d'équations simultanées.

Mots clés : financières, structure de capital, R-D, innovation, dynamique de firmes, failles de marché, données panel, effets aléatoires corrélés, fonctions de contrôle, attentes a posteriori.

Using a unique panel data of Dutch innovation and financial variables we empirically investigate how financing and innovation vary across firm characteristics. The study also tries to gauge the extent of market failure due to the presence of financing frictions. Our main findings can be summarized as follows. First, when firms face endogenous financial constraints, debt financing and innovation choices are not independent of firm characteristics such as age, size, and existing leverage. In the absence of financial constraints, however, firms, almost uniformly across firm characteristics, become less inclined – as compared to firms facing constraints - to engage in innovative activity by raising debt. Second, small, young, highly leveraged, and firms with lower collateralizable assets are more likely to be financially constrained. Third, large, young, and low leveraged firms are more likely to be innovators. Fourth, financial constraints adversely affect a firm's R&D intensity. Fifth, smaller and younger firms are more R&D intensive. A new estimator, that combines the method of “Correlated Random Effects” and “Control Function” to account for the endogeneity of regressors in a structural equations model, is developed.

Keywords: Financial Constraints, Capital Structure, R&D, Innovation, Firm Dynamics, Market Failure, Panel Data, Correlated Random Effects, Control Function, Expected a Posteriori.

Codes JEL : G30, O30, C30

* The authors gratefully acknowledge the financial support by The Netherlands Organization for Scientific Research (NWO) under grant 400-03-109. The authors wish to thank Aline Muller, seminar participants at HEC, University of Liege, and at ZEW, Mannheim, for their helpful comments. All remaining errors are our own.

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

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

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

^{**} Maastricht University, S.Loeff@maastrichtuniversity.nl.

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 differently with firm characteristics such as size, age, and leverage for financially constrained 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 constraints 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) model for R&D investment, where we try to assess the impact of financial constraints, as reported by the firms, on R&D investment, (b) a model for financial constraints, where we try to explain why certain firms report they are financially constrained, (b) a model for decision to innovate, where we study the financing choices of innovative firms, and (d) a system of reduced form equations of financing choice and other endogenous variables.

We find that (i) under binding financial constraints the marginal propensity to innovate by employing more long-term debt varies with characteristics such as age, size, and existing leverage, and that (ii) when financial constraints 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 constraints bind. Some of our other important empirical findings related to financing and innovative activity are that (iii) large and young firms are more likely to engage in innovative activity, (iv) that large and mature firms are less R&D intensive, (v) small and younger firms are more financially constrained. These findings suggest that decisions to innovate, financing choices, and firm dynamics are not independent.

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), 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 constraints 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 Klette and Kortum (2004) 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 of financial market inefficiency, such as by Cooley and Quadrini (2001), where financial frictions introduced in a standard model of industry dynamics generate results that match the empirical regularities in the financial and company level investment characteristics of firms that are related to their size and age. Albuquerque and Hopenhayn (2004) (henceforth AH) and Clementi and Hopenhayn (2006) (henceforth

CH) in their respective papers develop models of endogenous borrowing constraints 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 constraints and firm growth and survival. Our results suggests that modeling firm dynamics with R&D and innovation that incorporates borrowing constraints in a dynamic financial contract framework could be an important area of research.

Secondly, our results are in congruence with the papers that provide empirical evidence that R&D intensive firms are less leveraged than those that are not. Brown, Fazzari, and Petersen (2009) (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. We find that, *ceteris paribus*, innovative firms are likely to maintain higher levels of liquidity reserve, and less likely to pay out dividends, which tells us that external financing could be more costly for innovative firms.

Thirdly, in our fully specified model of endogenous financing and innovation choices and endogenous financial constraints 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)) 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 constraints. The finding is in line with prediction made by AH and CH in models of firm growth and survival with endogenous borrowing constraints. 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 constraints.

Fourthly, our paper contributes to the empirical literature that seeks to test for financing friction and quantifying the extent of market failure due to existence of financing friction. 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. Hennessy and Whited (2007) find that existence of costly external funds depresses the path of investment. In a recent paper Hajivassiliou and Savignac (2011), using a similar data set for France, find that financial constraints do adversely affect innovation output. Our aim here is to assess by how much R&D investment is hampered given that a firm faces financial constraints.

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” (CRE) and “control function” (CF) (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 constraints 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 discuss the empirical strategy employed, in Section 3 the data used and the definition of the variables, in Section 4 the results and finally in Section 5 we conclude. In Appendix A we discuss the identification of the structural parameters. The details of the econometric methodology are provided in appendices C, D, E, and F, which for reasons of space we have not included in the core of the paper, but which are available upon request.

2 Empirical Model

To study the effect of endogenous financial constraints on R&D expenditure, the endogenous financing decisions of firms, the endogenous decision to innovate, and to account for the fact that with our data R&D expenditure is known only for firms that opt to innovate, we consider the following system of equations:

$$\begin{aligned} r_{it} &= f_{it}(\beta_f + \mathbf{z}_{it}^r \beta_1 + \mathbf{x}_{it}^r \tilde{\beta}_1 + \mu_1 \tilde{\alpha}_i + \eta_{1it}) + (1 - f_{it})(\mathbf{z}_{it}^r \beta_0 + \mathbf{x}_{it}^r \tilde{\beta}_0 + \mu_0 \tilde{\alpha}_i + \eta_{0it}) \text{ if } s_{it} = 1 \\ &= f_{it} r_{1it} + (1 - f_{it}) r_{0it} \text{ if } s_{it} = 1, \end{aligned} \quad (1)$$

$$f_{it} = 1(f_{it}^* = \mathbf{z}_{it}^{f'} \varphi + \mathbf{x}_{it}^{f'} \tilde{\varphi} + \lambda \tilde{\alpha}_i + \zeta_{it} > 0), \quad (2)$$

$$s_{it} = 1(s_{it}^* = \mathbf{z}_{it}^{s'} \gamma + \mathbf{x}_{it}^{s'} \tilde{\gamma} + \theta \tilde{\alpha}_i + v_{it} > 0), \quad (3)$$

$$\mathbf{x}_{it} = \mathbf{Z}_{it}' \boldsymbol{\delta} + \tilde{\alpha}_i \boldsymbol{\kappa} + \boldsymbol{\epsilon}_{it}, \quad (4)$$

where r_{it} in equation (1) is the ratio of total R&D expenditure to total capital assets (tangible + intangible) of the firm. f_{it} is a binary variable that indicates, with value 1, if the firm reports that it is financially constrained with respect to innovation or R&D activities in period t , and f_{it}^* in (2) is the latent variable underlying f_{it} . We term equation (2) as the financial constraint

equation. Since R&D expenditure is observed only when the firm decides to innovate, that is when $s_{it} = 1$, to rule out possible sample selection bias that could arise because some component (observed or unobserved) of the decision to innovate also determines the amount of R&D expenditure, we specify a selection equation, equation (3), where the idiosyncratic error term is correlated with the idiosyncratic error terms appearing in the R&D and the financial constraint equations.

Conditional on unobserved heterogeneity/individual effect $\tilde{\alpha}_i$, \mathbf{z}_{it}^r is a vector of exogenous variables and \mathbf{x}_{it}^r is a vector of endogenous variables. To allow for the effect of \mathbf{z}_{it}^r and \mathbf{x}_{it}^r to be different in the two regimes, financially constrained and unconstrained, we model equation (1) as an endogenous switching regression model, where the financial constraint equation sorts the firms over the two different regimes. That is, $r_{it} = r_{1it} = \beta_f + \mathbf{z}_{it}^{r'}\beta_1 + \mathbf{x}_{it}^{r'}\tilde{\beta}_1 + \mu_1\tilde{\alpha}_i + \eta_{1it}$ if $f_{it} = 1$ and $s_{it} = 1$ and $r_{it} = r_{0it} = \mathbf{z}_{it}^{r'}\beta_0 + \mathbf{x}_{it}^{r'}\tilde{\beta}_0 + \mu_0\tilde{\alpha}_i + \eta_{0it}$ if $f_{it} = 0$ and $s_{it} = 1$. η_{1it} is the idiosyncratic term of the R&D equation for the firm that is financially constrained, regime 1, and η_{0it} denotes the idiosyncratic term when the firm is not financially constrained, regime 0. We term equation (1) as the R&D equation.

Similarly, conditional on unobserved heterogeneity $\tilde{\alpha}_i$, \mathbf{z}_{it}^f and \mathbf{z}_{it}^s are vectors of exogenous variables in the financial constraint and selection equations respectively. \mathbf{x}_{it}^f and \mathbf{x}_{it}^s are vectors of endogenous variables in equation (2) and (3) respectively. A detailed discussion of the specification for the all the three equations, (1), (2), and (3), is carried out later.

While equations (1), (2), and (3) are our structural equations of interest, equation (4) is the system of ‘ m ’ equations written in a reduced form for the endogenous variables $\mathbf{x}_{it} = (x_{1it}, \dots, x_{mit})'$, where every component of \mathbf{x}_{it}^r , \mathbf{x}_{it}^f , and \mathbf{x}_{it}^s is also a component of \mathbf{x}_{it} . $\mathbf{Z}_{it} = \text{diag}(\mathbf{z}_{1it}, \dots, \mathbf{z}_{mit})$ is the matrix of exogenous variables or instruments appearing in each of the m reduced form equations in (4) and $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \dots, \boldsymbol{\delta}'_m)'$. Let \mathbf{z}_{it} be the union of all exogenous variables appearing in each of \mathbf{z}_{it}^r , \mathbf{z}_{it}^f , and \mathbf{z}_{it}^s . For every $l \in (1, \dots, m)$, $\mathbf{z}_{lit} = \mathcal{Z}_{it} = (\mathbf{z}'_{lit}, \tilde{\mathbf{z}}'_{lit})'$, where the dimension of vector of instruments, $\tilde{\mathbf{z}}$, is greater than or equal to the dimension \mathbf{x} . Also define $\mathbf{X}_i = \{\mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT_i}\}'$ and $\mathcal{Z}_i = (\mathcal{Z}'_{i1} \dots \mathcal{Z}'_{iT_i})'$.

$\boldsymbol{\epsilon}_{it} = (\epsilon_{1it}, \dots, \epsilon_{mit})'$ is the vector of idiosyncratic component and $\tilde{\alpha}_i$ is the scalar unobserved individual effect for firm i , which we model as a random effect, is correlated with \mathcal{Z}_i , but is assumed to be mean independent of η_{it} , ζ_{it} , v_{it} , and $\boldsymbol{\epsilon}_{it}$. Also, conditional on $\tilde{\alpha}_i$, \mathcal{Z}_i is independent of η_{it} , ζ_{it} , v_{it} , and $\boldsymbol{\epsilon}_{it}$. Since the unobserved individual specific effect affects the endogenous regressors as well as the firm’s decision to innovate, it being financially constrained, and how much it spends on R&D, therefore we have different factor loadings, such as, $\{\kappa_1 \dots, \kappa_m\}$, that appear in the reduced form equations, and μ_1 , μ_0 , λ , and θ that appear in the structural equations (1), (2), and (3).

The above structural equations – (1), (2), and (3) – can be succinctly written as

$$\mathbf{y}_{it}^* = \mathbb{X}'_{it}\mathbf{B} + \tilde{\alpha}_i\mathbf{k} + \Upsilon_{it}, \quad (5)$$

where \mathbf{y}_{it}^* , suppressing the it subscript, is $\mathbf{y}^* = \{sf r_1, s(1-f)r_0, f^*, s^*\}'$, $\mathbb{X}_{it} = \text{diag}(\mathcal{X}'_{1it}, \mathcal{X}'_{0it}, \mathcal{X}'_{it}^f, \mathcal{X}'_{it}^s)$, where $\mathcal{X}'_{1it} = s_{it}f_{it}\{1, \mathbf{z}'_{it}, \mathbf{x}'_{it}\}'$, $\mathcal{X}'_{0it} = s_{it}(1-f_{it})\{\mathbf{z}'_{it}, \mathbf{x}'_{it}\}'$, $\mathcal{X}'_{it}^f = \{\mathbf{z}'_{it}, \mathbf{x}'_{it}\}'$, and $\mathcal{X}'_{it}^s = \{\mathbf{z}'_{it}, \mathbf{x}'_{it}\}'$. \mathbf{B} in (5) is given by $\mathbf{B} = \{\beta_f, \boldsymbol{\beta}'_1, \boldsymbol{\beta}'_0, \boldsymbol{\varphi}', \boldsymbol{\gamma}'\}'$, where $\boldsymbol{\beta}_1 = \{\beta'_1, \beta'_1\}$, $\boldsymbol{\beta}_0 = \{\beta'_0, \beta'_0\}$, $\boldsymbol{\varphi} = \{\varphi', \tilde{\varphi}'\}'$, and $\boldsymbol{\gamma} = \{\gamma', \tilde{\gamma}'\}'$. Finally, $\mathbf{k} = \{\mu_1, \mu_0, \lambda, \theta\}'$, and $\Upsilon_{it} = \{\eta_{1it}, \eta_{0it}, \zeta_{it}, v_{it}\}'$. The distribution of the error terms of the system of equations (1), (2), (3), and (4) is given by:

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

To estimate the structural equations in (5) of the above model we combine the method of “correlated random effect” and the “control function” approaches in a sequential estimation procedure. Our constructed control functions allow us to take account of the fact that the regressors are correlated with unobserved heterogeneity as well as the idiosyncratic component. The identification strategy that allows us to construct the control functions is based on the following conditional mean restriction:

$$\begin{aligned} E(\tilde{\alpha}_i \mathbf{k} + \Upsilon_{it} | \mathbf{X}_i, \mathcal{Z}_i, \tilde{\alpha}_i) &= \tilde{\alpha}_i \mathbf{k} + E(\Upsilon_{it} | \mathcal{Z}_i, \mathbf{X}_i, \tilde{\alpha}_i) = \tilde{\alpha}_i \mathbf{k} + E(\Upsilon_{it} | \mathcal{Z}_i, \boldsymbol{\epsilon}_i, \tilde{\alpha}_i) \\ &= \tilde{\alpha}_i \mathbf{k} + E(\Upsilon_{it} | \boldsymbol{\epsilon}_i, \tilde{\alpha}_i) = \tilde{\alpha}_i \mathbf{k} + E(\Upsilon_{it} | \boldsymbol{\epsilon}_i). \end{aligned} \quad (6)$$

According to the above, the mean dependence of the structural error term Υ_{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 $\boldsymbol{\epsilon}_i$. The expectation of Υ_{it} given $\boldsymbol{\epsilon}_i$ is given by

$$E(\Upsilon_{it} | \boldsymbol{\epsilon}_i) = E(\Upsilon_{it} | \boldsymbol{\epsilon}_{it}) = \Sigma_{\Upsilon\epsilon} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} = \tilde{\Sigma}_{\Upsilon\epsilon} \Sigma_{\epsilon\epsilon} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} = \tilde{\Sigma}_{\Upsilon\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it}, \quad (7)$$

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

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

and the $(m \times m)$ matrix Σ_{ϵ} is $\text{diag}(\sigma_{\epsilon_1}, \dots, \sigma_{\epsilon_m})$, 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 $E(\Upsilon_{it} | \boldsymbol{\epsilon}_{it}) = \tilde{\Sigma}_{\Upsilon\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it}$ since 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 (7) helps us distinguish the parameters that are estimated in the first stage as compared to 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}_{it} with respect to Υ_{it} .

Given (7) we can write the linear projections of $\tilde{\alpha}_i \mathbf{k} + \Upsilon_{it}$ in error form, given $\tilde{\alpha}_i$ and $\boldsymbol{\epsilon}_{it}$ as

$$\tilde{\alpha}_i \mathbf{k} + \Upsilon_{it} = \tilde{\alpha}_i \mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} + \tilde{\Upsilon}_{it} \quad (8)$$

where $\tilde{\Upsilon}_{it}$ is independent of \mathcal{Z}_i , \mathbf{X}_i , $\boldsymbol{\epsilon}_{it}$, and $\tilde{\alpha}_i$, and is normally distributed with expectation 0. The above then implies that the projections of \mathbf{y}_{it}^* in error form given $\tilde{\alpha}_i$ and $\boldsymbol{\epsilon}_{it}$ is given by

$$\mathbf{y}_{it}^* = \mathbb{X}_{it}' \mathbf{B} + \tilde{\alpha}_i \mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} + \tilde{\Upsilon}_{it}. \quad (9)$$

The structural parameters of interest in (9) are estimated sequentially. In the first step, the system of reduced form equations, equation (4), is estimated.

2.1 Estimation of the First Stage Reduced Form Equations

In the first stage we estimate the system of reduced form equations (4). Since $\tilde{\alpha}_i$ and \mathcal{Z}_i are correlated in order to estimate $\boldsymbol{\delta}$, $\Sigma_{\epsilon\epsilon}$, and σ_α consistently, we use Mundlak's (1978) correlated random effects formulation. We assume that

$$\tilde{\alpha}_i = \bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \alpha_i, \quad (10)$$

where $\bar{\mathcal{Z}}_i$ is the mean of time-varying variables in \mathcal{Z}_{it} . We also assume that α_i is normally distributed with mean 0 and variance σ_α^2 . Given the above, equation (4) can now be written as

$$\mathbf{x}_{it} = \mathbf{Z}'_{it} \boldsymbol{\delta} + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \alpha_i) \boldsymbol{\kappa} + \boldsymbol{\epsilon}_{it}. \quad (4a)$$

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

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

Moreover, as can be seen from the modified equation (4a), we also impose the restriction that $\bar{\boldsymbol{\delta}}$ remains the same across each of the m reduced form equations. In Appendix C we provide a note on the estimation strategy employed to estimate the parameters of the reduced form equations.

2.2 Estimation of the Structural Equations

Given equation (9) and (10), we can write the linear predictor of \mathbf{y}_{it}^* given \mathbf{X}_i , \mathcal{Z}_i , and α_i in error form as

$$\mathbf{y}_{it}^* = \mathbb{X}'_{it} \mathbf{B} + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \alpha_i) \mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it} + \tilde{\Upsilon}_{it}, \quad (11)$$

To estimate the system of equations in (11) the standard technique is to replace $\boldsymbol{\epsilon}_{it}$ by the residuals from the first stage reduced form regression, here equation (4a). However, the residuals $\mathbf{x}_{it} - \mathbb{E}(\mathbf{x}_{it} | \mathcal{Z}_i, \alpha_i) = \mathbf{x}_{it} - \mathbf{Z}'_{it} \boldsymbol{\delta} - (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \alpha_i) \boldsymbol{\kappa}$, remain unidentified because the α_i 's are unobserved even though the reduced form parameters, $\boldsymbol{\delta}$, $\bar{\boldsymbol{\delta}}$, and $\boldsymbol{\kappa}$, 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

$$\mathbb{E}(\mathbf{y}_{it}^* | \mathbf{X}_i, \mathcal{Z}_i) = \int \mathbb{E}(\mathbf{y}_{it}^* | \mathbf{X}_i, \mathcal{Z}_i, \alpha_i) f(\alpha_i | \mathbf{X}_i, \mathcal{Z}_i) d\alpha_i = \mathbb{X}'_{it} \mathbf{B} + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}, \quad (12)$$

where $\hat{\alpha}_i = \mathbb{E}(\alpha_i | \mathbf{X}_i, \mathcal{Z}_i)$, as discussed in Appendix A, is the Expected a Posteriori (EAP) value of α_i and $\hat{\boldsymbol{\epsilon}}_{it} = \mathbf{x}_{it} - \mathbb{E}(\mathbf{x}_{it} | \mathbf{X}_i, \mathcal{Z}_i) = \mathbf{x}_{it} - \mathbf{Z}'_{it} \boldsymbol{\delta} - \boldsymbol{\kappa} (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i)$. $\hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$ are the ‘‘control

functions” that correct for the bias when estimating the structural equations, which arises due to the correlation of \mathbf{x}_{it} with the unobserved individual effect α_i and with the idiosyncratic component Υ_{it} . The correlation of the exogenous variables \mathcal{Z}_{it} with the unobserved individual effect, $\tilde{\alpha}_i$, is accounted by $\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i$. In Appendix A we show how to construct the control functions $\hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$. Given (12) we can write the projection of \mathbf{y}_{it}^* given $\mathbf{X}_i, \mathcal{Z}_i$ in error form as

$$r_{it} = f_{it}(\mathcal{X}_{it}^{r'} \boldsymbol{\beta}_1 + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \mu_1 + \tilde{\Sigma}_{\eta_1 \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it} + \tilde{\eta}_{1it}) \\ + (1 - f_{it})(\mathcal{X}_{it}^{r'} \boldsymbol{\beta}_0 + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \mu_0 + \tilde{\Sigma}_{\eta_0 \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it} + \tilde{\eta}_{0it}) \text{ if } s_{it} = 1, \quad (13)$$

$$f_{it}^* = \mathcal{X}_{it}^{f'} \boldsymbol{\varphi} + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \lambda + \tilde{\Sigma}_{\zeta \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it} + \tilde{\zeta}_{it}, \quad (14)$$

$$s_{it}^* = \mathcal{X}_{it}^{s'} \boldsymbol{\gamma} + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \theta + \tilde{\Sigma}_{v \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it} + \tilde{v}_{it}, \quad (15)$$

where $\tilde{\Upsilon}_{it} = \{\tilde{\eta}_{1it}, \tilde{\eta}_{0it}, \tilde{\zeta}_{it}, \tilde{v}_{it}\}'$, defined in Appendix A, is normally distributed with mean 0, variance $\Sigma_{\tilde{\Upsilon} \tilde{\Upsilon}}$ and is independent of \mathbf{X}_i and \mathcal{Z}_i .

Given $\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$, while it may be possible to consistently estimate the structural parameters of interest by specifying a joint likelihood for r_{1it}, r_{0it}, f_{it} , and s_{it} , 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 (13) to (15) in two steps after the first stage reduced form estimation. In the second stage, given the estimates of the control functions $\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$, we estimate jointly the structural parameters, Θ_2 , of the financial constraint equation (14) and the innovation/selection equation (15).

2.2.1 The Second Stage: Estimation of the Financial Constraint and the Innovation Selection Equations

Given the financial constraint equation (14) and the innovation selection equation (15), the conditional log likelihood function for individual i in period t given $\mathbf{X}_i, \mathcal{Z}_i$, if the time period t corresponds to CIS3 and CIS3.5¹, is given by

$$\mathcal{L}_{it2}(\Theta_2 | \mathbf{X}_i, \mathcal{Z}_i, \hat{\alpha}_i, \hat{\boldsymbol{\epsilon}}_{it}) = f_{it} s_{it} \ln(\Phi_2(\varphi_{it}, \gamma_{it}, \rho_{\tilde{\zeta} \tilde{v}})) + (1 - f_{it}) s_{it} \ln(\Phi_2(-\varphi_{it}, \gamma_{it}, -\rho_{\tilde{\zeta} \tilde{v}})) \\ + f_{it} (1 - s_{it}) \ln(\Phi_2(\varphi_{it}, -\gamma_{it}, -\rho_{\tilde{\zeta} \tilde{v}})) + (1 - f_{it}) (1 - s_{it}) \ln(\Phi_2(-\varphi_{it}, -\gamma_{it}, \rho_{\tilde{\zeta} \tilde{v}})), \quad (16)$$

where $\Theta_2 = \{\boldsymbol{\varphi}', \lambda, \tilde{\Sigma}_{\zeta \epsilon}, \boldsymbol{\gamma}', \theta, \tilde{\Sigma}_{v \epsilon}, \rho_{\tilde{\zeta} \tilde{v}}\}'$, Φ_2 is the cumulative distribution function of a standard bivariate normal, $\rho_{\tilde{\zeta} \tilde{v}}$ is the correlation of $\tilde{\zeta}_{it}$ and \tilde{v}_{it} , and

$$\varphi_{it} = (\mathcal{X}_{it}^{f'} \boldsymbol{\varphi} + \lambda(\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{\zeta \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}) \frac{1}{\sigma_{\tilde{\zeta}}}, \quad (17)$$

$$\gamma_{it} = (\mathcal{X}_{it}^{s'} \boldsymbol{\gamma} + \theta(\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{v \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}) \frac{1}{\sigma_{\tilde{v}}}. \quad (18)$$

It should be noted that all the parameters of the structural equations (14) and (15) can only be identified up to a scale, the scaling factor for the financial constraint equation and selection

¹As discussed in next section, where we discuss the data, the empirical analysis is done using 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.

equation being respectively $\sigma_{\tilde{\zeta}}$ and $\sigma_{\tilde{v}}$. In what follows, with a slight abuse of notation, we will denote the scaled parameters of the second stage estimation by their original notation.

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

$$\begin{aligned} \mathcal{L}_{it2}(\Theta_2|\mathbf{X}_i, \mathcal{Z}_i, \hat{\alpha}_i, \hat{\epsilon}_{it}) = \\ f_{it}s_{it} \ln(\Phi_2(\varphi_{it}, \gamma_{it}, \rho_{\tilde{\zeta}\tilde{v}})) + (1 - f_{it})s_{it} \ln(\Phi_2(-\varphi_{it}, \gamma_{it}, -\rho_{\tilde{\zeta}\tilde{v}})) + (1 - s_{it}) \ln(\Phi(-\gamma_{it})), \end{aligned} \quad (19)$$

where $\Phi(-\gamma_{it}) = \Pr(s_{it} = 0|\mathbf{X}_i, \mathcal{Z}_i, \hat{\alpha}_i, \hat{\epsilon}_{it})$. The log likelihood of the second stage parameters is given by

$$\mathcal{L}_2() = \sum_{i=1}^N \sum_{t=1}^{T_i} \mathcal{L}_{it2}(). \quad (20)$$

Given the first stage estimates $\hat{\Theta}_1$, we can obtain the estimates of $\tilde{\mathcal{Z}}_i'\tilde{\delta} + \hat{\alpha}_i$ and $\hat{\epsilon}_{it}$, as shown in Appendix A, which can then be used in the above likelihood function to obtain consistent estimates for the second stage parameters. The true measure of the effect of a certain covariate 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 introduce the concept of APE and discuss tests for it in Appendix E.

2.2.2 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 (13) 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, one is to be financially constrained or not, and the other one is the choice to be innovative or not, determining whether R&D expenditure need to be declared or not. To consistently estimate the parameters of equation (13), 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 $\tilde{\mathcal{Z}}_i'\tilde{\delta} + \hat{\alpha}_i$ and $\hat{\epsilon}_{it}$, for each observation, we obtain consistent estimates of Θ_3 .

To this effect, consider the following conditional mean:

$$\begin{aligned} \mathbb{E}(r_{it}|f_{it}^*, s_{it}^* > 0, \mathbf{X}_i, \mathcal{Z}_i) = f_{it} \left(\beta_f + \mathcal{X}_{it}^{r'}\boldsymbol{\beta}_1 + \mu_1(\tilde{\mathcal{Z}}_i'\tilde{\delta} + \hat{\alpha}_i) + \tilde{\Sigma}_{\eta_1\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\epsilon}_{it} \right) \\ + (1 - f_{it}) \left(\mathcal{X}_{it}^{r'}\boldsymbol{\beta}_0 + \mu_0(\tilde{\mathcal{Z}}_i'\tilde{\delta} + \hat{\alpha}_i) + \tilde{\Sigma}_{\eta_0\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\epsilon}_{it} \right) \\ + f_{it}\mathbb{E}(\tilde{\eta}_{1it}|f_{it}^* > 0, s_{it}^* > 0, \mathbf{X}_i, \mathcal{Z}_i) + (1 - f_{it})\mathbb{E}(\tilde{\eta}_{0it}|f_{it}^* \leq 0, s_{it}^* > 0, \mathbf{X}_i, \mathcal{Z}_i) \end{aligned} \quad (21)$$

Now, we know that

$$\mathbb{E}(\tilde{\eta}_{1it}|f_{it}^* > 0, s_{it}^* > 0, \mathbf{X}_i, \mathcal{Z}_i) = \mathbb{E}[\tilde{\eta}_{1it}|\tilde{\zeta}_{it} > -\varphi_{it}, \tilde{v}_{it} > -\gamma_{it}],$$

and

$$\mathbb{E}(\tilde{\eta}_{0it}|f_{it}^* \leq 0, s_{it}^* > 0, \mathbf{X}_i, \mathcal{Z}_i) = \mathbb{E}[\tilde{\eta}_{0it}|\tilde{\zeta}_{it} \leq -\varphi_{it}, \tilde{v}_{it} > -\gamma_{it}],$$

where φ_{it} and γ_{it} are defined, respectively, in (17) and (18). In Appendix D we show that for an individual i

$$\begin{aligned} E[\tilde{\eta}_{1t} | \tilde{\zeta}_t > -\varphi_t, \tilde{v}_t > -\gamma_t] &= \sigma_{\tilde{\eta}_1} \rho_{\tilde{\eta}_1 \tilde{\zeta}} \phi(\varphi_t) \frac{\Phi\left(\frac{(\gamma_t - \rho_{\tilde{\zeta}\tilde{v}} \varphi_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{v}})} \\ &+ \sigma_{\tilde{\eta}_1} \rho_{\tilde{\eta}_1 \tilde{v}} \phi(\gamma_t) \frac{\Phi\left(\frac{(\varphi_t - \rho_{\tilde{\zeta}\tilde{v}} \gamma_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{v}})} \end{aligned} \quad (22)$$

and

$$\begin{aligned} E[\tilde{\eta}_{0t} | \tilde{\zeta}_t \leq -\varphi_t, \tilde{v}_t > -\gamma_t] &= -\sigma_{\tilde{\eta}_0} \rho_{\tilde{\eta}_0 \tilde{\zeta}} \phi(\varphi_t) \frac{\Phi\left(\frac{(\gamma_t - \rho_{\tilde{\zeta}\tilde{v}} \varphi_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(-\varphi_t, \gamma_t, -\rho_{\tilde{\zeta}\tilde{v}})} \\ &+ \sigma_{\tilde{\eta}_0} \rho_{\tilde{\eta}_0 \tilde{v}} \phi(\gamma_t) \frac{\Phi\left(\frac{(-\varphi_t + \rho_{\tilde{\zeta}\tilde{v}} \gamma_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(-\varphi_t, \gamma_t, -\rho_{\tilde{\zeta}\tilde{v}})}, \end{aligned} \quad (23)$$

where ϕ , Φ , and Φ_2 , respectively denote the density function of a standard normal distribution, the cumulative distribution function of a standard normal, and the cumulative distribution function of a standard bivariate normal. For any individual i , define the following

$$\begin{aligned} C_{11t} &\equiv \phi(\varphi_t) \frac{\Phi\left(\frac{(\gamma_t - \rho_{\tilde{\zeta}\tilde{v}} \varphi_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{v}})}, & C_{12t} &\equiv \phi(\gamma_t) \frac{\Phi\left(\frac{(\varphi_t - \rho_{\tilde{\zeta}\tilde{v}} \gamma_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(\varphi_t, \gamma_t, \rho_{\tilde{\zeta}\tilde{v}})}, \\ C_{01t} &\equiv -\phi(\varphi_t) \frac{\Phi\left(\frac{(\gamma_t - \rho_{\tilde{\zeta}\tilde{v}} \varphi_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(-\varphi_t, \gamma_t, -\rho_{\tilde{\zeta}\tilde{v}})}, & \text{and } C_{02t} &\equiv \phi(\gamma_t) \frac{\Phi\left(\frac{(-\varphi_t + \rho_{\tilde{\zeta}\tilde{v}} \gamma_t) / \sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}{\sqrt{1 - \rho_{\tilde{\zeta}\tilde{v}}^2}}\right)}{\Phi_2(-\varphi_t, \gamma_t, -\rho_{\tilde{\zeta}\tilde{v}})}. \end{aligned}$$

Given estimates of $\hat{\alpha}_i$ and $\hat{\epsilon}_{it}$, φ_{it} , γ_{it} , and $\rho_{\tilde{\zeta}\tilde{v}}$ we can construct the above control functions. With the above defined, we can now write the R&D switching equations in (14), conditional on f_{it}^* , $s_{it}^* > 0$, \mathbf{X}_i , \mathbf{Z}_i as

$$\begin{aligned} r_{it} &= f_{it} \left(\beta_f + \mathcal{X}_{it}^{r'} \beta_1 + \mu_1(\bar{\mathbf{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{\eta_1 \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\epsilon}_{it} + \sigma_{\tilde{\eta}_1} \rho_{\tilde{\eta}_1 \tilde{\zeta}} C_{11it} + \sigma_{\tilde{\eta}_1} \rho_{\tilde{\eta}_1 \tilde{v}} C_{12it} + \underline{\eta}_{1it} \right) \\ &+ (1 - f_{it}) \left(\mathcal{X}_{it}^{r'} \beta_0 + \mu_0(\bar{\mathbf{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{\eta_0 \epsilon} \tilde{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\epsilon}_{it} + \sigma_{\tilde{\eta}_0} \rho_{\tilde{\eta}_0 \tilde{\zeta}} C_{01it} + \sigma_{\tilde{\eta}_0} \rho_{\tilde{\eta}_0 \tilde{v}} C_{02it} + \underline{\eta}_{0it} \right), \end{aligned} \quad (24)$$

where $\underline{\eta}_{1it}$ and $\underline{\eta}_{0it}$ conditional on f_{it}^* , s_{it}^* , \mathbf{X}_i , \mathbf{Z}_i 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.

Estimating the parameters of the second and third 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 F we provide analytical expression for the error adjusted covariance matrix for the estimates of the structural parameters.

To measure the magnitude by which R&D intensity is affected due to the presence of financial constraints we have to compute the average partial effect (APE) of f_{it} . For an individual, i , in time period, t , given $\mathcal{X}_{it} = \bar{\mathcal{X}}$, where \mathcal{X}_{it} is the union of elements appearing in \mathcal{X}_{it}^r , \mathcal{X}_{it}^f , and \mathcal{X}_{it}^s , 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 $\hat{\alpha}$ and $\hat{\epsilon}$. The APE of financial constraint on R&D expenditure, conditional on $s_{it} = 1$, is given by

$$\begin{aligned} \Delta_f E(r_{it}|\bar{\mathcal{X}}) &= \int E(r_{1it}|\bar{\mathcal{X}}, f_{it} = 1, s_{it} = 1, \hat{\alpha}, \hat{\epsilon}) dG(\hat{\alpha}, \hat{\epsilon}) \\ &\quad - \int E(r_{0it}|\bar{\mathcal{X}}, f_{it} = 0, s_{it} = 1, \hat{\alpha}, \hat{\epsilon}) dG(\hat{\alpha}, \hat{\epsilon}). \end{aligned} \quad (25)$$

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

3 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 $s = 1, 2, \dots, S$. 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 $\Theta_s = \{1, 2, \dots, i, \dots, i_s\}$ 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

$$\hat{x}_f = \sum_{s=1}^S (N_{fs} - n_{fs}) \bar{x}_s + \sum_{s=1}^S \sum_{k=1}^{n_{fs}} x_{fsk}, \quad (26)$$

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 ². Table 1 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.

Table 1: **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.

The sample of firms used in the estimation are only those for which at least one R&D performing enterprise is present in the innovation surveys. Enterprises belonging to firms not present in the SF data had to be dropped. N_f was obtained from the Frame Population constructed by the CBS and n_f from the CIS surveys. The percentage of firms in the sample for which imputation, using equation (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 other firms happen to be single enterprises: 78.97, 74.01, and 73.87 respectively for CIS2.5, CIS3, and CIS3.5. The exact count of firms for which $N_f = n_f$ or for which, $(N_f - n_f) > 0$ can be found in Table B1 in Appendix B.

The two variables of interest 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 as 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

²Proof:

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 (26), is $\sum_{s=1}^S n_{fs} \mu_{xs}$. Taking expectations in (26) and substituting the expected value of $E(\sum_{s=1}^S \sum_{k=1}^{n_{fs}} x_{fsk}) = \sum_{s=1}^S n_{fs} \mu_{xs}$ and noting that $E(\sum_{s=1}^S n_{fs} \bar{x}_s) = \sum_{s=1}^S n_{fs} \mu_{xs}$, we get $E(\hat{x}_f) = \mu_{xf} = \sum_{s=1}^S N_{fs} \mu_{xs}$.

the definition criteria for an innovator are exhaustive, we can reasonably assume 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, based on the CIS survey, is classified as an innovating stratum.

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 an enterprise belongs to is available from the General Business Register for every year.

In Table 2 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.

Table 2: Innovating/Non-Innovating and Financially Constrained/Unconstrained Firms

CIS2.5 (1996-98)			
	Financially Constrained	Financially Unconstrained	Total
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.

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_{it} : R&D intensity defined as the ratio of R&D expenditure to total (tangible+ intangible) capital assets.
- (2) f_{it} : Binary variable equal to one if the firm is financially constrained.
- (3) s_{it} : Binary variable equal to one if the firm is an innovator.
- (4) $DEBT_{it}$: 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_{it} : Liquidity reserve including cash, bills of exchange, cheques, deposit accounts, current accounts, and other short-term receivables.
- (6) DIV_{it} : Dividend payments to shareholders, group companies, and cooperative societies.
- (7) $SIZE_{it}$: Logarithm of the number of people employed.
- (8) $RAINT_{it}$: Ratio of intangible assets to total (tangible+ intangible) capital assets.
- (9) $SINS_{it}$: Share of sales in the total sales of the firm which is due to newly introduced products.
- (10) CF_{it} : Cashflow defined as operating profit after tax, interest payment, and preference dividend plus the provision for depreciation of assets.
- (11) $MKSH_{it}$: Market share defined as the ratio of firms sales to the total industry sales.
- (12) $DNFC_{it}$: Dummy variable that takes value one for negative realization of cashflows
- (13) $DMULTI_{it}$: Dummy that takes value one if a firm has multiple enterprises.
- (14) AGE_{it} : Age of the firm³.
- (15) Industry dummies.
- (16) Year dummies.

To minimize heteroscedasticity we scale long-term debt ($DEBT_{it}$), cashflows (CF_{it}), liquidity reserve (LQ_{it}), and dividend payout DIV_{it} by total capital assets. Henceforth whenever we refer to these variables it would mean the scaled value of these variables. The lagged values of some of these variables are used as instruments.

³We 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.

Our set of endogenous regressors, \mathbf{x}_{it} , that appear in the structural equations are (1) long-term debt ($DEBT_{it}$) (2) liquidity reserve (LQ_{it}), (3) dividend payout (DIV_{it}), (4) the logarithm of the number of people employed ($SIZE_{it}$), (5) ratio of intangible assets to total assets ($RAINT_{it}$), (6) the share of sales in the total sales of the firm ($SINS_{it}$). $SINS_{it}$ is only observed for innovators. For the purpose of estimating the reduced form equation we assume that $SINS$ 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.

The vector of exogenous – exogenous conditional on unobserved heterogeneity $\tilde{\alpha}_i$ – variables, \mathbf{z}_{it} , that appear in the structural and reduced form equation, and which are included in \mathcal{Z}_{it} are: (1) cashflows of the firm (CF_{it}), (2) a dummy for negative realization of cashflows ($DNFC_{it}$), (3) the market share of the firm ($MKSH_{it}$), (4) the age of the firm (AGE_{it}), (5) a dummy that takes value 1 if the firm consists of multiple enterprises ($DMULTI_{it}$), (6) industry dummies, and (7) year dummies.

Our additional set of instruments, $\tilde{\mathbf{z}}_{it}$, that are also part of \mathcal{Z}_{it} are: (1) cashflows in period $t-1$ ($CF_{i,t-1}$), (2) market share in period $t-1$ ($MKSH_{i,t-1}$), (3) dummy for negative cashflows ($DNFC_{i,t-1}$), (4) square of cashflows in period t (CF_{it}^2), (5) square of cashflows in period $t-1$ ($CF_{i,t-1}^2$), (6) a dummy that takes value 1 if the firm consists of multiple enterprises in period $t-1$ ($DMULTI_{i,t-1}$), and (7) dummy if the firm existed prior to 1967 ($DAGE_{it}$).

Here we would like to point out that variables included in \mathcal{Z}_{it} may or may not be strictly exogenous, but are very likely to be exogenous conditional on unobserved individual effects. For example, cashflows of a firm will most likely have a component that is correlated with firm characteristics, but is also driven to a large extent by exogenous events. To the extent that we take into account the correlation between \mathcal{Z}_{it} and $\tilde{\alpha}_i$, the presence of these variables in the specification of the structural equations will not lead to any inconsistency of the results.

[Table 3 about here]

4 Results

4.1 Financial constraints

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 4 and Table 5. While Table 4 has the coefficient estimates, in Table 5 the Average Partial Effects (APE) of the covariates are reported. In Specification 2 and Specification 3 in Table 4 and 5 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.

[Table 4 about here]

[Table 5 about here]

Let us first discuss the specification and the results of the financial constraint equation. A firm may be constrained both because of high cost of external funds and/or because of high need for external funds, see Hennessy and Whited (2007). Accordingly, we interpret the latent

variable f_{it}^* in equation (2) as reflecting both the premium on scarce external finance and the inability to access external funds. The premium, for example, could reflect bankruptcy cost or the cost of floating equity as in Hennessy and Whited (2007) and Gale and Hellwig (1985), or could reflect higher repayment schedule to lenders as compared to profits during such time as when firms face borrowing constraints and short-term capital advancement are low as in AH and CH. We hypothesize that the premium on external finance and the gap in financing are a function of observable variables, as in Whited and Wu (2005) and Gomes, Yaron, and Zhang (2006), where the shadow price of scarce external finance in the firm’s optimization problem is assumed to be a function of observable variables. Our specification for the financial constraint equation is rich enough to capture both aspects, high cost as well as high need for external finance⁴.

In the specification for the financial constraint equation we use cashflows (CF) and share of innovative sales in the total sales of the firm ($SINS$) as proxies to control for future expected profitability for R&D investment. Ceteris paribus, higher expectation of profits will drive up the demand for R&D investment and cause the firm to report itself being financially constrained. Since we do not have any information on the market valuation of the firms, we cannot construct an average “ q ” for our firms. In any case, Moyen (2004) finds that Tobin’s “ q ” is a poor proxy for investment opportunities, while cashflow is an excellent proxy, positively related to the income shock. Hence, we expect cashflows (CF) to signal investment opportunities and expect it to have a positive sign in the financial constraint equation. We use the share of innovative sales in the total sales of the firm ($SINS$) since the amount of $SINS$ can signal demand for R&D related activity within a firm. We note here that while CF , which is largely driven by exogenous shocks and is exogenous conditional on $\tilde{\alpha}_i$, $SINS$ is an outcome of current and past R&D efforts. Therefore we endogenise $SINS$. We find that both CF and $SINS$ have a significant positive sign in the financial constraint equation. This suggests that both cashflows 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.

In our specification we also have a dummy for negative cashflows ($DNCF$), which is found to have a positive and significant coefficient. It seems that variations over time from negative to positive in the cashflows are more indicative of positive “shifts” in the supply of internal equity finance that relax financial constraints than variation in cashflows 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 constraints. 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 Table 5, 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.

⁴ The paper by Hennessy and Whited (2007) has a detailed discussion on constraint proxies that reflect high cost or high need for external finance.

With higher liquidity reserve firms can counter bad shocks by draining it. 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, if only to avoid laying off their knowledge workers. This implies that when a firm is not sure about a steady supply of positive cashflow it is likely to practice precautionary savings to reduce its risks of being financially constrained during periods of bad shocks.

Our results suggest that dividends (*DIV*) paying firms are less likely to be financially constrained. Hennessy and Whited (2007) also find low dividend paying firms face high costs of external funds. Besides, AH and CH show that when a firm faces borrowing constraints 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 5, is very high, our results too 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. Hennessy and Whited (2007) 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 constraints. Besides, 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 constraints. Since secondary markets for intangible asset is fraught with more frictions and generally does not exist, hence 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 3, 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 4 we find 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.

4.2 Innovation

In the innovation/selection equation we do not include cashflows and the dummy for negative cashflows since cashflows do not have any bearing on the decision to innovate, except maybe through its effect on financial constraints, and we have included both $DNCF$ and CF in the financial constraint equation. We do not include the share of innovative sales in total sales because it is observed only for innovators.

We find that firms with higher long-term debt, $DEBT$, in their capital structure are less likely to take up innovative activity. This is consistent with the findings of BFP and others who find that equity rather than debt may be more suitable to finance innovative activity. Holmstrom (1989) points out that 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.

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.

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 equation⁵. This suggests that firms that pay out dividends are less likely to innovate.

⁵As stated earlier, we assumed $SINS_{it}$ to be zero for the non-innovators when estimating the system of reduced form equation. Therefore, like $SINS_{it}$ the correction term for $SINS_{it}$ will highly correlated with s_{it} , the decision to innovate. Hence, the significance of correction term/control function for Share of Innovative Sales in Specification 1 and Specification 2 should not come as a surprise.

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 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 constraints 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) and 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 a firm is more incited to innovate if it enjoys a monopoly position to prevent entry of potential rivals, 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. This was to be expected since firms that engage in innovative activity have more intangible assets in their asset base. But also, 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 enterprises 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.

The factor loadings, λ and θ , which are the coefficients of $\bar{Z}'_i\bar{\delta} + \hat{\alpha}_i$ in the financial constraint and the selection equation respectively are significant. This, and the facts that control functions pertaining various endogenous regressors to correct for the bias in the structural equations, and $\rho_{\zeta v}$, the correlation between the idiosyncratic component of the financial constraint equation and the innovation equation, are all significant, suggest simultaneity in the decision to innovate, the financing choices made, and the financial constraints faced.

Our analysis allows us to examine how the average partial effect of certain exploratory variables depends on the levels of other variables. In Figure 1 we plot the APE of leverage on the propensity to innovate conditional on being financially *constrained* and conditional being financially *unconstrained*. We plot the APE of leverage 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 leverage on innovation to be negative and almost constant over the distribution of size, age and leverage. In contrast, the APE of leverage 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 leverage on innovation conditional on not being financially constrained. This indicates that under no financial constraints innovative firms, regardless of size, maturity, and existing level of debt, would almost *uniformly* be less inclined to finance themselves with debt. In other words, when borrowing constraints 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 to do it for innovative activity or to be an innovating firm in the first place. 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 retained earnings. This confirms the findings of BFP that, in the absence of constraints, 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 borrow as is reflected in the relatively higher marginal propensity to innovate with respect to marginal increase in leverage as compared to when firms are unconstrained. Now, under financial constraints, 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 propensity to innovate with respect to leverage is relatively higher than under no financial constraints, 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 marginal propensity to innovate with respect to leverage varies with size, age, and existing leverage. This is because under financial constraints, 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 leverage on innovation conditional on financial constraint against size of the firm. We see that under financial constraints 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 constraints 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 constraints 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, Hennessy and Whited (2007) 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 leverage 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 constraints young firms are more willing to finance themselves by increasing their leverage than matured firms. Consequently, we find the marginal propensity to innovate with respect to leverage of young firms is higher compared to a matured firm. This also makes the young firms more prone to default as discussed in Cooley and Quadriani (2001). However, the difference in APE of leverage 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 leverage on engaging in innovation does not vary much with age.

Lastly, under financial constraints, we find that APE of leverage on innovation declines with higher leverage, which only shows that, *ceteris paribus*, for reasons stated earlier, the borrowing constraints get tighter with higher debt in the capital structure, and the firm becomes more reluctant to engage in innovative activity by increasing leverage.

4.3 R&D Switching Regression Model

In the third stage we estimate the R&D switching regression model, given in equation (25), 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_{it}^* , underlying f_{it} 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 6 about here]

The results of the third stage switching regression estimates are presented in Table 6. 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, s_{it} , and financial constraints, f_{it} , 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 6 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 constraints, f_{it} , 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 4.1, where we discussed the results of the second stage estimation, we include cashflows, 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 to asset ratio or cash reserves to asset ratio. 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 constraints affecting the firm.

Now, even though cashflows turn 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 cashflows for constrained and unconstrained firms, but through the test of the effect of reported financial constraints on R&D investment. While sensitivity of R&D investment to cashflows can indicate the existence of financing frictions, as BFP claim, it could be possible that cashflows are correlated with the R&D investment opportunity set and provide information about future investment opportunities, hence, R&D investment-cashflow sensitivity may equally occur because firms respond to demand signals that cashflows 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 cashflows. Moyen (2004) too finds that cashflow is an excellent proxy for investment opportunity, and that cashflow is an increasing function of the income shock. Hennessy and Whited (2007) also discuss mechanisms, that are related to costs of issuing new equity, bankruptcy costs, and curvature of profit functions, that drive investment-cashflow sensitivity. However, it is beyond the scope of this paper to test for exact mechanism that drives the results on R&D investment-cashflow 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 Klette and Kortum (2004) 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. Besides, $SINS$ is clearly endogenous as is reflected in the significance of correction term for share of innovative sales.

Here, we want to test whether financing friction, as summarized by f_{it} , 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_{it} 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 in Table 6, 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, $\Delta_f E(r_{it}|\bar{\mathcal{X}})$, defined in equation (25).

The other variables included in the specification are size of the firm (*SIZE*), market share (*MKSH*), age of the firm (*AGE*), dummy (*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 Klette and Kortum (2004) who model firm dynamics with R&D and where R&D intensity is independent of firm size. This is because Klette and Kortum (2004) 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 Cooley and Quadrini (2001) 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 Cooley and Quadrini (2001), 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 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, corroborating the result of the second stage, where we found them to more likely to engage in innovative activity as compared to mature firms. We also find that at a given age constrained firms are less likely to invest. 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 likely to be 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 constraints, 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 $\bar{Z}_i'\delta + \hat{\alpha}_i$ 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.

5 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' endogenous decision to innovate and the endogenous 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 constraints, 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) model for R&D investment, where we try to assess the impact of financial constraints on R&D investment, (b) a model for financial constraints, where we try to explain why certain firms report they are financially constrained, (c) a model for decision to innovate, where we try to explain how incentives to innovate are shaped, and (2) a system of reduced form equations of financing choice and other endogenous variables. The first (I) structural part of the analysis, that is, the study of firms' endogenous decision to innovate and the endogenous financial constraint faced, was carried out in the second step, conditional on the first stage reduced form estimation. The second (II) part, where we study the effect of financial constraints on R&D investment, was carried out in the third step, 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 constraints, 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 un-

der no financial constraints, barely varies with firm characteristics such as maturity, size and leverage, under financial constraints 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 is 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 constraints 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. 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 constraints. 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 does 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, D., AND B. LEV (2000): "Information Asymmetry, R&D, and Insider Gains," *Journal of Finance*, 55, 2747 – 2766.
- ALBUQUERQUE, R., AND H. A. HOPENHAYN (2004): "Optimal Lending Contracts and Firm Dynamics," *Review of Economic Studies*, 71, 285–315.
- ALMEIDA, H., AND M. CAMPELLO (2007): "Financial Constraints, Asset Tangibility, and Corporate Investment," *Review of Financial Studies*, 20, 1429 – 1460.
- AUDRETSCH, D. (1995): *Innovation and Industry Evolution*. MIT Press, Cambridge, Massachusetts.
- BAYER, C. (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, J. B., R. C. GREEN, AND V. NAIK (2004): “Valuation and Return Dynamics of New Ventures,” *Review of Financial Studies*, 17, 1–35.
- BIØRN, E. (2004): “Regression Systems for Unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure,” *Journal of Econometrics*, 122, 281–291.
- BLUNDELL, R., AND J. POWELL (2003): “Endogeneity in Nonparametric and Semiparametric Regression Models,” in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II*, ed. by L. H. M. Dewatripont, and S. Turnovsky. Cambridge University Press, Cambridge.
- BROWN, J. R., S. M. FAZZARI, AND B. C. PETERSEN (2009): “Financing Innovation and Growth: Cash Flow, External Equity, and the 1990’s R&D Boom,” *Journal of Finance*, 64, 151–185.
- CASSIMAN, B., M. COLOMBO, P. GARRONE, AND R. VEUGELERS (2005): “The impact of M&A on the R&D process,” *Research Policy*, 34, 195–220.
- CHAMBERLAIN, G. (1984): “Panel Data,” in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2. Elsevier.
- CLEMENTI, G. L., AND H. A. HOPENHAYN (2006): “A Theory of Financing Constraints and Firm Dynamics,” *The Quarterly Journal of Economics*, 121, 229–265.
- COOLEY, T. F., AND V. QUADRINI (2001): “Financial Markets and Firm Dynamics,” *American Economic Review*, 91, 1286–1310.
- GALE, D., AND M. HELLWIG (1985): “Incentive Compatible Debt Contracts: The One Period Problem,” *Review of Economic Studies*, 52, 647–663.
- GAMBA, A., AND A. TRIANTIS (2008): “The value of financial flexibility,” *Journal of Finance*, 63, 2263–2296.
- GOMES, J. F. (2001): “Financing Investment,” *American Economic Review*, 91, 1263–1285.
- GOMES, J. F., A. YARON, AND L. ZHANG (2006): “Asset Pricing Implications of Firms’ Financing Constraints,” *Review of Financial Studies*, 19, 1321–1356.
- HAJIVASSILIOU, V., AND F. 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, B., AND J. LERNER (2010): “The Financing of R&D and Innovation,” in *Handbook of the Economics of Innovation*, ed. by B. H. Hall, and N. Rosenberg. Cambridge University Press, Cambridge, forthcoming.
- HENNESSY, C., AND T. M. WHITED (2007): “How Costly is External Financing? Evidence from a Structural Estimation,” *Journal of Finance*, 62, 1705–1745.
- HOLMSTROM, B. (1989): “Agency Costs and Innovation,” *Journal of Economic Behavior and Organization*, 12, 305–327.

- HUERGO, E., AND J. JAUMANDREU (2004): “How does Probability of Innovation change with Firm Age?,” *Small Business Economics*, 22, 193–207.
- KLEPPER, S., AND P. THOMPSON (2006): “Submarkets and the evolution of market structure,” *RAND Journal of Economics*, 37, 861–886.
- KLETTE, T. J., AND S. KORTUM (2004): “Innovating Firms and Aggregate Innovation,” *Journal of Political Economy*, 112, 986–1018.
- LAMBRECHT, B., AND S. C. MYERS (2008): “Debt and Managerial Rents in a Real-Options Model of the Firm,” *Journal of Financial Economics*, 89, 209–231.
- LELAND, H. E., AND D. H. PYLE (1977): “Informational Asymmetries, Financial Structure, and Financial Intermediation,” *Journal of Finance*, 32, 371–387.
- MOYEN, N. (2004): “Investment-Cash Flow Sensitivities: Constrained versus Unconstrained Firms,” *Journal of Finance*, 69, 2061–2092.
- MYERS, S. (1977): “Determinants of corporate borrowing,” *Journal of Financial Economics*, 5, 147–175.
- PAPKE, L. E., AND J. M. WOOLDRIDGE (2008): “Panel data methods for fractional response variables with an application to test pass rates,” *Journal of Econometrics*, 145, 121–133.
- RAYMOND, W., P. MOHNEN, F. PALM, AND S. SCHIM VAN DER LOEFF (2010): “Persistence of innovation in Dutch manufacturing: Is it spurious?,” *Review of Economics and Statistics*, 92, 495–504.
- ROBERTS, M. R., AND T. WHITED (2010): “Endogeneity in Empirical Corporate Finance,” in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, vol. 2. Elsevier, Amsterdam.
- SEMYKINA, A., AND J. WOOLDRIDGE (2010): “Estimating panel data models in the presence of endogeneity and selection,” *Journal of Econometrics*, 157, 375–380.
- TITMAN, S., AND R. WESSELS (1988): “The Determinants of Capital Structure Choice,” *Journal of Finance*, 43, 1–19.
- WHITED, T. (2006): “External Finance Constraints and the Intertemporal Pattern of Intermittent Investment,” *Journal of Financial Economics*, 81, 467–502.
- WHITED, T. M., AND G. WU (2005): “Financial Constraints Risk,” *Review of Financial Studies*, 19, 531–559.

Appendix A: Identification of Structural Parameters with Expected a Posteriori Values of Individual Effects

We began with a set of structural equations

$$\mathbf{y}_{it}^* = \mathbb{X}'_{it}\mathbf{B} + \tilde{\alpha}_i\mathbf{k} + \Upsilon_{it}, \quad (\text{A-1})$$

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

$$\mathbf{x}_{it} = \mathbf{Z}'_{it}\boldsymbol{\delta} + \tilde{\alpha}_i\boldsymbol{\kappa} + \boldsymbol{\epsilon}_{it}, \quad (\text{A-2})$$

a specification for the conditional expectation and distribution of the individual effects $\tilde{\alpha}_i$,

$$\tilde{\alpha}_i = \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \alpha_i, \quad (\text{A-3})$$

and specification of the joint distribution of idiosyncratic components Υ_{it} and $\boldsymbol{\epsilon}_{it}$. In order to identify the structural parameters of interest in (A-1) we had posited a set of conditional mean restriction, given in equations (6) and (7), which led to the following modified set of structural equations

$$\mathbf{y}_{it}^* = \mathbb{X}'_{it}\mathbf{B} + (\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \alpha_i)\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} + \tilde{\Upsilon}_{it}. \quad (\text{A-4})$$

We had argued that in order to estimate the system of equations in (A-4) the standard technique of the control function approach is to replace $\boldsymbol{\epsilon}_{it}$ by the residuals from the first stage reduced form regression. However, the residuals $\mathbf{x}_{it} - \mathbb{E}(\mathbf{x}_{it}|\mathcal{Z}_i, \alpha_i) = \mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - (\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \alpha_i)\boldsymbol{\kappa}$, remain unidentified because the α_i '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 (4a).

However, it can still be possible to estimate the structural parameters if we can integrate out the α_i s. But given that α_i s are correlated with the endogenous regressors we have to integrate it out with respect to its conditional distribution. Given (A-4), we have

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

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 individual, i , taking expectation of the above with respect to the conditional distribution of α , $\mathbf{f}(\alpha|\mathbf{X}, \mathcal{Z}_i)$ we obtain

$$\begin{aligned} E(\mathbf{y}_{it}^*|\mathbf{X}_i, \mathcal{Z}_i) &= \int \mathbb{E}(\mathbf{y}_{it}^*|\mathbf{X}_i, \mathcal{Z}_i, \alpha_i)\mathbf{f}(\alpha_i|\mathbf{X}_i, \mathcal{Z}_i)d\alpha_i \\ &= \mathbb{X}'_{it}\mathbf{B} + \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + \int (\mathbf{k} - \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa})\alpha_i\mathbf{f}(\alpha_i|\mathbf{X}_i, \mathcal{Z}_i)d\alpha_i \\ &= \mathbb{X}'_{it}\mathbf{B} + \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + \int (\mathbf{k} - \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa})\alpha_i\mathbf{f}(\alpha_i|\mathbf{X}_i)d\alpha_i \\ &= \mathbb{X}'_{it}\mathbf{B} + \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}\boldsymbol{\kappa}) + (\mathbf{k} - \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa})\hat{\alpha}_i \\ &= \mathbb{X}'_{it}\mathbf{B} + (\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - (\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)\boldsymbol{\kappa}) \\ &= \mathbb{X}'_{it}\mathbf{B} + (\bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}, \end{aligned} \quad (\text{A-5})$$

where the second equality follows from the fact that \mathcal{Z}_i and α_i 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 α_i , and Θ_1 is the set of first stage reduced form parameters.

To obtain (A-5), using Bayes rule we can write $\mathbf{f}(\alpha|\mathbf{X}, \mathcal{Z})$ as

$$\mathbf{f}(\alpha|\mathbf{X}) = \frac{\mathbf{f}(\mathbf{X}|\alpha)\mathbf{g}(\alpha)}{\mathbf{h}(\mathbf{X})} = \frac{\mathbf{f}(\mathbf{X}, \mathcal{Z}|\alpha)\mathbf{g}(\alpha)}{\mathbf{h}(\mathbf{X}, \mathcal{Z})}, \quad (\text{A-6})$$

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, α_i s are independent of the exogenous variables \mathcal{Z} , hence $\mathbf{p}(\mathcal{Z}|\alpha) = \mathbf{p}(\mathcal{Z})$, that is,

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

Hence,

$$\begin{aligned} \int \alpha \mathbf{f}(\alpha|\mathbf{X}, \mathcal{Z})d(\alpha) &= \int \frac{\alpha \mathbf{f}(\mathbf{X}|\mathcal{Z}, \alpha)\mathbf{g}(\alpha)d\alpha}{\int \mathbf{f}(\mathbf{X}|\mathcal{Z}, \alpha)\mathbf{g}(\alpha)d\alpha} \\ &= \frac{\int \alpha \prod_{t=1}^T \mathbf{f}(\mathbf{x}_t|\mathcal{Z}, \alpha)\mathbf{g}(\alpha)d\alpha}{\int \prod_{t=1}^T \mathbf{f}(\mathbf{x}_t|\mathcal{Z}, \alpha)\mathbf{g}(\alpha)d\alpha} \\ &= \hat{\alpha}(\mathbf{X}, \mathcal{Z}_i, \Theta_1) \end{aligned} \quad (\text{A-8})$$

where the second equality follow from the fact that conditional on \mathcal{Z} and α , each of the \mathbf{x}_t , $\mathbf{x}_t \in \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ 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 $\mathbf{a} = \frac{\alpha}{\sigma_\alpha}$ follows a standard normal distribution. The functional form of $\hat{\alpha}_i(\mathbf{x}_i, \mathcal{Z}_i, \Theta_1)$ is given by:

$$\begin{aligned} \hat{\alpha}_i(\mathbf{x}_i, \mathcal{Z}_i, \Theta_1) &= \\ &= \frac{\int \sigma_\alpha \mathbf{a} \prod_{t=1}^T \exp(-\frac{1}{2}(\mathbf{x}_t - \mathbf{Z}'_t\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha \mathbf{a})\boldsymbol{\kappa})^T \Sigma_{\epsilon\epsilon}^{-1} (\mathbf{x}_t - \mathbf{Z}'_t\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha \mathbf{a})\boldsymbol{\kappa})) \phi(\mathbf{a}) d\mathbf{a}}{\int \prod_{t=1}^T \exp(-\frac{1}{2}(\mathbf{x}_t - \mathbf{Z}'_t\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha \mathbf{a})\boldsymbol{\kappa})^T \Sigma_{\epsilon\epsilon}^{-1} (\mathbf{x}_t - \mathbf{Z}'_t\boldsymbol{\delta} - (\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \sigma_\alpha \mathbf{a})\boldsymbol{\kappa})) \phi(\mathbf{a}) d\mathbf{a}}. \end{aligned} \quad (\text{A-9})$$

The right hand side of (A-9) is the expected a posteriori value of α_i . $\hat{\alpha}_i(\mathbf{x}_i, \mathcal{Z}_i, \hat{\Theta}_1)$ is the estimated expected a posteriori value of α_i , 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 $\hat{\alpha}_i(\mathbf{x}_i, \mathcal{Z}_i, \hat{\Theta}_1)$ converges a.s. to $\hat{\alpha}_i(\mathbf{x}_i, \mathcal{Z}_i, \Theta_1)$, where $\hat{\Theta}_1$ is the consistent first stage estimates.

Proof of Lemma 1 Given in Section A.1.

Lemma 1 implies that

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

Therefore, if the reduced form population parameters, Θ_1 , are known, the above implies that we could write the linear predictor of y_{it}^* , given \mathbf{X}_i and \mathcal{Z}_i in error form as

$$\mathbf{y}_{it}^* = \mathbb{X}'_{it}\mathbf{B} + (\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)\mathbf{k} + \tilde{\Sigma}_{\Upsilon\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it} + \tilde{\Upsilon}_{it}, \quad (\text{A-10})$$

where conditional of \mathbf{X}_i and \mathcal{Z}_i , $\tilde{\Upsilon}_{it}$ is i.i.d. with mean 0. For liner models, say if all the variables in \mathbf{y}_{it}^* were continuous and observed, with estimates of expected a posteriori values, $\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i$, and the estimates of $\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}$ 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 , $k \in \{1, \dots, n\}$, $n = 4$ in our model, $\tilde{\Sigma}_{\Upsilon_k\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}$ take the form

$$\rho_{\Upsilon_k\epsilon_1}\sigma_{\Upsilon_k}f_1(\Sigma_{\epsilon\epsilon}, \hat{\epsilon}_{1it}, \dots, \hat{\epsilon}_{mit}) + \dots + \rho_{\Upsilon_k\epsilon_m}\sigma_{\Upsilon_k}f_m(\Sigma_{\epsilon\epsilon}, \hat{\epsilon}_{1it}, \dots, \hat{\epsilon}_{mit})$$

where each of the f 's above are linear in $\hat{\boldsymbol{\epsilon}}_{it}$. The estimates $\rho_{\Upsilon_k\epsilon_l}\sigma_{\Upsilon_k}$, $l \in \{1, \dots, m\}$, 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. Also, let, given the first stage population parameters Θ_1 , the linear predictor of f_{it}^* , given \mathbf{X}_i and \mathcal{Z}_i , in error form given by (A-10) be

$$f_{it}^* = \mathcal{X}'_{it}\boldsymbol{\varphi} + \lambda(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it} + \tilde{\zeta}_{it},$$

where $\tilde{\zeta}_{it}$ is i.i.d. and normally distributed with mean 0 and variance $\sigma_{\tilde{\zeta}}^2$. Since in probit models the parameters are identified only up to a scale, for an individual i , the probability of $f_t = 1$ given \mathbf{X}_i and \mathcal{Z}_i is given by

$$\begin{aligned} \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}) &= \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}) = \Pr(y_t^* > 0|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}) \\ &= \Phi\left(\left(\mathcal{X}'_{it}\boldsymbol{\varphi} + \lambda(\bar{\mathcal{Z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}\right)\frac{1}{\sigma_{\tilde{\zeta}}}\right), \end{aligned} \quad (\text{A-11})$$

where the first equality follows from the fact that $\hat{\boldsymbol{\epsilon}}_t$ and $\hat{\alpha}$ is a function of \mathbf{X} and \mathcal{Z} and $\Phi(\cdot)$ is the cumulative standard normal density function. However, $\Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}) = \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, E(\boldsymbol{\epsilon}_t|\mathbf{X}, \mathcal{Z}), E(\tilde{\alpha}|\mathbf{X}, \mathcal{Z}))$ is generally not equal to $\Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \tilde{\alpha})$. Our measure of interest, however, is $\int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \tilde{\alpha})dg(\boldsymbol{\epsilon}_t, \tilde{\alpha})$, since for any individual i , the average partial effect (APE) of changing a variable, say z_k , in time period t from z_{kt} to $z_{kt} + \Delta_{z_k}$ is given by

$$\begin{aligned} \frac{\Delta \Pr(f_t = 1)}{\Delta_{z_k}} &= \left[\int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}_{-z_k}, z_{k-t}, (z_{kt} + \Delta_{z_k}), \boldsymbol{\epsilon}_t, \tilde{\alpha})dg(\boldsymbol{\epsilon}_t, \tilde{\alpha}) \right. \\ &\quad \left. - \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}_{-z_k}, z_{k-t}, z_{kt}, \boldsymbol{\epsilon}_t, \tilde{\alpha})dg(\boldsymbol{\epsilon}_t, \tilde{\alpha}) \right] / \Delta_{z_k}, \end{aligned} \quad (\text{A-12})$$

where $g(\boldsymbol{\epsilon}_t, \tilde{\alpha})$ is the joint distribution function of $\boldsymbol{\epsilon}_t$ and $\tilde{\alpha}$. To recover the above measure in (A-15), like Chamberlain (1984), we make an assumption about the conditional distribution for $\lambda\tilde{\alpha}_i$ and $\tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t$, where the conditioning variable are \mathbf{X}_i and \mathcal{Z}_i . We assume that

$$\lambda\tilde{\alpha}_i = E(\lambda\tilde{\alpha}_i|\mathbf{X}_i, \mathcal{Z}_i) + \bar{\alpha}_{\zeta i} \text{ and } \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it} = E(\tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_{it}|\mathbf{X}_i, \mathcal{Z}_i) + \bar{\epsilon}_{\zeta it} \quad (\text{A-13})$$

where $\bar{\alpha}_{\zeta i}$ and $\bar{\epsilon}_{\zeta it}$, a scalar, are normally distributed with mean 0, and are independent of everything else. Now, for any individual i , we have shown that

$$E(\tilde{\alpha}|\mathbf{X}, \mathcal{Z}) = \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \int \alpha f(\alpha|\mathbf{X}, \mathcal{Z})d(\alpha) = \bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha}(\mathbf{X}, \mathcal{Z}, \Theta_1) \text{ and } E(\tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t|\mathbf{X}_i, \mathcal{Z}_i) = \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t.$$

Hence, under the assumption about the conditional distribution of $\lambda\tilde{\alpha}_i$ and $\tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t$, we can write the linear projection of f_t^* in error form in (A-4), as

$$f_t^* = \mathcal{X}_t^{f'}\boldsymbol{\varphi} + \lambda(\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha}) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t + \omega_{\zeta t} + \bar{\zeta}_t, \quad (\text{A-14})$$

where $\omega_{\zeta t} = \bar{\alpha}_{\zeta} + \bar{\epsilon}_{\zeta t}$ and $\tilde{\zeta}_t$ in (A-10) becomes $\tilde{\zeta}_t = \omega_{\zeta t} + \bar{\zeta}_t$.

Now, having assumed the conditional distribution of $\lambda\tilde{\alpha}_i$ and $\tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\epsilon}_t$, for any individual i , we now have

$$\Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \tilde{\alpha}) = \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t})$$

and

$$\int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \tilde{\alpha})dg(\boldsymbol{\epsilon}_t, \tilde{\alpha}) = \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t})dF(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t}),$$

where $F(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t})$ is the joint distribution function of the arguments. Now

$$\begin{aligned} \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t})dF(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t}) &= \int \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t})d(h(\omega_{\zeta t}, |\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}))dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}) \\ &= \int \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}, \omega_{\zeta t})d(h(\omega_{\zeta t}))dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}) \\ &= \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha})dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}), \end{aligned} \quad (\text{A-15})$$

where $G(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha})$ is the distribution of $\hat{\boldsymbol{\epsilon}}_t$ and $\hat{\alpha}$, and $h(\omega_{\zeta t}, |\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha})$ is the conditional distribution of $\omega_{\zeta t}$ given $\hat{\boldsymbol{\epsilon}}_t$ and $\hat{\alpha}$. The second equality above follows from the fact that $\hat{\boldsymbol{\epsilon}}_t$ and $\hat{\alpha}$ are independent of $\omega_{\zeta t}$. Thus we have shown that

$$\begin{aligned} \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \boldsymbol{\epsilon}_t, \tilde{\alpha})dg(\boldsymbol{\epsilon}_t, \tilde{\alpha}) &= \int \Pr(f_t = 1|\mathbf{X}, \mathcal{Z}, \hat{\boldsymbol{\epsilon}}_t, \hat{\alpha})dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}) \\ &= \int \Phi\left(\left\{\mathcal{X}_t^{f'}\boldsymbol{\varphi} + \lambda(\bar{\mathcal{Z}}'\bar{\boldsymbol{\delta}} + \hat{\alpha}) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_t\right\}\frac{1}{(\sigma_{\omega_{\zeta}}^2 + \sigma_{\bar{\zeta}}^2)^{1/2}}\right)dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}), \end{aligned} \quad (\text{A-16})$$

where $\sigma_{\omega_{\zeta}}^2$ is the variance of $\omega_{\zeta t}$ and $\sigma_{\bar{\zeta}}^2$ the variance of $\bar{\zeta}$. To obtain the sample analog of RHS of (A-16) for any fixed $\mathcal{X}_{it}^f = \bar{\mathcal{X}}^f$ we can compute

$$\frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \Pr(f_{it} = 1|\bar{\mathcal{X}}^f, \hat{\boldsymbol{\epsilon}}_{it}, \hat{\alpha}_i). \quad (\text{A-17})$$

With (A-17) we can now compute (A-12), the mean effect or the average partial effect (APE), of changing a variable, say w , from $w_t = \bar{w}$ to $\bar{w} + \Delta_w$, given $\mathcal{X}_{-wt}^f = \bar{\mathcal{X}}_{-w}^f$. In the limit when Δ_w tends to zero, and since the integrand is a smooth function of its arguments we can change the order of differentiation and integration in (A-16) to get

$$\frac{\partial \Pr(f_t = 1)}{\partial w} = \int \varphi_w \phi \left(\bar{\mathcal{X}}^{f'} \boldsymbol{\varphi} + \lambda(\bar{\mathcal{Z}}' \bar{\boldsymbol{\delta}} + \hat{\alpha}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \right) dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}), \quad (\text{A-18})$$

where ϕ is the density function of a standard normal and the coefficient $-\boldsymbol{\varphi}$, λ , and $\tilde{\Sigma}_{\zeta\epsilon}$ – with a slight abuse of notation, denote the scaled coefficients. Then, for any fixed $\mathcal{X}_{it}^f = \bar{\mathcal{X}}^f$, an estimate of the APE of w , the sample analog of the RHS in (A-18), can be computed as follows:

$$\frac{\partial \widehat{\Pr}(f_{it} = 1)}{\partial w} = \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \hat{\varphi}_w \phi \left(\bar{\mathcal{X}}^{f'} \hat{\boldsymbol{\varphi}} + \lambda(\bar{\mathcal{Z}}_i' \hat{\boldsymbol{\delta}} + \hat{\alpha}_i) + \hat{\Sigma}_{\zeta\epsilon} \hat{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it} \right), \quad (\text{A-19})$$

which converges in probability to its true value in (A-18) as $\sum_{i=1}^N T_i = \mathcal{N} \rightarrow \infty$.

Suppose, w is dummy variable taking values 0 and 1, then the APE of change of w_{it} from 0 to 1, at population parameters, on the probability of $y_{it} = 1$, given other covariates, is given by

$$\int \left[\Phi \left(\bar{\mathcal{X}}_{-w}^{f'} \boldsymbol{\varphi}_{-w} + \varphi_w + \lambda(\bar{\mathcal{Z}}' \bar{\boldsymbol{\delta}} + \hat{\alpha}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \right) - \Phi \left(\bar{\mathcal{X}}_{-w}^{f'} \boldsymbol{\varphi}_{-w} + \lambda(\bar{\mathcal{Z}}' \bar{\boldsymbol{\delta}} + \hat{\alpha}) + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_t \right) \right] dG(\hat{\boldsymbol{\epsilon}}_t, \hat{\alpha}), \quad (\text{A-20})$$

whose sample analog, given $\mathcal{X}_{it-w}^f = \bar{\mathcal{X}}_{-w}^f$, the estimated first and second stage coefficients, $\hat{\boldsymbol{\epsilon}}_{it}$, and $\hat{\alpha}_i$, can be computed by employing (A-17) for $w = 1$ and $w = 0$. In this Appendix we have shown the identification of structural parameters only for the financial constraint equation. A similar equation as (A-14) also holds for the R&D and selection equations. That is $\tilde{\mathbf{Y}}_{it}$ in equations (13) to (15) in the main text is $\tilde{\mathbf{Y}}_{it} = \{\tilde{\eta}_{1it}, \tilde{\eta}_{0it}, \tilde{\zeta}_{it}, \tilde{v}_{it}\}' = \{\omega_{\eta_{1it}} + \tilde{\eta}_{1it}, \omega_{\eta_{0it}} + \tilde{\eta}_{0it}, \omega_{\zeta_{it}} + \tilde{\zeta}_{it}, \omega_{v_{it}} + \tilde{v}_{it}\}'$, where $\omega_{\eta_{1it}}$, $\omega_{\eta_{0it}}$, and $\omega_{v_{it}}$ are defined in the same way as $\omega_{\zeta_{it}}$ is defined in (A-14) and $\tilde{\mathbf{Y}}_{it} = \{\tilde{\eta}_{1it}, \tilde{\eta}_{0it}, \tilde{\zeta}_{it}, \tilde{v}_{it}\}'$ in equations (8) and (9) in the main text.

A.1 Proof of Lemma 1

Proof: Let Θ_1^* be true value of first stage reduced form parameters. Now, for an individual 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, $|\alpha|$, $|\cdot|$ being the absolute value of its argument, is continuous in α and $|\alpha| \geq \alpha \exp(-\frac{1}{2}r(\Theta_1, \alpha)) \forall \Theta_1 \in \boldsymbol{\Theta}_1$. 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 Lebesgue 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 Lebesgue 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 surely 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: Information on Firms for which Imputations was done in from moving to Enterprise to Firm Level Data

Table B.1: 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 3: 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
Cashflow*	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 Cashflow	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 4: Second Stage Estimates: Financial Constraints and Innovation

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

Total Number of Observations: 13032

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

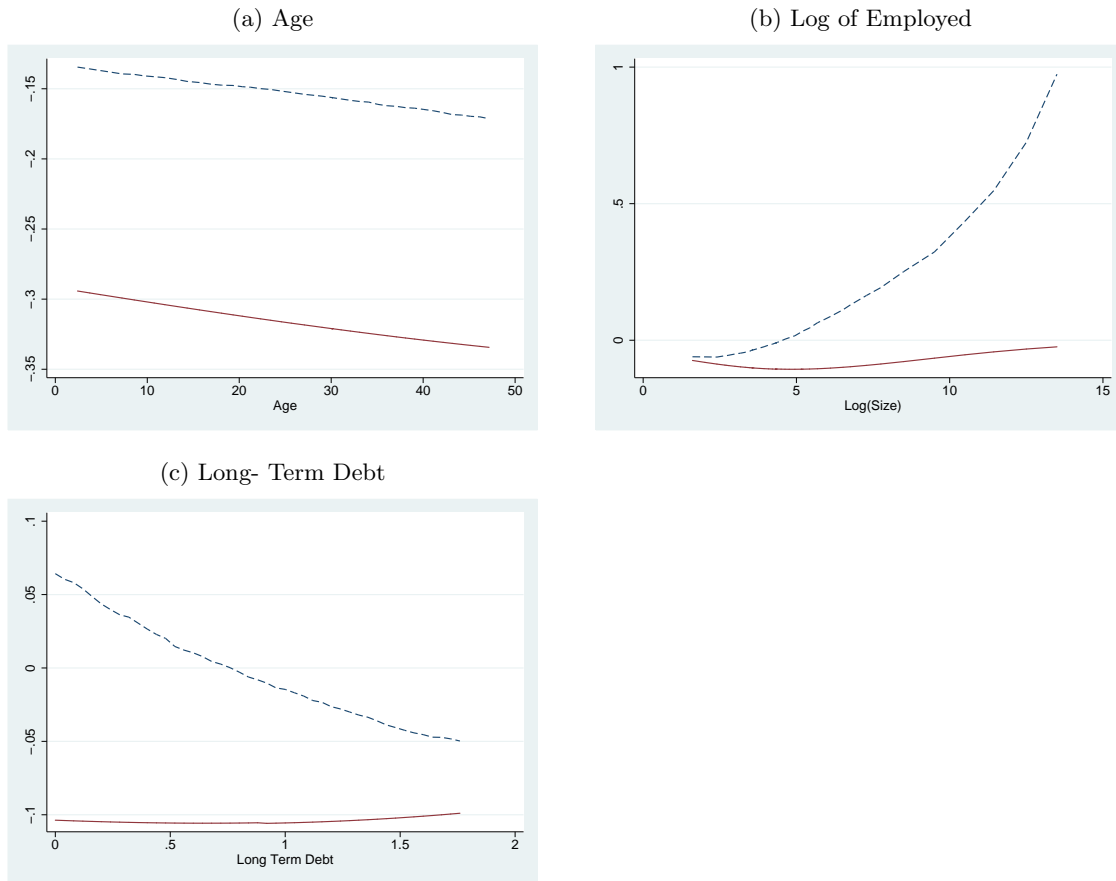
† The estimated coefficient for the Control Functions for Share of Innovative Sales, Long-term Debt, Dividends, Liquidity Reserve, Ratio of Intangible to total Assets, and Size correspond to the estimated terms in $\bar{\Sigma}_{\zeta\epsilon} = \{\rho_{\zeta\epsilon_1}\sigma_\zeta, \dots, \rho_{\zeta\epsilon_m}\sigma_\zeta\}$ of equation (14) and $\bar{\Sigma}_{v\epsilon} = \{\rho_{v\epsilon_1}\sigma_v, \dots, \rho_{v\epsilon_m}\sigma_v\}$ of equation (15).

Table 5: 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*** (0.004)		0.028*** (0.003)		0.028*** (0.003)	
Long Term debt	0.107*** (0.034)	-0.091*** (0.025)	0.108*** (0.034)	-0.091*** (0.027)	0.108*** (0.034)	-0.3*** (0.01)
Cashflows	0.043*** (0.006)		0.043*** (0.006)		0.043*** (0.006)	
Dummy for Negative Cashflows	0.163*** (0.02)		0.166*** (0.019)		0.166*** (0.019)	
Liquidity Reserve	-0.036*** (0.013)	0.127*** (0.023)	-0.041*** (0.005)	0.127*** (0.024)	-0.041*** (0.005)	0.199*** (0.013)
Dividends	-0.497*** (0.066)	0.005 (0.005)	-0.502*** (0.062)	0.005 (0.005)	-0.502*** (0.062)	-0.013*** (0.002)
Size	-0.067*** (0.009)	0.072*** (0.009)	-0.066*** (0.009)	0.072*** (0.008)	-0.066*** (0.009)	0.097*** (0.005)
Market Share	0.001 (0.001)	0.032*** (0.005)	0.001 (0.001)	0.032*** (0.005)	0.001 (0.001)	0.008*** (0.003)
Age	-0.001** (0.001)	-0.003*** (0)	-0.002*** (0.001)	-0.003*** (0)	-0.002*** (0.001)	-0.002*** (0)
Ratio of Intangible Assets to Total Assets	0.006 (0.004)	-0.064*** (0.008)	0.008*** (0.002)	-0.064*** (0.008)	0.008*** (0.002)	0.023*** (0.003)
Dummy for Multiple Enterprise Firms	0.011 (0.023)	0.555*** (0.097)		0.621*** (0.013)		0.866*** (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, $\frac{\partial \Pr(s_{it}=1|f_{it}=1)}{\partial DEBT_{it}}$, or *not* Financially Constrained, $\frac{\partial \Pr(s_{it}=1|f_{it}=0)}{\partial DEBT_{it}}$, against Age, Size, and Leverage.



$$\frac{\partial \Pr(s_{it} = 1|f_{it} = 1)}{\partial DEBT_{it}} \text{ --- --- --- } , \frac{\partial \Pr(s_{it} = 1|f_{it} = 0)}{\partial DEBT_{it}} \text{ --- --- ---}$$

Table 6: 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-term Debt	0.525** (0.213)	0.511** (0.215)
f * Cashflow	0.07* (0.041)	0.071* (0.041)	Dividends	-1.296*** (0.39)	-1.232*** (0.363)
$(1 - f)$ * Cashflow	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 ($C_{11}(\cdot)_t$)	0.967*** (0.319)	0.83*** (0.209)
$(1 - f)$ *Market share	0.011 (0.012)	0.005 (0.004)	Selection ($C_{12}(\cdot)_t$)	0.636* (0.326)	0.589* (0.306)
f *Size	-0.494*** (0.118)	-0.431*** (0.071)	Individual effects ($\tilde{Z}_i\tilde{\delta} + \hat{\alpha}_i$)	-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-term 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 ($C_{01}(\cdot)_t$)	-0.277 (0.198)	-0.186** (0.065)
			Selection ($C_{02}(\cdot)_t$)	-0.883*** (0.324)	-0.745*** (0.114)
			Individual effects ($\tilde{Z}_i\tilde{\delta} + \hat{\alpha}_i$)	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 coefficient for the Control Functions for Share of Innovative Sales, Long-term Debt, Dividends, Liquidity Reserve, Ratio of Intangible to total Assets, and Size correspond to the estimated terms in $\tilde{\Sigma}_{\eta_1}\epsilon = \{\rho_{\eta_1\epsilon_1}\sigma_{\eta_1}, \dots, \rho_{\eta_1\epsilon_m}\sigma_{\eta_1}\}$ for the firms that are financially constrained and $\tilde{\Sigma}_{\eta_0}\epsilon = \{\rho_{\eta_0\epsilon_1}\sigma_{\eta_0}, \dots, \rho_{\eta_0\epsilon_m}\sigma_{\eta_0}\}$ for the firms that are not financially constrained of the R&D equation (13).

Supplement to “Microeconomic Evidence of Financing Frictions and Innovative Activity”: Supplementary Appendix

Amaresh K Tiwari*, Pierre Mohnen†, Franz C. Palm‡
 Sybrand Schim van der Loeff§

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

Appendix C: Maximum Likelihood Estimation of the Reduced Form Equations

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

The system of m reduced form equations in equation (4a) is given by

$$\mathbf{x}_{it} = \mathbf{Z}'_{it} \boldsymbol{\delta} + \bar{\mathbf{Z}}'_i \bar{\boldsymbol{\delta}} \boldsymbol{\kappa} + \alpha_i \boldsymbol{\kappa} + \boldsymbol{\epsilon}_{it} = \mathbf{Z}'_{it} \boldsymbol{\delta} + \bar{\mathbf{Z}}'_i \bar{\boldsymbol{\delta}} \boldsymbol{\kappa} + \mathbf{u}_{it}, \quad (\text{C-1})$$

where $\mathbf{x}_{it} = (x_{1it}, \dots, x_{mit})'$ and $\mathbf{Z}_{it} = \text{diag}(\mathbf{z}_{1it}, \dots, \mathbf{z}_{mit})$ is the matrix of exogenous variables appearing in each of the m reduced form equation in (C-1). $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \dots, \boldsymbol{\delta}'_m)'$, $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_m)'$, 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:

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

*University of Liege, A.Tiwari@ulg.ac.be

†Maastricht University, UNU-MERIT and CIRANO, P.Mohnen@MERIT.unu.edu

‡Maastricht University and CESifo fellow, F.Palm@maastrichtuniversity.nl

§Maastricht University, S.Loeff@maastrichtuniversity.nl

ϵ_{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 $\mathbf{x}_{i(p)} = (\mathbf{x}'_{i1}, \dots, \mathbf{x}'_{ip})'$, $\mathbf{Z}_{i(p)} = (\mathbf{Z}'_{i1}, \dots, \mathbf{Z}'_{ip})'$ and $\boldsymbol{\epsilon}_{i(p)} = (\boldsymbol{\epsilon}'_{i1}, \dots, \boldsymbol{\epsilon}'_{ip})'$ and write the model as

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

$$\text{E}(\mathbf{u}_{i(p)}\mathbf{u}'_{i(p)}) = I_p \otimes \Sigma_{\epsilon\epsilon} + E_p \otimes \Sigma_{\alpha} = K_p \otimes \Sigma_{\epsilon\epsilon} + J_p \otimes \Sigma_{(p)} = \Omega_{u(p)} \quad (\text{C-3})$$

where

$$\Sigma_{(p)} = \Sigma_{\epsilon\epsilon} + p\Sigma_{\alpha}, \quad p = 1, \dots, P \quad (\text{C-4})$$

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 $\Omega_{u(p)}$.

C.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 $Q_{i(p)} = \mathbf{u}'_{i(p)}\Omega_{u(p)}^{-1}\mathbf{u}_{i(p)}$, 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

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

with $\mathbf{u}_{i(p)} = \mathbf{x}_{i(p)} - \mathbf{Z}'_{i(p)}\boldsymbol{\delta} - (e_p \otimes \bar{\mathbf{Z}}'_i\bar{\boldsymbol{\delta}}\boldsymbol{\kappa})$. Had we not imposed the restriction that $\tilde{\boldsymbol{\delta}}$ 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 (2004).

C.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 ϵ_{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 $\mathbf{u}_{i(p)} = (e_p \otimes \alpha_i\boldsymbol{\kappa}) + \boldsymbol{\epsilon}_{i(p)} \sim IIN(0_{mp,1}, \Omega_{u(p)})$. The log-likelihood function of all \mathbf{x} 's conditional on all \mathbf{Z} 's for an individual in group p and for all individuals then become, respectively,

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

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

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

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

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 $\tilde{\delta}$ 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 (C-5) with respect to $\tilde{\delta}$ and (C-7) with respect to $\Sigma_{\epsilon\epsilon}$, $\boldsymbol{\kappa}$, and σ_α^2 and iterating until convergence. Biörn (2004) 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 D: Derivation of the Correction Terms for the Third Stage Switching Regression Model

To avoid complicating the notations, we denote the idiosyncratic error components, $\tilde{\eta}_1$, $\tilde{\eta}_0$, $\tilde{\zeta}$ and \tilde{v} in equations (13) to (15) respectively as η_1 , η_0 , ζ and v . We know that the conditional expectation of η , where η is either η_1 or η_0 , given ζ and v , $E[\eta|\zeta, v]$, is given by

$$E[\eta|\zeta, v] = \mu_\eta + \frac{\sigma_\eta(\rho_{\eta\zeta} - \rho_{\eta v}\rho_{\zeta v})(\zeta - \mu_\zeta)}{\sigma_\zeta(1 - \rho_{\zeta v}^2)} + \frac{\sigma_\eta(\rho_{\eta v} - \rho_{\eta\zeta}\rho_{\zeta v})(v - \mu_v)}{\sigma_v(1 - \rho_{\zeta v}^2)}.$$

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

$$E[\eta|\zeta, v] = \frac{\sigma_\eta(\rho_{\eta\zeta} - \rho_{\eta v}\rho_{\zeta v})(\zeta)}{\sigma_\zeta(1 - \rho_{\zeta v}^2)} + \frac{\sigma_\eta(\rho_{\eta v} - \rho_{\eta\zeta}\rho_{\zeta v})(v)}{\sigma_v(1 - \rho_{\zeta v}^2)}.$$

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

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

which can be written as

$$E[\eta|\zeta, v] = \frac{\sigma_\eta\rho_{\eta\zeta}}{(1 - \rho_{\zeta v}^2)}(\bar{\zeta} - \rho_{\zeta v}\bar{v}) + \frac{\sigma_\eta\rho_{\eta v}}{(1 - \rho_{\zeta v}^2)}(\bar{v} - \rho_{\zeta v}\bar{\zeta}). \quad (\text{D-1})$$

Hence,

$$\begin{aligned}
\mathbb{E}[\eta|\zeta > -a, v > -b] &= \mathbb{E}[\eta|\bar{\zeta} > \frac{-a}{\sigma_\zeta}, \bar{v} > \frac{-b}{\sigma_v}] = \frac{\int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \mathbb{E}[\eta|\bar{\zeta}, \bar{v}] \phi_2(\bar{\zeta}, \bar{v}, \rho_{\zeta v}) d\bar{\zeta} d\bar{v}}{\Phi_2\left(\frac{a}{\sigma_\zeta}, \frac{b}{\sigma_v}, \rho_{\zeta v}\right)} \\
&= \frac{1}{\Phi_2\left(\frac{a}{\sigma_\zeta}, \frac{b}{\sigma_v}, \rho_{\zeta v}\right)} \frac{\sigma_\eta \rho_{\eta\zeta}}{(1 - \rho_{\zeta v}^2)} \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} (\bar{\zeta} - \rho_{\zeta v} \bar{v}) \phi_2(\bar{\zeta}, \bar{v}, \rho_{\zeta v}) d\bar{\zeta} d\bar{v} \\
&+ \frac{1}{\Phi_2\left(\frac{a}{\sigma_\zeta}, \frac{b}{\sigma_v}, \rho_{\zeta v}\right)} \frac{\sigma_\eta \rho_{\eta v}}{(1 - \rho_{\zeta v}^2)} \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} (\bar{v} - \rho_{\zeta v} \bar{\zeta}) \phi_2(\bar{\zeta}, \bar{v}, \rho_{\zeta v}) d\bar{\zeta} d\bar{v}, \tag{D-2}
\end{aligned}$$

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

$$\begin{aligned}
&\int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} (\bar{\zeta} - \rho_{\zeta v} \bar{v}) \phi(\bar{\zeta}) \frac{1}{\sqrt{(1 - \rho_{\zeta v}^2)}} \phi\left(\frac{\bar{v} - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) d\bar{\zeta} d\bar{v} = \\
&\int_{\frac{-a}{\sigma_\zeta}}^{\infty} \bar{\zeta} \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\frac{-b}{\sigma_v} - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right)\right) d\bar{\zeta} - \rho_{\zeta v} \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \bar{v} \phi(\bar{\zeta}) \frac{1}{\sqrt{(1 - \rho_{\zeta v}^2)}} \phi\left(\frac{\bar{v} - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) d\bar{\zeta} d\bar{v}. \tag{D-3}
\end{aligned}$$

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

$$\begin{aligned}
&\int_{\frac{-a}{\sigma_\zeta}}^{\infty} \bar{\zeta} \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\frac{-b}{\sigma_v} - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right)\right) d\bar{\zeta} - \rho_{\zeta v} \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \frac{y \sqrt{(1 - \rho_{\zeta v}^2)} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}} \phi(\bar{\zeta}) \phi(y) d\bar{\zeta} dy \\
&= \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \bar{\zeta} \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\frac{-b}{\sigma_v} - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right)\right) d\bar{\zeta} \\
&- \rho_{\zeta v} \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \frac{y \sqrt{(1 - \rho_{\zeta v}^2)} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}} \phi(\bar{\zeta}) \phi(y) d\bar{\zeta} dy - \rho_{\zeta v}^2 \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \bar{\zeta} \phi(\bar{\zeta}) \phi(y) d\bar{\zeta} dy \tag{D-4}
\end{aligned}$$

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

$$= (1 - \rho_{\zeta v}^2) \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \bar{\zeta} \phi(\bar{\zeta}) \Phi\left(\frac{\frac{b}{\sigma_v} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) d\bar{\zeta} - \rho_{\zeta v} \sqrt{(1 - \rho_{\zeta v}^2)} \int_{\frac{-b - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} y \phi(\bar{\zeta}) \phi(y) d\bar{\zeta} dy. \quad (\text{D-5})$$

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 (D-5) can now be written as

$$\begin{aligned} & (1 - \rho_{\zeta v}^2) \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \bar{\zeta} \phi(\bar{\zeta}) \Phi\left(\frac{\frac{b}{\sigma_v} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) d\bar{\zeta} = (1 - \rho_{\zeta v}^2) \int_{\frac{-a}{\sigma_\zeta}}^{\infty} -d\phi(\bar{\zeta}) \Phi\left(\frac{\frac{b}{\sigma_v} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) \\ & = -(1 - \rho_{\zeta v}^2) \phi(\bar{\zeta}) \Phi\left(\frac{\frac{b}{\sigma_v} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) \Big|_{\frac{-a}{\sigma_\zeta}}^{\infty} + \rho_{\zeta v} \sqrt{(1 - \rho_{\zeta v}^2)} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \phi(\bar{\zeta}) \phi\left(\frac{\frac{b}{\sigma_v} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) d\bar{\zeta} \\ & = (1 - \rho_{\zeta v}^2) \phi\left(\frac{a}{\sigma_\zeta}\right) \Phi\left(\frac{\frac{b}{\sigma_v} - \rho_{\zeta v} \frac{a}{\sigma_\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) + \rho_{\zeta v} \sqrt{(1 - \rho_{\zeta v}^2)} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \phi(\bar{\zeta}) \phi\left(\frac{\frac{b}{\sigma_v} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) d\bar{\zeta}. \quad (\text{D-6}) \end{aligned}$$

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

$$\begin{aligned} & -\rho_{\zeta v} \sqrt{(1 - \rho_{\zeta v}^2)} \int_{\frac{-b - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}}^{\infty} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} y \phi(\bar{\zeta}) \phi(y) d\bar{\zeta} dy = \rho_{\zeta v} \sqrt{(1 - \rho_{\zeta v}^2)} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \int_{\frac{-b - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}}^{\infty} d\phi(y) \phi(\bar{\zeta}) d\bar{\zeta} \\ & = \rho_{\zeta v} \sqrt{(1 - \rho_{\zeta v}^2)} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \phi(y) \Big|_{\frac{-b - \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}}^{\infty} \phi(\bar{\zeta}) d\bar{\zeta} = -\rho_{\zeta v} \sqrt{(1 - \rho_{\zeta v}^2)} \int_{\frac{-a}{\sigma_\zeta}}^{\infty} \phi\left(\frac{\frac{b}{\sigma_v} + \rho_{\zeta v} \bar{\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) \phi(\bar{\zeta}) d\bar{\zeta}. \quad (\text{D-7}) \end{aligned}$$

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

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

Similarly, it can be shown that

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

Hence,

$$\mathbb{E}[\eta | \bar{\zeta} > \frac{-a}{\sigma_\zeta}, \bar{v} > \frac{-b}{\sigma_v}] = \frac{\sigma_\eta \rho_{\eta \zeta} \phi\left(\frac{a}{\sigma_\zeta}\right)}{\Phi_2\left(\frac{a}{\sigma_\zeta}, \frac{b}{\sigma_v}, \rho_{\zeta v}\right)} \Phi\left(\frac{\frac{b}{\sigma_v} - \rho_{\zeta v} \frac{a}{\sigma_\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) + \frac{\sigma_\eta \rho_{\eta v} \phi\left(\frac{b}{\sigma_v}\right)}{\Phi_2\left(\frac{a}{\sigma_\zeta}, \frac{b}{\sigma_v}, \rho_{\zeta v}\right)} \Phi\left(\frac{\frac{a}{\sigma_\zeta} - \rho_{\zeta v} \frac{b}{\sigma_v}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right). \quad (\text{D-8})$$

Now, consider

$$\begin{aligned}
\mathbb{E}[\eta|\zeta \leq -a, v > -b] &= \mathbb{E}[\eta|\bar{\zeta} \leq \frac{-a}{\sigma_\zeta}, \bar{v} > \frac{-b}{\sigma_v}] = \frac{\int_{\frac{-b}{\sigma_v}}^{\infty} \int_{-\infty}^{\frac{-a}{\sigma_\zeta}} \mathbb{E}[\eta|\bar{\zeta}, \bar{v}] \phi_2(\bar{\zeta}, \bar{v}, \rho_{\zeta v}) d\bar{\zeta} d\bar{v}}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_v}, -\rho_{\zeta v}\right)} \\
&= \frac{1}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_v}, -\rho_{\zeta v}\right)} \frac{\sigma_\eta \rho_{\eta\zeta}}{(1 - \rho_{\zeta v}^2)} \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{-\infty}^{\frac{-a}{\sigma_\zeta}} (\bar{\zeta} - \rho_{\zeta v} \bar{v}) \phi_2(\bar{\zeta}, \bar{v}, \rho_{\zeta v}) d\bar{\zeta} d\bar{v} \\
&+ \frac{1}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_v}, -\rho_{\zeta v}\right)} \frac{\sigma_\eta \rho_{\eta v}}{(1 - \rho_{\zeta v}^2)} \int_{\frac{-b}{\sigma_v}}^{\infty} \int_{-\infty}^{\frac{-a}{\sigma_\zeta}} (\bar{v} - \rho_{\zeta v} \bar{\zeta}) \phi_2(\bar{\zeta}, \bar{v}, \rho_{\zeta v}) d\bar{\zeta} d\bar{v}. \tag{D-9}
\end{aligned}$$

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

$$\mathbb{E}[\eta|\bar{\zeta} \leq \frac{-a}{\sigma_\zeta}, \bar{v} > \frac{-b}{\sigma_v}] = \frac{-\sigma_\eta \rho_{\eta\zeta} \phi\left(\frac{a}{\sigma_\zeta}\right)}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_v}, -\rho_{\zeta v}\right)} \Phi\left(\frac{\frac{b}{\sigma_v} - \rho_{\zeta v} \frac{a}{\sigma_\zeta}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right) + \frac{\sigma_\eta \rho_{\eta v} \phi\left(\frac{b}{\sigma_v}\right)}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_v}, -\rho_{\zeta v}\right)} \Phi\left(\frac{\frac{-a}{\sigma_\zeta} + \rho_{\zeta v} \frac{b}{\sigma_v}}{\sqrt{(1 - \rho_{\zeta v}^2)}}\right). \tag{D-10}$$

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 financial constraint equation and the selection equation,

$$\begin{aligned}
f_{it}^* &= \mathcal{X}_{it}^{f'} \boldsymbol{\varphi} + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \lambda + \tilde{\Sigma}_{\zeta\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it} + \tilde{\zeta}_{it}, \\
s_{it}^* &= \mathcal{X}_{it}^{s'} \boldsymbol{\gamma} + (\bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i) \theta + \tilde{\Sigma}_{v\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it} + \tilde{v}_{it},
\end{aligned}$$

given in equations (14) and (15) 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 $w = x_k$, say long-term debt to asset ratio, on the probability of firm being an innovator, $s_{it} = 1$, conditional on being financially constrained, $f_{it} = 1$, as compared to the APE of w ,

on the probability of $s_{it} = 1$, conditional on $f_{it} = 0$. We know that for an individual i

$$\Pr(s_t = 1|f_t = 1, \mathbf{X}, \mathcal{Z}, \hat{\alpha}, \hat{\boldsymbol{\epsilon}}_t) = \frac{\Pr(s_t = 1, f_t = 1, \mathbf{X}, \mathcal{Z}, \hat{\alpha}, \hat{\boldsymbol{\epsilon}}_t)}{\Pr(f_t = 1, \mathbf{X}, \mathcal{Z}, \hat{\alpha}, \hat{\boldsymbol{\epsilon}}_t)} = \frac{\Phi_2(\varphi_t, \gamma_t, \rho_{\zeta\bar{v}})}{\Phi(\varphi_t)},$$

$$\Pr(s_t = 1|f_t = 0, \mathbf{X}, \mathcal{Z}, \hat{\alpha}, \hat{\boldsymbol{\epsilon}}_t) = \frac{\Pr(s_t = 1, f_t = 0, \mathbf{X}, \mathcal{Z}, \hat{\alpha}, \hat{\boldsymbol{\epsilon}}_t)}{\Pr(f_t = 0, \mathbf{X}, \mathcal{Z}, \hat{\alpha}, \hat{\boldsymbol{\epsilon}}_t)} = \frac{\Phi_2(\varphi_t, -\gamma_t, -\rho_{\zeta\bar{v}})}{1 - \Phi(\varphi_t)},$$

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

$$\varphi_{it} = (\mathcal{X}_{it}^{f'}\boldsymbol{\varphi} + \lambda(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{\zeta\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it})\frac{1}{\sigma_{\zeta}}, \text{ and } \gamma_{it} = (\mathcal{X}_{it}^{s'}\boldsymbol{\gamma} + \theta(\bar{\mathcal{Z}}_i'\bar{\boldsymbol{\delta}} + \hat{\alpha}_i) + \tilde{\Sigma}_{v\epsilon}\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it})\frac{1}{\sigma_{\bar{v}}}.$$

Hence, for an individual i we have

$$\frac{\partial \Pr(s_t = 1|f_t = 1)}{\partial w} = \int \frac{\partial}{\partial w} \left(\frac{\Phi_2(\varphi_t, \gamma_t, \rho_{\zeta\bar{v}})}{\Phi(\varphi_t)} \right) dG(\hat{\alpha}, \hat{\boldsymbol{\epsilon}}). \quad (\text{E-1})$$

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 in (E-1)

$$\frac{\partial}{\partial w} \left(\frac{\Phi_2(\varphi_t, \gamma_t, \rho_{\zeta\bar{v}})}{\Phi(\varphi_t)} \right) = \frac{1}{\Phi(\varphi_t)} \left[g_s \gamma_w + \left(g_f - \Phi_2(\cdot) \frac{\phi(\varphi_t)}{\Phi(\varphi_t)} \right) \varphi_w \right], \quad (\text{E-2})$$

where

$$g_f = \phi(\varphi_t) \Phi \left(\frac{\gamma_t - \rho_{\zeta\bar{v}} \varphi_t}{\sqrt{1 - \rho_{\zeta\bar{v}}^2}} \right) \text{ and } g_s = \phi(\gamma_t) \Phi \left(\frac{\varphi_t - \rho_{\zeta\bar{v}} \gamma_t}{\sqrt{1 - \rho_{\zeta\bar{v}}^2}} \right). \quad (\text{E-3})$$

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-2) is estimated at $\mathcal{X}_{it}^f = \bar{\mathcal{X}}^f$ and $\mathcal{X}_{it}^s = \bar{\mathcal{X}}^s$, given the estimates $\hat{\alpha}_i$ and $\hat{\boldsymbol{\epsilon}}_{it}$, 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-19) in Appendix A we know that estimated APE of w on the unconditional probability of being, say, financially constrained is given by

$$\frac{\partial \widehat{\Pr}(f_{it} = 1)}{\partial w} = \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \hat{\varphi}_w \phi(\bar{\mathcal{X}}_{it}^{f'} \hat{\boldsymbol{\varphi}}),$$

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

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

where V_{2f}^* is the second stage error adjusted covariance matrix, shown in Appendix E, of $\hat{\underline{\varphi}}$. The RHS of (E-4) turns out to be

$$\left[\frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \phi(\bar{\mathbb{X}}_{it}^f \hat{\underline{\varphi}}) [e_w - (\hat{\underline{\varphi}}' \bar{\mathbb{X}}_{it}^f) \hat{\varphi}_w \bar{\mathbb{X}}_{it}^f] \right] V_{2f}^* \left[\frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \phi(\bar{\mathbb{X}}_{it}^f \hat{\underline{\varphi}}) [e_w - (\hat{\underline{\varphi}}' \bar{\mathbb{X}}_{it}^f) \hat{\varphi}_w \bar{\mathbb{X}}_{it}^f] \right]', \quad (\text{E-5})$$

where e_w is a row vector having the dimension of $\underline{\varphi}'$ and with 1 at the position of φ_w in $\underline{\varphi}$ and zeros elsewhere. The estimated asymptotic covariance matrix of the APE of all the continuous variables in \mathcal{X}^f on the probability of being financially constrained can be obtained as above.

If w is a dummy variable then from (A-20) we know that the estimated APE of w on the probability of being financially constrained is given by

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

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

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

where

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

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 (selected) given the firm is financially constrained or not financially constrained. Let $\bar{\mathbb{X}}_{2it} = \{\bar{\mathbb{X}}_{it}^f, \bar{\mathbb{X}}_{it}^{st}\}'$ and $\Theta_2 = \{\underline{\varphi}', \underline{\gamma}', \rho_{\zeta\bar{v}}\}'$, where $\bar{\mathbb{X}}_{it}^{st} = \{\bar{\mathcal{X}}^{st}, \hat{\alpha}_i, (\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\epsilon}_{it})\}'$ and $\underline{\gamma} = \{\underline{\gamma}', \theta, \tilde{\Sigma}'_{v\epsilon}\}'$ and denote the right hand side of (E-2) as $\Lambda_{(s=1|f=1),w}(\bar{\mathbb{X}}_{2it}, \Theta_2)$. Then the APE of w on the conditional probability of being an innovator given that the firm is financially constrained is given by

$$\frac{\partial \widehat{\Pr}(s_t = 1 | f_t = 1)}{\partial w} = \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \Lambda_{(s=1|f=1),w}(\bar{\mathbb{X}}_{2it}, \hat{\Theta}_2)$$

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

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

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

E.2 Average Partial Effects for the Third Stage

One of the main purposes of this exercise is to measure the effect of financial constraints, $f_{it} = 1$, on R&D expenditure. For an individual, i , in time period, t , given $\mathcal{X}_{it} = \bar{\mathcal{X}}$, where \mathcal{X}_{it} is the union of elements appearing in \mathcal{X}_{it}^r , \mathcal{X}_{it}^f , and \mathcal{X}_{it}^s , 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 $\hat{\alpha}$ and $\hat{\epsilon}$. The conditional, conditional on being an innovator ($s_{it} = 1$), APE of financial constraint on R&D expenditure is given by

$$\begin{aligned} \Delta_f E(r_{it}|\bar{\mathcal{X}}) &= \int E(r_{1it}|\bar{\mathcal{X}}, f_{it} = 1, s_{it} = 1, \hat{\alpha}, \hat{\epsilon}) dG(\hat{\alpha}, \hat{\epsilon}) \\ &\quad - \int E(r_{0it}|\bar{\mathcal{X}}, f_{it} = 0, s_{it} = 1, \hat{\alpha}, \hat{\epsilon}) dG(\hat{\alpha}, \hat{\epsilon}). \end{aligned} \quad (\text{E-8})$$

From the discussion of the third stage estimation we know that for an individual i

$$\begin{aligned} E(r_{1t}|\bar{\mathcal{X}}, f_t = 1, s_t = 1, \hat{\alpha}, \hat{\epsilon}_t) &= \\ \beta_f + \bar{\mathcal{X}}^r \beta_1 + \mu_1(\bar{\mathcal{Z}}\bar{\delta} + \hat{\alpha}) + \tilde{\Sigma}_{\eta_1\epsilon} \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\epsilon}_t + \sigma_{\tilde{\eta}_1} \rho_{\tilde{\eta}_1\zeta} C_{11}(\hat{\alpha}, \hat{\epsilon}_t) + \sigma_{\tilde{\eta}_1} \rho_{\tilde{\eta}_1\bar{v}} C_{12}(\hat{\alpha}, \hat{\epsilon}_t) \end{aligned} \quad (\text{E-9})$$

if $f_{it}^* > 0$, and

$$\begin{aligned} E(r_{0t}|\bar{\mathcal{X}}, f_t = 0, s_t = 1, \hat{\alpha}, \hat{\epsilon}_t) &= \\ \bar{\mathcal{X}}_t^r \beta_0 + \mu_0(\bar{\mathcal{Z}}\bar{\delta} + \hat{\alpha}) + \Sigma_{\eta_0\epsilon} \Sigma_{\epsilon\epsilon}^{-1} \hat{\epsilon}_t + \sigma_{\tilde{\eta}_0} \rho_{\tilde{\eta}_0\zeta} C_{01}(\hat{\alpha}, \hat{\epsilon}_t) + \sigma_{\tilde{\eta}_0} \rho_{\tilde{\eta}_0\bar{v}} C_{02}(\hat{\alpha}, \hat{\epsilon}_t) \end{aligned} \quad (\text{E-10})$$

if $f_{it}^* \leq 0$, and where the correction terms $-C_{11}(\bar{\mathcal{X}}^f, \bar{\mathcal{X}}^s, \hat{\alpha}, \hat{\epsilon}_t)$, $C_{12}(\bar{\mathcal{X}}^f, \bar{\mathcal{X}}^s, \hat{\alpha}, \hat{\epsilon}_t)$, $C_{01}(\bar{\mathcal{X}}^f, \bar{\mathcal{X}}^s, \hat{\alpha}, \hat{\epsilon}_t)$, and $C_{02}(\bar{\mathcal{X}}^f, \bar{\mathcal{X}}^s, \hat{\alpha}, \hat{\epsilon}_t)$ – are defined at the given $\mathcal{X}_{it}^f = \bar{\mathcal{X}}^f$ and $\mathcal{X}_{it}^s = \bar{\mathcal{X}}^s$. 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 $s_{it} = 1$.

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 second 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 will be a function of the variance of the estimated second stage coefficients.

To see this, consider the conditional APE of the financial constraint on the R&D

expenditure, which is given by

$$\begin{aligned} \Delta_f \hat{E}(r_t | \bar{\mathcal{X}}) = & \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \left[s_{it} \left(\hat{\beta}_f + \bar{\mathcal{X}}^{r'} (\hat{\beta}_1 - \hat{\beta}_0) + (\hat{\mu}_1 - \hat{\mu}_0) (\bar{\mathcal{Z}} \hat{\delta} + \hat{\alpha}) + (\hat{\Sigma}_{\eta_1 \epsilon_k} - \hat{\Sigma}_{\eta_0 \epsilon_k}) \hat{\Sigma}_{\epsilon \epsilon}^{-1} \hat{\epsilon}_t \right. \right. \\ & \left. \left. + \widehat{\sigma}_{\eta_1} \rho_{\eta_1 \zeta} C_{11}(\hat{\alpha}, \hat{\epsilon}_t) + \widehat{\sigma}_{\eta_1} \rho_{\eta_1 \tilde{v}} C_{12}(\hat{\alpha}, \hat{\epsilon}_t) - \widehat{\sigma}_{\eta_0} \rho_{\eta_0 \zeta} C_{01}(\hat{\alpha}, \hat{\epsilon}_t) - \widehat{\sigma}_{\eta_0} \rho_{\eta_0 \tilde{v}} C_{02}(\hat{\alpha}, \hat{\epsilon}_t) \right) \right] \end{aligned} \quad (\text{E-11})$$

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

$$\text{Asy. Var}[\Delta_f \hat{E}(r_t | \bar{\mathcal{X}})] = \left[\frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \frac{\partial \Delta_f \hat{E}(r_t | \bar{\mathcal{X}})}{\partial \Theta_s} \right]' V_s^* \left[\frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \frac{\partial \Delta_f \hat{E}(r_t | \bar{\mathcal{X}})}{\partial \Theta_s} \right], \quad (\text{E-12})$$

where V_s^* , the error corrected asymptotic covariance matrix of $\hat{\Theta}_s$, has been derived in Appendix F. 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 .

Appendix F: 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:

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

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

and

$$E(\mathcal{L}_{i(p)3\Theta_3}) = E \left[\sum_{t=1}^p s_{it} \mathbb{X}_{it}^r (r_{it} - \mathbb{X}_{it}^r \Theta_3) \right] = 0 \quad (\text{F-3})$$

where $L_{i(p)1}(\Theta_1)$ is the likelihood function for individual i belonging to the group p , $p \in \{1, \dots, P\}$, for the first step system of reduced form equations. The notation p was introduced

in Appendix C. p is the number of time period an individual 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 (F-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, $s_{it} = 1$, where

$$\begin{aligned} r_{it} &= f_{it} r_{it} \\ \mathbb{X}_{it}^r &= f_{it} \{f_{it}, \mathcal{X}_{it}^{r'}, \bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i(\Theta_1), (\Sigma_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}(\Theta_1))', C_{11it}(\Theta_1, \Theta_2), C_{12it}(\Theta_1, \Theta_2)\}' \end{aligned}$$

if $f_{it} = 1$, else

$$\begin{aligned} r_{it} &= (1 - f_{it}) r_{it} \\ \mathbb{X}_{it}^r &= (1 - f_{it}) \{f_{it}, \mathcal{X}_{it}^{r'}, \bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i(\Theta_1), (\Sigma_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}(\Theta_1))', C_{01it}(\Theta_1, \Theta_2), C_{02it}(\Theta_1, \Theta_2)\}' \end{aligned}$$

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

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

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

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

where $\mathcal{L}_{it2}(\Theta_1, \Theta_2)$ is given by equations (19) and (20) in the main text and $\Theta_2 = \{\boldsymbol{\varphi}', \boldsymbol{\gamma}', \rho_{\zeta\bar{v}}\}'$ was defined in Appendix E. Finally, the sample analog of the population for the third step estimation is given by

$$\frac{1}{N} \mathcal{L}_{3\Theta_3}(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3) = \frac{1}{N} \sum_{p=1}^P \sum_{i \in I_{(p)}} \sum_{t=1}^p s_{it} \mathbb{X}_{it}^r (r_{it} - \mathbb{X}_{it}^{r'} \hat{\Theta}_3) \quad (\text{F-6})$$

In Appendix A, we had shown that with α_i substituted by their EAP values $\hat{\alpha}_i(\mathbf{X}_i, \mathcal{Z}_i, \Theta_1)$ 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 (1984) 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

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

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 $\mathcal{L}_{3\Theta_3}(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3)$ around Θ^* gives

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

In matrix notation the above can be written as

$$B_{\Theta\Theta_N} \sqrt{N}(\hat{\Theta} - \Theta) = -\frac{1}{\sqrt{N}} \Lambda_{\Theta_N}, \quad (\text{F-8})$$

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, $\mathbb{L}_{2\Theta_2\Theta_1} = E(\mathcal{L}_{i(p)2\Theta_2\Theta_1})$. $\frac{1}{\sqrt{N}}\Lambda_N$ in (F-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 matrix¹ given by $B_*^{-1}A_*B_*^{-1'}$.

$$\sqrt{N}(\hat{\Theta} - \Theta^*) \overset{a}{\sim} N[(0), (B_*^{-1}A_*B_*^{-1'})] \quad (\text{F-9})$$

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

The moment conditions for every individual, 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 MLE, at Θ_1^* and Θ_2^* to which $\hat{\Theta}_1$ and $\hat{\Theta}_2$ converge, we have

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

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 individuals, 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 $\mathcal{L}_{i(p)3\Theta_3}(\Theta_1, \Theta_2, \Theta_3)$ with respect to Θ_1 and the derivative of $\mathcal{L}_{i(p)3\Theta_3}(\Theta_1, \Theta_2, \Theta_3)$ with respect to Θ_2 . But first we begin by deriving the first order conditions of the log likelihood function of the first stage.

F.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, \mathbf{u}_{it} , given in (C-1), for an individual i , $i \in I_p$, in the $(m \times p)$ matrix $\tilde{E}_{i(p)} = [\mathbf{u}_{i1}, \dots, \mathbf{u}_{ip}]$, write $\mathbf{u}_{i(p)} = \text{vec}(E_{i(p)})$, where ‘vec()’ is the vectorization operator, and define

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

where J_p and K_p defined earlier in Appendix C 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

$$\begin{aligned} \frac{\partial \mathcal{L}_{i(p)}}{\partial \boldsymbol{\delta}} &= 2\mathbf{Z}_{i(p)} \Omega_{u(p)}^{-1} \mathbf{u}_{i(p)}, \\ \frac{\partial \mathcal{L}_{i(p)}}{\partial \boldsymbol{\delta}} &= -2\bar{\mathbf{Z}}_i \boldsymbol{\kappa}' \left[\Sigma_{(p)}^{-1} \tilde{E}_{i(p)} J_p + \Sigma_{\epsilon\epsilon}^{-1} \tilde{E}_{i(p)} K_p \right] e_p, \\ \frac{\partial \mathcal{L}_{i(p)}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} &= -\frac{1}{2} \text{vech} \left(\Sigma_{(p)}^{-1} + (p-1)\Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{(p)}^{-1} B_{ui(p)} \Sigma_{(p)}^{-1} - \Sigma_{\epsilon\epsilon}^{-1} W_{ui(p)} \Sigma_{\epsilon\epsilon}^{-1} \right), \\ \frac{\partial \mathcal{L}_{i(p)}}{\partial \boldsymbol{\kappa}} &= -p\sigma_\alpha^2 [\Sigma_{(p)}^{-1} - \Sigma_{(p)}^{-1} B_{ui(p)} \Sigma_{(p)}^{-1}] \boldsymbol{\kappa} + \bar{\mathbf{Z}}_i' \boldsymbol{\delta} \left[\Sigma_{(p)}^{-1} \tilde{E}_{i(p)} J_p + \Sigma_{\epsilon\epsilon}^{-1} \tilde{E}_{i(p)} K_p \right] e_p, \\ \text{and} \\ \frac{\partial \mathcal{L}_{i(p)}}{\partial \sigma_\alpha^2} &= -\frac{1}{2} p [\text{vec}(\Sigma_{(p)}^{-1})' - \text{vec}(\Sigma_{(p)}^{-1} B_{ui(p)} \Sigma_{(p)}^{-1})'] \text{vec}(\Sigma_\kappa), \end{aligned} \quad (\text{F-11})$$

where ‘ $\text{vech}()$ ’ operator is column-wise vectorization of the lower triangle of the symmetric matrix $\Sigma_{\epsilon\epsilon}$ ².

To derive the above we utilize the following matrix results:

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

The log-likelihood for an individual 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

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

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

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

From (c) we get

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

Using (F-10) we obtain

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

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

and from (d) we get

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

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

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

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

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

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

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

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

where the second equality follows from employing (d), and the fact that $\Sigma_{\epsilon\epsilon}$ being given, $d\Sigma_{(p)} = d(\Sigma_{\epsilon\epsilon} + p\sigma_\alpha^2\Sigma_\kappa) = pd(\sigma_\alpha^2\Sigma_\kappa)$. 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 (F-16) we can conclude that

$$\frac{\partial Q_{i(p)}(\cdot)}{\partial \boldsymbol{\kappa}} = -2p\sigma_\alpha^2[\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1}]\boldsymbol{\kappa} - 2\bar{\mathbf{Z}}_i'\bar{\boldsymbol{\delta}}\left[\Sigma_{(p)}^{-1}\tilde{E}_{i(p)}J_p + \Sigma_{\epsilon\epsilon}^{-1}\tilde{E}_{i(p)}K_p\right]e_p. \quad (\text{F-17})$$

Similarly, from (F-15) we obtain

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

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

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

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

F.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 selection equation Θ_{2s} and $\rho_{\zeta\tilde{v}}$, the correlation between the idiosyncratic components of the financial constraint and the selection equation. In matrix form we can write

$$\mathcal{L}_{i(p)2\Theta_2\Theta_1} = \frac{\partial \mathcal{L}_{i(p)2\Theta_2}}{\partial \Theta_1} = \sum_{t=1}^p \frac{\partial \mathcal{L}_{it2\Theta_2}}{\partial \Theta_1} = \sum_{t=1}^p \begin{bmatrix} \frac{\partial \mathcal{L}_{it2\Theta_{2f}}}{\partial \Theta_1} \\ \frac{\partial \mathcal{L}_{it2\Theta_{2s}}}{\partial \Theta_1} \\ \frac{\partial \mathcal{L}_{it2\rho_{\zeta\tilde{v}}}}{\partial \Theta_1} \end{bmatrix},$$

where the score functions, $\mathcal{L}_{it2\Theta_{2f}}$, $\mathcal{L}_{it2\Theta_{2s}}$, and $\mathcal{L}_{it2\rho_{\zeta\tilde{v}}}$, 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

$$\mathcal{L}_{it2\Theta_{2f}}(\Theta_1, \Theta_2) = \frac{qitfgitf}{\Phi_2} \mathbb{X}_{it}^f, \quad \mathcal{L}_{it2\Theta_{2s}}(\Theta_1, \Theta_2) = \frac{qitsgits}{\Phi_2} \mathbb{X}_{it}^s, \quad \text{and} \quad \mathcal{L}_{it2\rho_{\zeta\tilde{v}}}(\Theta_1, \Theta_2) = \frac{qitfqits\phi_2}{\Phi_2} \quad (\text{F-20})$$

where $\mathbb{X}_{it}^f = \{\mathcal{X}_{it}^{f'}, \bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i, (\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it})'\}'$, $\mathbb{X}_{it}^s = \{\mathcal{X}_{it}^{s'}, \bar{\mathcal{Z}}_i' \bar{\boldsymbol{\delta}} + \hat{\alpha}_i, (\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it})'\}'$, $q_{itf} = 2f_{it} - 1$, $q_{its} = 2s_{it} - 1$. g_{itf} and g_{its} in (F-20) are defined as

$$g_{itf} = \phi(\varphi_{it}) \Phi\left(\frac{\gamma_{it} - \rho_{\bar{\zeta}\bar{v}}^* \varphi_{it}}{\sqrt{(1 - \rho_{\bar{\zeta}\bar{v}}^{*2})}}\right), \text{ and } g_{its} = \phi(\gamma_{it}) \Phi\left(\frac{\varphi_{it} - \rho_{\bar{\zeta}\bar{v}}^* \gamma_{it}}{\sqrt{(1 - \rho_{\bar{\zeta}\bar{v}}^{*2})}}\right),$$

where $\rho_{\bar{\zeta}\bar{v}}^* = q_{itf} q_{its} \rho_{\bar{\zeta}\bar{v}}$, $\varphi_{it} = \mathbb{X}_{it}^{f'} \Theta_{2f}$, and $\gamma_{it} = \mathbb{X}_{it}^{s'} \Theta_{2s}$. However, for CIS2.5, where we do not observe f_{it} when $s_{it} = 0$, the score functions remain the same as in (F-20) when $s_{it} = 1$, but are

$$\mathcal{L}_{it2\Theta_{2f}}(\Theta_1, \Theta_2) = \mathbf{0}_{\Theta_{2f}}, \quad \mathcal{L}_{it2\Theta_{2s}}(\Theta_1, \Theta_2) = -\frac{\phi(-\mathbb{X}_{it}^{s'} \Theta_{2s})}{\Phi(-\mathbb{X}_{it}^{s'} \Theta_{2s})} \mathbb{X}_{it}^s, \text{ and } \mathcal{L}_{it2\rho_{\bar{\zeta}\bar{v}}}(\Theta_1, \Theta_2) = 0 \quad (\text{F-21})$$

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

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

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

where $j \in \{f, s\}$ and

$$\frac{\partial \mathcal{L}_{2\rho_{\bar{\zeta}\bar{v}}}(\Theta_1, \Theta_2)}{\partial \Theta_1'} = q_f q_s \left\{ \frac{\partial}{\partial \Theta_1'} \left(\frac{\phi_2}{\Phi_2} \right) \right\} = q_f q_s \left\{ \frac{\partial(\phi_2/\Phi_2)}{\partial \varphi_{it}} \frac{\partial \mathbb{X}_{it}^{f'}}{\partial \Theta_1'} \Theta_{2f} + \frac{\partial(\phi_2/\Phi_2)}{\partial \gamma_{it}} \frac{\partial \mathbb{X}_{it}^{s'}}{\partial \Theta_1'} \Theta_{2s} \right\} \quad (\text{F-23})$$

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

$$\begin{aligned} \frac{\partial \mathcal{L}_{2\Theta_{2s}}(\Theta_1, \Theta_2)}{\partial \Theta_1'} &= -\left\{ \frac{\partial}{\partial \Theta_1'} \left(\frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})} \right) \mathbb{X}_{it}^s + \frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})} \frac{\partial \mathbb{X}_{it}^s}{\partial \Theta_1'} \right\} \\ &= -\left\{ \frac{\partial}{\partial \gamma_{it}} \left(\frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})} \right) \frac{\partial \mathbb{X}_{it}^{s'}}{\partial \Theta_1'} \Theta_{2s} \mathbb{X}_{it}^s + \frac{\phi(-\gamma_{it})}{\Phi(-\gamma_{it})} \frac{\partial \mathbb{X}_{it}^s}{\partial \Theta_1'} \right\} \end{aligned} \quad (\text{F-24})$$

To obtain expressions for (F-22), (F-23), and (F-24) we need the derivative of $\frac{g_j}{\Phi_2}$, $j \in \{f, s\}$, 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}^s with respect to Θ_1 .

$$\frac{\partial \mathbb{X}_{it}^j}{\partial \Theta'_1} = \begin{bmatrix} \frac{\partial \mathcal{X}_{it}^j}{\partial \delta'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \bar{\delta}'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \boldsymbol{\kappa}'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} & \frac{\partial \mathcal{X}_{it}^j}{\partial \sigma_\alpha^2} \\ \frac{\partial(\bar{\mathcal{Z}}'_i \bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \delta'} & \frac{\partial(\bar{\mathcal{Z}}'_i \bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \bar{\delta}'} & \frac{\partial(\bar{\mathcal{Z}}'_i \bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \boldsymbol{\kappa}'} & \frac{\partial(\bar{\mathcal{Z}}'_i \bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} & \frac{\partial(\bar{\mathcal{Z}}'_i \bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial \sigma_\alpha^2} \\ \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \delta'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \bar{\delta}'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \boldsymbol{\kappa}'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} & \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \sigma_\alpha^2} \end{bmatrix},$$

where $j \in \{f, s\}$. While $\frac{\partial \mathcal{X}_{it}^j}{\partial \Theta'_1} = \mathbf{0}$, below we show that

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

$$\begin{aligned} \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \delta'} &= -\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \mathbf{Z}'_{it} + \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa}}{U_{dr}^2} \sum_{t=1}^p \left[U_{nr}^2 - U_{dr} F_{dr} \right] \boldsymbol{\kappa}' \Sigma_{\epsilon\epsilon}^{-1} \mathbf{Z}'_{it} \\ \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \bar{\delta}'} &= -\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa} \bar{\mathcal{Z}}'_i + \frac{p \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa}}{U_{dr}^2} \left[U_{nr}^2 - U_{dr} F_{dr} \right] \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 \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\} \\ \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\boldsymbol{\epsilon}}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} &= \left[(\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})' \right] L'_m \\ &\quad - \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa}}{2(U_{dr})^2} \sum_{t=1}^p \left[(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})' \right] (\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}), \end{aligned} \quad (\text{F-26})$$

where

$$\begin{aligned} U_{nr} &= \int \alpha \exp\left(-\frac{1}{2} \sum_{t=1}^p \boldsymbol{\epsilon}'_{it} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it}\right) \phi(\alpha) d\alpha, & F_{nr} &= \int \alpha^3 \exp\left(-\frac{1}{2} \sum_{t=1}^p \boldsymbol{\epsilon}'_{it} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it}\right) \phi(\alpha) d\alpha, \\ U_{dr} &= \int \exp\left(-\frac{1}{2} \sum_{t=1}^p \boldsymbol{\epsilon}'_{it} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it}\right) \phi(\alpha) d\alpha, & F_{dr} &= \int \alpha^2 \exp\left(-\frac{1}{2} \sum_{t=1}^p \boldsymbol{\epsilon}'_{it} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\epsilon}_{it}\right) \phi(\alpha) d\alpha, \end{aligned}$$

and $\mathbf{r}_{it} = \mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \boldsymbol{\kappa}'\bar{\mathbf{z}}_i\bar{\boldsymbol{\delta}}$. We note here that while construction of $\hat{\alpha}_i$ to estimate the structural parameters involved estimation of U_{nr} and U_{dr} , in order to estimate the covariance matrix of the structural parameters F_{nr} and F_{dr} will also have to be estimated.

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

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

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

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

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

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

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

where

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

Hence we have

$$\frac{\partial(\bar{\mathbf{z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial\boldsymbol{\delta}'} = -\frac{1}{U_{dr}^2} \sum_{t=1}^p \left[U_{nr}^2 - U_{dr}F_{dr} \right] \boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\mathbf{Z}'_{it}, \quad (\text{F-29})$$

and

$$\frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\boldsymbol{\epsilon}}_{it}}{\partial\boldsymbol{\delta}'} = \frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}(\mathbf{x}_{it} - \mathbf{Z}'_{it}\boldsymbol{\delta} - \bar{\mathbf{z}}'_i\bar{\boldsymbol{\delta}}\boldsymbol{\kappa})}{\partial\boldsymbol{\delta}'} - \frac{\partial\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\alpha}_i\boldsymbol{\kappa}}{\partial\boldsymbol{\delta}'} = -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\mathbf{Z}'_{it} + \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}}{U_{dr}^2} \sum_{t=1}^p \left[U_{nr}^2 - U_{dr}F_{dr} \right] \boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\mathbf{Z}'_{it}. \quad (\text{F-30})$$

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

$$\frac{\partial(\bar{\mathbf{z}}'_i\bar{\boldsymbol{\delta}} + \hat{\alpha}_i)}{\partial\boldsymbol{\delta}'} = \bar{\mathbf{z}}'_i - \frac{1}{U_{dr}^2} \sum_{t=1}^p \left[U_{nr}^2 - U_{dr}F_{dr} \right] \boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\bar{\mathbf{z}}'_i = \bar{\mathbf{z}}'_i - \frac{p}{U_{dr}^2} \left[U_{nr}^2 - U_{dr}F_{dr} \right] \boldsymbol{\kappa}'\Sigma_{\epsilon\epsilon}^{-1}\boldsymbol{\kappa}\bar{\mathbf{z}}'_i, \quad (\text{F-31})$$

and

$$\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\epsilon}_{it}}{\partial \delta'} = \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} (\mathbf{x}_{it} - \mathbf{Z}'_{it} \boldsymbol{\delta} - \bar{\mathbf{Z}}'_i \bar{\boldsymbol{\delta}} \boldsymbol{\kappa})}{\partial \delta'} - \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\alpha}_i \boldsymbol{\kappa}}{\partial \delta'} = -\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa} \bar{\mathbf{Z}}'_i + \frac{p \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa}}{U_{dr}^2} \left[U_{nr}^2 - U_{dr} F_{dr} \right] \boldsymbol{\kappa}' \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa} \bar{\mathbf{Z}}'_i. \quad (\text{F-32})$$

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

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

which after simplification can be written as

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

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

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

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

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

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

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

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

$$\frac{\partial \hat{\alpha}_i}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} = \frac{1}{2U_{dr}^2} \sum_{t=1}^p \left[(U_{nr}^2 - U_{dr}F_{dr}) \text{vec}(\boldsymbol{\kappa}\mathbf{r}_{it}' + \mathbf{r}_{it}\boldsymbol{\kappa}')' + (U_{dr}F_{nr} - U_{nr}F_{dr}) \text{vec}(\Sigma_{\kappa})' \right] (\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})' L'_m. \tag{F-37}$$

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

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

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

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

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

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

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

$$\begin{aligned}
\frac{\Sigma_{\epsilon} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa} \partial (\hat{\alpha}_i)}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} &= \frac{\Sigma_{\epsilon} \Sigma_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa}}{2U_{dr}^2} \sum_{t=1}^p \left[(U_{nr}^2 - U_{dr}F_{dr}) \text{vec}(\boldsymbol{\kappa}\mathbf{r}_{it}' + \mathbf{r}_{it}\boldsymbol{\kappa}')' \right. \\
&\quad \left. + (U_{dr}F_{nr} - U_{nr}F_{dr}) \text{vec}(\Sigma_{\kappa})' \right] (\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})' L'_m.
\end{aligned} \tag{F-41}$$

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

$$\begin{aligned} \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1} \hat{\epsilon}_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})'} &= \left[(\epsilon'_{it} \Sigma_{\epsilon\epsilon}^{-1} \otimes I_m) \text{vec}((\text{dg}(\Sigma_{\epsilon\epsilon}))^{-1/2})' - (\epsilon_{it} \otimes \Sigma'_{\epsilon})' (\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})' \right] L'_m \\ &- \frac{\tilde{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\kappa}}{2(U_{dr})^2} \sum_{t=1}^p \left[(U_{nr}^2 - U_{dr} F_{dr}) \text{vec}(\boldsymbol{\kappa} \mathbf{r}'_{it} + \mathbf{r}_{it} \boldsymbol{\kappa}')' + (U_{dr} F_{nr} - U_{nr} F_{dr}) \text{vec}(\Sigma_{\kappa})' \right] (\Sigma_{\epsilon\epsilon}^{-1} \otimes \Sigma_{\epsilon\epsilon}^{-1})' L'_m. \end{aligned} \quad (\text{F-42})$$

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

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

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

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

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

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

F.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 (F-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

$$\begin{aligned} \frac{\partial \mathcal{L}_{i(p)3\Theta_3}}{\partial \Theta'_j} &= \sum_{t=1}^p \frac{\partial \mathcal{L}_{it3\Theta_3}}{\partial \Theta'_j} = \sum_{t=1}^p \frac{\partial}{\partial \Theta'_j} s_{it} [\mathbb{X}_{it}^r(\Theta_1, \hat{\Theta}_2)(r_{it} - \mathbb{X}_{it}^r(\cdot)' \Theta_3)] \\ &= \sum_{t=1}^p s_{it} \left[\frac{\mathbb{X}_{it}^r(\cdot)}{\partial \Theta'_j} (r_{it} - \mathbb{X}_{it}^r(\cdot)' \hat{\Theta}_3) + \mathbb{X}_{it}^r(\hat{\Theta}_1, \hat{\Theta}_2) \frac{\mathbb{X}_{it}^r(\cdot)'}{\partial \Theta'_j} \hat{\Theta}_3 \right] \quad j \in \{1, 2\}, \end{aligned} \quad (\text{F-45})$$

where

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

And $\mathcal{X}_{it}^r = \{\mathcal{X}_{1it}^{r'}, \mathcal{X}_{0it}^{r'}\}'$ where \mathcal{X}_{1it}^r and \mathcal{X}_{0it}^r have been defined in equation (6) in the main text.

We know that $\frac{\mathcal{X}_{it}^r}{\Theta'_1} = \frac{\mathcal{X}_{it}^r}{\Theta'_2} = \mathbf{0}$, that $\frac{\bar{Z}'_i \bar{\delta} + \hat{\alpha}_i}{\Theta'_2} = \frac{\Sigma_{\epsilon\epsilon}^{-1} \hat{\epsilon}_{it}}{\Theta'_2} = \mathbf{0}$ and $\frac{\bar{Z}'_i \bar{\delta} + \hat{\alpha}_i}{\Theta'_1}$ and $\frac{\Sigma_{\epsilon\epsilon}^{-1} \hat{\epsilon}_{it}}{\Theta'_1}$ 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

$$\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \Theta'_1} = \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \varphi_{it}} \frac{\partial \mathbb{X}_{it}^{f'}}{\partial \Theta'_1} + \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \gamma_{it}} \frac{\partial \mathbb{X}_{it}^{s'}}{\partial \Theta'_1}, \quad j \in \{0, 1\}, k \in \{1, 2\}. \quad (\text{F-46})$$

Given the functional form of $C_{jk}(\Theta_1, \Theta_2)$ in equations (24) and (25), its derivative with respect to φ_{it} and γ_{it} can be easily obtained. The partial derivatives $\frac{\partial \mathbb{X}_{it}^{f'}}{\partial \Theta'_1}$ and $\frac{\partial \mathbb{X}_{it}^{s'}}{\partial \Theta'_1}$ have been worked out above. Now consider the derivative of $C_{jk}(\Theta_1, \Theta_2)$ with respect to $\Theta_2 = \{\Theta'_{2f}, \Theta'_{2s}, \rho_{\zeta\bar{v}}\}'$.

$$\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \Theta'_2} = \begin{bmatrix} \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \varphi_{it}} \frac{\partial \mathbb{X}_{it}^{f'}}{\partial \Theta_{2f}} \\ \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \gamma_{it}} \frac{\partial \mathbb{X}_{it}^{s'}}{\partial \Theta_{2s}} \\ \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \rho_{\zeta\bar{v}}} \end{bmatrix}' = \begin{bmatrix} \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \varphi_{it}} \mathbb{X}_{it}^f \\ \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \gamma_{it}} \mathbb{X}_{it}^s \\ \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \rho_{\zeta\bar{v}}} \end{bmatrix}' \quad (\text{F-47})$$

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. If only to note, depending on the particular combination of j and k , the derivatives stated above involve taking derivatives of $\Pr(f_{it} = 1, s_{it} = 1)$ and $\Pr(f_{it} = 0, s_{it} = 1)$ with respect to φ_{it} , γ_{it} and $\rho_{\zeta\bar{v}}$, and these are stated in Greene (2002).

References

- BIØRN, E. (2004): "Regression Systems for Unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure," *Journal of Econometrics*, 122, 281–291.
- GREENE, W. H. (2002): *Econometric Analysis*. Prentice Hall, Upper Saddle River, 4th edn.
- LUTKEPOHL, H. (1996): *Handbook of Matrices*. Wiley, Chichester.
- MAGNUS, J.R., H. N. (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.
- NEWBY, W. K. (1984): "A Method of Moment Interpretation of Sequential Estimators," *Economics Letters*, 14, 201–206.