Microeconometric evidence of financing frictions and innovative activity

Citation for published version (APA):

Document status and date:
Published: 01/01/2012

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

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

Link to publication

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

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

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:
repository@maastrichtuniversity.nl
providing details and we will investigate your claim.

Download date: 24 Aug. 2019
#2012-062

Microeconometric evidence of financing frictions and innovative activity
Amaresh K Tiwari, Pierre Mohnen, Franz C. Palm, Sybrand Schim van der Loeff

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

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

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499
Microeconometric Evidence of Financing Frictions and Innovative Activity *

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

Abstract

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.

JEL Classification: G30, O30, C30
Key Words: Financial Constraints, Capital Structure, R&D, Innovation, Firm Characteristics, Panel Data, Correlated Random Effects, Control Function, Expected a Posteriori.

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

*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
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 suggest 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 condi-
tional 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:

\[ r_{it} = f_{it} \beta f + z_{it}^r \beta_1 + x_{it}^r \beta_1 + \mu_1 \hat{\alpha}_i + \eta_{1it} + (1 - f_{it})(z_{it}^r \beta_0 + x_{it}^r \beta_0 + \mu_0 \hat{\alpha}_i + \eta_{0it}) \text{ if } s_{it} = 1 \]

\[ f_{it} = 1(f^*_{it} = z_{it}^f \varphi + x_{it}^f \tilde{\varphi} + \lambda \hat{\alpha}_i + \xi_{it} > 0), \]

\[ s_{it} = 1(s^*_{it} = z_{it}^s \gamma + x_{it}^s \tilde{\gamma} + \theta \hat{\alpha}_i + \nu_{it} > 0), \]

\[ x_{it} = Z_{it}^x \delta + \tilde{\alpha}_i \kappa + \epsilon_{it}, \]

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 \( \hat{\alpha}_i \), \( z_{it}^r \) is a vector of exogenous variables and \( x_{it}^r \), is a vector of endogenous variables. To allow for the effect of \( z_{it}^r \) and \( x_{it}^r \) 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, 

\[ r_{it} = r_{1it} = \beta_f + z_{it}^f \beta_1 + x_{it}^f \beta_2 + \mu_1 \alpha_i + \eta_{1it} \]

if \( f_{it} = 1 \) and 

\[ r_{it} = r_{0it} = z_{it}^0 \beta_0 + x_{it}^0 \beta_3 + \mu_0 \alpha_i + \eta_{0it} \]

if \( f_{it} = 0 \) and \( s_{it} = 1 \). \( \eta_{1it} \) is the idiosyncratic term of the R&D equation for the firm that is financially constrained, regime 1, and \( \eta_{0it} \) 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 \( \tilde{\alpha}_i, z_{it}^f \) and \( z_{it}^s \) are vectors of exogenous variables in the financial constraint and selection equations respectively. \( x_{it}^f \) and \( x_{it}^s \) 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 \( x_{it} = (x_{1it}, \ldots, x_{mit})' \), where every component of \( x_{1it}, x_{2it}, \ldots, x_{mit} \) is also a component of \( x_{it} \). \( Z_{it} = \text{diag}(z_{1it}, \ldots, z_{mit}) \) is the matrix of exogenous variables or instruments appearing in each of the \( m \) reduced form equations in (4) and \( \delta = (\delta_1', \ldots, \delta_m')' \). Let \( z_{it} \) be the union of all exogenous variables appearing in each of \( z_{1it}, z_{2it}, \ldots, z_{mit} \). For every \( l \in (1, \ldots, m) \), \( z_{lit} = Z_{it} = (z_{1it}, z_{2it}')' \), where the dimension of vector of instruments, \( z \), is greater than or equal to the dimension \( x \). Also define \( \mathbf{X}_i = (x_{1i}, \ldots, x_{mi})' \) and \( Z_i = (Z_{i1}, \ldots, Z_{mi})' \).

\( \mathbf{e}_{it} = (e_{1it}, \ldots, e_{mit})' \) 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 \( Z_i \), but is assumed to be mean independent of \( \eta_{it}, \zeta_{it}, v_{it} \), and \( \mathbf{e}_{it} \). Also, conditional on \( \tilde{\alpha}_i, Z_i \) is independent of \( \eta_{it}, \zeta_{it}, v_{it} \), and \( \mathbf{e}_{it} \). 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, \( (\kappa_1, \ldots, \kappa_m) \), that appear in the reduced form equations, and \( \mu_1, \mu_0, \lambda, \) and \( \theta \) that appear in the structural equations (1), (2), and (3).

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

\[ \mathbf{y}_{it}^s = X_{it}^0 \mathbf{B} \tilde{\alpha}_i + \mathbf{k} + \Upsilon_{it}, \]  

(5)

where \( \mathbf{y}_{it}^s \), suppressing the \( it \) subscript, is \( \mathbf{y}^s = (sf_{r1}, s(1-f)r_0, f*, s*)' \), \( X_{it} = \text{diag}(X_{1it}^c, X_{0it}^c, X_{it}^f, X_{it}^s) \), where \( X_{1it}^c = s_{it} f_{it} \{1, z_{it}^f, x_{it}^f\}', X_{0it}^c = s_{it} (1 - f_{it}) \{z_{it}^s, x_{it}^s\}' \), \( X_{it}^f = \{z_{it}^f, x_{it}^f\}' \), and \( X_{it}^s = \{z_{it}^s, x_{it}^s\}' \). \( \mathbf{B} \) in (5) is given by \( \mathbf{B} = (\beta_1', \beta_0', \phi', \gamma')' \), where \( \beta_0 = (\beta'_0, \beta'_0)' \), \( \phi = (\phi', \phi')' \), and \( \gamma = (\gamma', \gamma')' \). Finally, \( \mathbf{k} = (\mu_1, \mu_0, \lambda, \theta)' \), and \( \Upsilon_{it} = (\eta_{it}, \eta_{0it}, \zeta_{it}, v_{it})' \). The distribution of the error terms of the system of equations (1), (2), (3), and (4) is given by:

\[ \begin{pmatrix} \Upsilon_{it} \\ \mathbf{e}_{it} \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\Upsilon \Upsilon} & \Sigma_{\Upsilon e} \\ \Sigma_{e \Upsilon} & \Sigma_{ee} \end{pmatrix} \right]. \]

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:

\[
E(\alpha_i k + Y_{it} | X_i, Z_i, \alpha_i) = \alpha_i k + E(Y_{it} | Z_i, X_i, \alpha_i) = \tilde{\alpha}_i k + E(Y_{it} | Z_i, \epsilon_i, \alpha_i)
\]

\[
= \tilde{\alpha}_i k + E(Y_{it} | \epsilon_i, \alpha_i) = \tilde{\alpha}_i k + E(Y_{it} | \epsilon_i).
\]

(6)

According to the above, the mean dependence of the structural error term \(Y_{it}\) on the vector of regressors \(X_i, Z_i\), and \(\tilde{\alpha}_i\) is completely characterized by the reduced form error vectors \(\epsilon_i\). The expectation of \(Y_{it}\) given \(\epsilon_i\) is given by

\[
E(Y_{it} | \epsilon_i) = E(Y_{it} | \epsilon_{it}) = \Sigma_{\epsilon} \Sigma_{\epsilon}^{-1} \epsilon_{it} = \Sigma_{\epsilon} \Sigma_{\epsilon}^{-1} \epsilon_{it} = \Sigma_{\epsilon} \Sigma_{\epsilon}^{-1} \epsilon_{it}.
\]

(7)

where the first equality follows from the assumption that conditional on \(\epsilon_{it}\), \(Y_{it}\) is independent of \(\epsilon_{i-1}\). This assumption has also been made in Papke and Wooldridge (2008), and Semykina and Wooldridge (2010). The \((4 \times m)\) matrices \(\Sigma_{\epsilon}\) in the fourth equality is

\[
\Sigma_{\epsilon} = \begin{pmatrix}
\rho_{\eta_{11}} \sigma_{\eta_1} & \cdots & \rho_{\eta_{1m}} \sigma_{\eta_1} \\
\rho_{\eta_{21}} \sigma_{\eta_2} & \cdots & \rho_{\eta_{2m}} \sigma_{\eta_2} \\
\rho_{\xi_{11}} \sigma_{\xi_1} & \cdots & \rho_{\xi_{1m}} \sigma_{\xi_1} \\
\rho_{\xi_{21}} \sigma_{\xi_2} & \cdots & \rho_{\xi_{2m}} \sigma_{\xi_2}
\end{pmatrix}
\]

and the \((m \times m)\) matrix \(\Sigma_{\epsilon}\) is diag(\(\sigma_{\epsilon_1}, \ldots, \sigma_{\epsilon_m}\)), so that \(\Sigma_{\epsilon} \Sigma_{\epsilon}^{-1} = \Sigma_{\epsilon}\). Finally, in the last equality \(\Sigma_{\epsilon}^{-1} = \Sigma_{\epsilon} \Sigma_{\epsilon}^{-1}\). We prefer to write the above conditional expectation as \(E(Y_{it} | \epsilon_{it}) = \Sigma_{\epsilon} \Sigma_{\epsilon}^{-1} \epsilon_{it}\) since the elements of \(\Sigma_{\epsilon}\) 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 \(\Sigma_{\epsilon}\), which are estimated in the subsequent stages that give us the test of exogeneity of \(x_{it}\) with respect to \(Y_{it}\).

Given (7) we can write the linear projections of \(\tilde{\alpha}_i k + Y_{it}\) in error form, given \(\tilde{\alpha}_i\) and \(\epsilon_{it}\) as

\[
\tilde{\alpha}_i k + Y_{it} = \tilde{\alpha}_i k + \Sigma_{\epsilon} \Sigma_{\epsilon}^{-1} \epsilon_{it} + Y_{it}
\]

(8)

where \(\tilde{Y}_{it}\) is independent of \(Z_i, X_i, \epsilon_{it}\), and \(\tilde{\alpha}_i\), and is normally distributed with expectation 0. The above then implies that the projections of \(y^*_{it}\) in error form given \(\tilde{\alpha}_i\) and \(\epsilon_{it}\) is given by

\[
y^*_{it} = X'_i B + \tilde{\alpha}_i k + \Sigma_{\epsilon} \Sigma_{\epsilon}^{-1} \epsilon_{it} + \tilde{Y}_{it}.
\]

(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 \(Z_i\) are correlated in order to estimate \(\beta, \Sigma_{\epsilon}\), and \(\sigma_{\alpha}\) consistently, we use Mundlak’s (1978) correlated random effects formulation. We assume that

\[
\tilde{\alpha}_i = Z'_i \beta + \alpha_i,
\]

(10)
where $\bar{Z}_t$, is the mean of time-varying variables in $Z_{it}$. We also assume that $\alpha_i$ is normally distributed with mean 0 and variance $\sigma_\alpha^2$. Given the above, equation (4) can now be written as

$$x_{it} = Z_{it}'\delta + (\bar{Z}_t'\delta + \alpha_i)k + \epsilon_{it}. \tag{4a}$$

To consistently estimate the reduced form parameters, $\Theta_1 = \{\delta', \bar{Z}_t'\delta, \text{vech}(\Sigma_{ee})', k', \sigma_\alpha\}'$, 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 $\Sigma_{\alpha} = \{\alpha_{1i}, \ldots, \alpha_{mi}\}'$, where each of the $\alpha_{li}$, $l \in \{1, \ldots, m\}$, is unrestricted, we place the restriction $\alpha_{li} = \kappa_i \alpha_i$. This implies that

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

Moreover, as can be seen from the modified equation (4a), we also impose the restriction that $\delta$ 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 $y_{it}^\ast$ given $X_i$, $Z_i$, and $\alpha_i$ in error form as

$$y_{it}^\ast = X_{it}'B + (\bar{Z}_t'\delta + \alpha_i)k + \bar{\Sigma}_{\hat{\epsilon}_{te}}\bar{\Sigma}_{\hat{\epsilon}_{te}}^{-1}\epsilon_{it} + \hat{\gamma}_{it}, \tag{11}$$

To estimate the system of equations in (11) the standard technique is to replace $\epsilon_{it}$ by the residuals from the first stage reduced form regression, here equation (4a). However, the residuals $x_{it} - E(x_{it}|Z_i, \alpha_i) = x_{it} - Z_{it}'\delta - (\bar{Z}_t'\delta + \alpha_i)k$, remain unidentified because the $\alpha_i$‘s are unobserved even though the reduced form parameters, $\delta$, $\delta$, and $k$, 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

$$E(y_{it}^\ast | X_i, Z_i) = \int E(y_{it}^\ast | X_i, Z_i, \alpha_i) f(\alpha_i| X_i, Z_i) d\alpha_i = X_{it}'B + (\bar{Z}_t'\delta + \hat{\alpha}_i)k + \bar{\Sigma}_{\hat{\epsilon}_{te}}\bar{\Sigma}_{\hat{\epsilon}_{te}}^{-1}\epsilon_{it}, \tag{12}$$

where $\hat{\alpha}_i = E(\alpha_i| X_i, Z_i)$, as discussed in Appendix A, is the Expected a Posteriori (EAP) value of $\alpha_i$ and $\epsilon_{it} = x_{it} - E(x_{it}|X_i, Z_i) = x_{it} - Z_{it}'\delta - k(\bar{Z}_t'\delta + \hat{\alpha}_i)$. $\hat{\alpha}_i$ and $\epsilon_{it}$ are the “control functions” that correct for the bias when estimating the structural equations, which arises due to the correlation of $x_{it}$ with the unobserved individual effect $\alpha_i$ and with the idiosyncratic component $\hat{\gamma}_{it}$. The correlation of the exogenous variables $Z_{it}$ with the unobserved individual effect, $\hat{\alpha}_i$, is accounted by $\bar{Z}_t'\delta + \hat{\alpha}_i$. In Appendix A we show how to construct the control functions $\hat{\alpha}_i$ and $\epsilon_{it}$. Given (12) we can write the projection of $y_{it}^\ast$ given $X_i$, $Z_i$ in error form.
as
\begin{align*}
    r_{it} &= f_{it}(\lambda'_{it} \beta_1 + (Z'_{it} \delta + \tilde{\alpha}_i) + \Sigma_{\eta_{it}}^{-1} \tilde{\eta}_{it}) \\
    &+ (1 - f_{it})(\lambda'_{it} \beta_0 + (Z'_{it} \delta + \tilde{\alpha}_i)) + \Sigma_{\eta_{it}}^{-1} \tilde{\eta}_{it} + \tilde{\eta}_{not} \) \\
    f_{it}^* &= \lambda_{it}^{\alpha'} \gamma + (Z_{it} \delta + \tilde{\alpha}_i) \lambda + \Sigma_{\xi_{it}}^{-1} \tilde{\xi}_{it} + \tilde{\xi}_{it}, \quad (14) \\
    s_{it}^* &= \lambda_{it}^{\gamma} \gamma + (Z_{it} \delta + \tilde{\alpha}_i) \theta + \Sigma_{\nu_{it}}^{-1} \tilde{\nu}_{it} + \tilde{\nu}_{it}, \quad (15)
\end{align*}

where \( \tilde{Y}_{it} = (\tilde{\eta}_{it}, \tilde{\eta}_{not}, \tilde{\xi}_{it}, \tilde{\xi}_{it})' \), defined in Appendix A, is normally distributed with mean 0, variance \( \Sigma_{\tilde{Y}_{it}} \) and is independent of \( X_i \) and \( Z_i \).

Given \( Z_{it} \delta + \tilde{\alpha}_i \) and \( \tilde{\epsilon}_{it} \), while it may be possible to consistently estimate the structural parameters of interest by specifying a joint likelihood for \( r_{it}, r_{not}, f_{it}, \) 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 \( Z_{it} \delta + \tilde{\alpha}_i \) and \( \tilde{\epsilon}_{it} \), we estimate jointly the structural parameters, \( \Theta_2 \), 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 \( X_i, Z_i \), if the time period \( t \) corresponds to CIS3 and CIS3.5, is given by
\begin{align*}
    L_{it2}(\Theta_2 | X_i, Z_i, \tilde{\alpha}_i, \tilde{\epsilon}_{it}) &= f_{it} s_{it} \ln(\Phi_2(\varphi_{it}, \gamma_{it}, \rho_{\xi C})) \\
    &+ f_{it}(1 - s_{it}) \ln(\Phi_2(-\varphi_{it}, -\gamma_{it}, -\rho_{\xi C})) \\
    &+ (1 - f_{it}) s_{it} \ln(\Phi_2(-\varphi_{it}, -\gamma_{it}, -\rho_{\xi C})) \\
    &+ (1 - f_{it})(1 - s_{it}) \ln(\Phi_2(-\varphi_{it}, -\gamma_{it}, -\rho_{\xi C})), \quad (16)
\end{align*}

where \( \Theta_2 = (\varphi', \lambda, \tilde{\Sigma}_{\xi}, \gamma', \theta, \tilde{\Sigma}_{\nu}, \rho_{\xi C}') \), \( \Phi_2 \) is the cumulative distribution function of a standard bivariate normal, \( \rho_{\xi C} \) is the correlation of \( \tilde{\xi}_{it} \) and \( \tilde{\xi}_{it} \), and
\begin{align*}
    \varphi_{it} &= (\lambda'_{it} \varphi + \lambda(Z_{it} \delta + \tilde{\alpha}_i) + \tilde{\Sigma}_{\xi}^{-1} \tilde{\xi}_{it}) \frac{1}{\sigma_{\xi}} \\
    \gamma_{it} &= (\lambda_{it}^{\alpha'} \gamma + \theta(Z_{it} \delta + \tilde{\alpha}_i) + \tilde{\Sigma}_{\nu}^{-1} \tilde{\nu}_{it}) \frac{1}{\sigma_{\nu}}. \quad (17, 18)
\end{align*}

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 equation being respectively \( \sigma_{\xi} \) and \( \sigma_{\nu} \). 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{align*}
    L_{it2}(\Theta_2 | X_i, Z_i, \tilde{\alpha}_i, \tilde{\epsilon}_{it}) &= f_{it} s_{it} \ln(\Phi_2(\varphi_{it}, -\rho_{\xi C})) \\
    &+ (1 - f_{it}) s_{it} \ln(\Phi_2(-\varphi_{it}, -\rho_{\xi C})) \\
    &+ (1 - s_{it}) \ln(\Phi(-\gamma_{it})), \quad (19)
\end{align*}

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

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

(20)

Given the first stage estimates $\hat{\Theta}_1$, we can obtain the estimates of $\hat{\Theta}_2$ 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, $\Theta_3$, 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 $\hat{\Theta}_2$ and $\hat{\epsilon}_{it}$, for each observation, we obtain consistent estimates of $\Theta_3$.

To this effect, consider the following conditional mean:

$$
E(r_{it} | f^*_{it}, s^*_{it} > 0, X_i, Z_i) = f_{it} \left( \beta_f + X^t_{it} \beta_1 + \mu_1 (Z^t_\delta + \hat{\alpha}_i) + \hat{\Sigma}_{\eta_{it}} \hat{\Sigma}_{\epsilon_{it}}^{-1} \hat{\epsilon}_{it} \right)
+ (1 - f_{it}) \left( X^t_{it} \theta_0 + \mu_0 (Z^t_\delta + \hat{\alpha}_i) + \hat{\Sigma}_{\eta_{it}} \hat{\Sigma}_{\epsilon_{it}}^{-1} \hat{\epsilon}_{it} \right)
+ f_{it} E(\tilde{\eta}_{it} | f^*_{it} > 0, s^*_{it} > 0, X_i, Z_i) + (1 - f_{it}) E(\tilde{\eta}_{it} | f^*_{it} \leq 0, s^*_{it} > 0, X_i, Z_i)
$$

(21)

Now, we know that

$$
E(\tilde{\eta}_{it} | f^*_{it} > 0, s^*_{it} > 0, X_i, Z_i) = E[\tilde{\eta}_{it} | \tilde{\zeta}_{it} > -\varphi_{it}, \tilde{v}_{it} > -\gamma_{it}],
$$

and

$$
E(\tilde{\eta}_{it} | f^*_{it} \leq 0, s^*_{it} > 0, X_i, Z_i) = E[\tilde{\eta}_{it} | \tilde{\zeta}_{it} \leq -\varphi_{it}, \tilde{v}_{it} > -\gamma_{it}],
$$

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

$$
E[\tilde{\eta}_{it} | \tilde{\zeta}_{it} > -\varphi_{it}, \tilde{v}_{it} > -\gamma_{it}] = \Phi \left( \frac{(\gamma_{it} - \rho_{\tilde{\zeta}_{it}} \varphi_{it})}{\sqrt{1 - \rho^2_{\tilde{\zeta}_{it}}} \Phi_2(\varphi_{it}, \gamma_{it}, \rho_{\tilde{\zeta}_{it}})} \right)
+ \Phi \left( \frac{(\varphi_{it} - \rho_{\tilde{\zeta}_{it}} \gamma_{it})}{\sqrt{1 - \rho^2_{\tilde{\zeta}_{it}}} \Phi_2(\varphi_{it}, \gamma_{it}, \rho_{\tilde{\zeta}_{it}})} \right)
$$

(22)
and

\[
E[\eta_{it}|\tilde{c}_t \leq -\phi_{it}, \tilde{v}_t > -\gamma_t] = -\sigma_{\eta_{it}}\rho_{\eta_{it}}\tilde{\phi}(\phi_{it}) \Phi\left(\frac{\gamma_t - \rho_{zz}\phi_{it}}{\sqrt{1 - \rho_{zz}^2}}\right) \Phi_2(-\phi_{it}, \gamma_t, -\rho_{zz}) \\
+ \sigma_{\eta_{it}}\rho_{\eta_{it}}\tilde{\phi}(\phi_{it}) \Phi(\phi_{it} + \rho_{zz}\gamma_t) / \left(1 - \rho_{zz}^2\right) \Phi_2(-\phi_{it}, \gamma_t, -\rho_{zz}),
\]

(23)

where \(\tilde{\phi}(\cdot)\), \(\Phi(\cdot)\), and \(\Phi_2(\cdot)\), 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

\[
C_{11t} = \phi(\phi_{it}) \Phi\left(\frac{\gamma_t - \rho_{zz}\phi_{it}}{\sqrt{1 - \rho_{zz}^2}}\right) \Phi_2(-\phi_{it}, \gamma_t, -\rho_{zz}), \quad C_{12t} = \phi(\gamma_t) \Phi(-\phi_{it}, \gamma_t, -\rho_{zz}),
\]

\[
C_{01t} = -\phi(\phi_{it}) \Phi_2(-\phi_{it}, \gamma_t, -\rho_{zz}), \quad C_{02t} = \phi(\gamma_t) \Phi_2(-\phi_{it}, \gamma_t, -\rho_{zz}).
\]

Given estimates of \(\hat{\alpha}_i\) and \(\hat{\epsilon}_{it}, \phi_{it}, \gamma_{it}, \rho_{zz}\) we can construct the above control functions. With the above defined, we can now write the R&D switching equations in (14), conditional on \(f_{it}^*, s_{it}^* > 0, X_i, Z_i\) as

\[
r_{it} = f_{it}\left(\beta + X_{it}'\beta_1 + \mu_1(Z_i'\delta + \hat{\alpha}_i) + \sum_{it}^{-1}\Phi_{it} \hat{\epsilon}_{it} + \sigma_{\eta_{it}}\rho_{\eta_{it}}\tilde{\phi}(\phi_{it}) C_{11it} + \sigma_{\eta_{it}}\rho_{\eta_{it}}\tilde{\phi}(\phi_{it}) C_{12it} + \eta_{1it}\right) \\
+ (1 - f_{it})\left(X_{it}'\beta_0 + \mu_0(Z_i'\delta + \hat{\alpha}_i) + \sum_{it}^{-1}\Phi_{it} \hat{\epsilon}_{it} + \sigma_{\eta_{it}}\rho_{\eta_{it}}\tilde{\phi}(\phi_{it}) C_{01it} + \sigma_{\eta_{it}}\rho_{\eta_{it}}\tilde{\phi}(\phi_{it}) C_{02it} + \eta_{2it}\right),
\]

(24)

where \(\eta_{1it}\) and \(\eta_{2it}\) conditional on \(f_{it}^*, s_{it}^*, X_i, Z_i\) is distributed with mean zero. With the additional correction terms \(-C_{11}, C_{12}, C_{01}, C_{02}\) – 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 \(\Theta_1\) been known. To obtain correct inference about the structural parameters, \(\Theta_2\) and \(\Theta_3\), 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 \(X_{it} = \bar{X}\), where \(X_{it}\) is the union of elements appearing in \(X_{ir}, X_{it}, X_{is}\), 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

$$
\Delta_f E(r_{it}|\bar{X}) = \int E(r_{1it}|\bar{X}, f_{it} = 1, s_{it} = 1, \hat{\alpha}, \hat{\epsilon})dG(\hat{\alpha}, \hat{\epsilon})
- \int E(r_{0it}|\bar{X}, f_{it} = 0, s_{it} = 1, \hat{\alpha}, \hat{\epsilon})dG(\hat{\alpha}, \hat{\epsilon}).
$$

(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 $s = 1, 2, \cdots , S$. Then, $\sum_{s=1}^{S} N_s = N$, where $N_s$ is the population size of R&D performing enterprise belonging to stratum $s$. Let $n_s$ be the sample size of each stratum and let $\Theta_s = \{1, 2, \cdots , i_s\}$ 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

$$
\hat{x}_{f} = \sum_{s=1}^{S} \frac{(N_{fs} - n_{fs})\bar{x}_s}{\sum_{s=1}^{S} \sum_{k=1}^{n_{fs}} x_{fsk}},
$$

(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 \( f \). 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>
<thead>
<tr>
<th>Stratum</th>
<th>Total no. of enterprises</th>
<th>Total no. of strata</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIS2.5</td>
<td>13465</td>
<td>240</td>
</tr>
<tr>
<td>CSI3</td>
<td>10750</td>
<td>249</td>
</tr>
<tr>
<td>CIS3.5</td>
<td>10533</td>
<td>280</td>
</tr>
</tbody>
</table>

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 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 of 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,

---

**Proof:**
The proof is based on the assumption that the expected value of \( x \) is the same for each enterprise in a particular stratum. Let \( \mu_{sf} \) be the population mean of \( x \) for the firm \( f \) and let \( \mu_{xs} \) 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 \( \mu_{xs} \), that \( \mu_{sf} = \sum_{s=1}^{S} N_{fs} \mu_{xs} \), 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 \( E(\hat{x}_f) = \mu_{sf} = \sum_{s=1}^{S} N_{fs} \mu_{xs} \).

---

12
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

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

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) \( RAIN_{it} \): 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) \( DNF_C_{it} \): 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) \( AG_{Eit} \): Age of the firm\(^3\).

(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.

Our set of endogenous regressors, \( x_{it} \), 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 (\( RAIN_{it} \)), (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, \( z_{it} \), that appear in the structural and reduced form equation, and which are included in \( Z_{it} \) are: (1) cashflows of the firm (\( CF_{it} \)), (2) a dummy for negative realization of cashflows (\( DNF_C_{it} \)), (3) the market share of the firm (\( MKSH_{it} \)), (4) the age of the firm (\( AG_{Eit} \)), (5) 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.
(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, $z_{it}$, that are also part of $Z_{it}$ 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 $t$ ($CF_{i,t}^2$), (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 $Z_{it}$ 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 $Z_{it}$ 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 finance.

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. 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.

\[\text{constraint equation} \]

\[\text{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.}

\[\text{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.}

\[\text{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.}

\[\text{We find that firms that maintain higher liquidity reserve, (\(LQ\)), are less likely to be constrained. Gamba and Triantis (2008) point out that cash balances, which give financial flexibility to firms, are held when external finance is costly and/or income uncertainty is high. With higher liquidity reserve firms can counter bad shocks by draining it. 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.}

\[\text{4 The paper by Hennessy and Whited (2007) has a detailed discussion on constraint proxies that reflect high cost or high need for external finance.}\]
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 equation\(^5\). This suggests that firms that pay out dividends are less likely to innovate. Now, given the nature of R&D activity that makes borrowing costly, internal funds may be more preferable. Therefore, innovative firms, ceteris paribus, are less likely to distribute cash as dividends.

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

\(^5\) As 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.
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 scale in the R&D process through their merger, and become more active R&D performers after being merged or acquired.

The factor loadings, $\lambda$ and $\theta$, which are the coefficients of $\bar{Z}_i \delta + \hat{\alpha}_i$ 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 $\rho_{\varepsilon \mu}$, 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 Quadrini (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, \( f_{i}^{*} \), underlying \( f_{i} \) 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.

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 - \( C11, C12, C01, C02 \) - 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, \( \Delta_{f}E(r_{it}|\bar{X}) \), 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 $\bar{Z}'\bar{\delta} + \hat{\alpha}_i$ 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 under 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 it not a preferred means to finance innovative activity, firms do show a propensity to engage in innovative activity by raising debt. However, this propensity is influenced both by the incentives to innovate and the capacity to raise debt; both of which vary with firm characteristics.
These findings draw our attention to the fact that innovation and financing policy are not independent of firms dynamics of survival, exit, and growth.

The results of the third stage R&D switching regression imply that financial 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


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

We began with a set of structural equations

\[ y_{it}^* = X_{it}'B + \alpha_i k + \Upsilon_{it}, \]  

(A-1)

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

\[ x_{it} = Z_{it}'\delta + \alpha_i k + \epsilon_{it}, \]  

(A-2)

a specification for the conditional expectation and distribution of the individual effects \( \bar{\alpha}_i \),

\[ \bar{\alpha}_i = \bar{Z}_i'\bar{\delta} + \alpha_i, \]  

(A-3)

and specification of the joint distribution of idiosyncratic components \( \Upsilon_{it} \) and \( \epsilon_{it} \). 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

\[ y_{it}^* = X_{it}'B + (\bar{Z}_i'\bar{\delta} + \alpha_i)k + \bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}\epsilon_{it} + \bar{\Upsilon}_{it}. \]  

(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 \( \epsilon_{it} \) by the residuals from the first stage reduced form regression. However, the residuals \( x_{it} - E(x_{it} | Z_i, \alpha_i) = x_{it} - Z_{it}'\delta - (\bar{Z}_i'\bar{\delta} + \alpha_i)k \), remain unidentified because the \( \alpha_i \)'s are unobserved even though the reduced form parameters, \( \delta, \bar{\delta}, \) and \( \kappa \), 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 \( \alpha_i \)'s. But given that \( \alpha_i \)'s are correlated with the endogenous regressors we have to integrate it out with respect to its conditional distribution. Given (A-4), we have

\[ E(y_{it}^* | X_i, Z_i, \alpha_i) = X_{it}'B + (\bar{Z}_i'\bar{\delta} + \alpha_i)k + \bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}\epsilon_{it}. \]

Let \( f(\alpha_i | X_i, Z_i) \) be the conditional distribution of time invariant individual effect \( \alpha_i \) conditional on \( X_i \) and \( Z_i \). For any individual, \( i \), taking expectation of the above with respect to the conditional distribution of \( \alpha \), \( f(\alpha_i | X_i, Z_i) \) we obtain

\[
E(y_{it}^* | X_i, Z_i) = \int E(y_{it}^* | X_i, Z_i, \alpha_i)f(\alpha_i | X_i, Z_i)d\alpha_i
\]

\[
= X_{it}'B + \bar{Z}_i'\bar{\delta}k + \bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}(x_{it}-Z_{it}'\delta-\bar{Z}_i'\bar{\delta}k) + \int (k-\bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}\kappa)\alpha_i f(\alpha_i | X_i, Z_i)d\alpha_i
\]

\[
= X_{it}'B + \bar{Z}_i'\bar{\delta}k + \bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}(x_{it}-Z_{it}'\delta-\bar{Z}_i'\bar{\delta}k) + \int (k-\bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}\kappa)\alpha_i f(\alpha_i | X_i)d\alpha_i
\]

\[
= X_{it}'B + \bar{Z}_i'\bar{\delta}k + \bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}(x_{it}-Z_{it}'\delta-\bar{Z}_i'\bar{\delta}k) + (k-\bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}\kappa)\bar{\alpha}_i
\]

\[
= X_{it}'B + (\bar{Z}_i'\bar{\delta} + \alpha_i)k + \bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}(x_{it}-Z_{it}'\delta-\bar{Z}_i'\bar{\delta}k)
\]

\[
= X_{it}'B + (\bar{Z}_i'\bar{\delta} + \alpha_i)k + \bar{\Sigma}_{\Upsilon_e}\bar{\Sigma}_{\epsilon e}^{-1}\epsilon_{it}, \quad \text{(A-5)}
\]
where the second equality follows from the fact that \( \mathcal{Z}_t \) and \( \alpha_i \) are independent. \( \hat{\alpha}_i = \hat{\alpha}_i(x_i, \mathcal{Z}_i, \Theta_1) \) is the expected a posteriori (EAP) value of time invariant individual effects \( \alpha_i \), and \( \Theta_1 \) is the set of first stage reduced form parameters.

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

\[
f(\alpha|X, \mathcal{Z}) = \frac{f(X|\alpha)g(\alpha)}{h(X, \mathcal{Z})} = \frac{f(X, \mathcal{Z}|\alpha)g(\alpha)}{h(X, \mathcal{Z})}, \tag{A-6}
\]

where \( g \) and \( h \) are density functions. The above can be written as

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

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

\[
f(\alpha|X, \mathcal{Z}) = \frac{f(X|\mathcal{Z}, \alpha)g(\alpha)}{h(X|\mathcal{Z})} = \frac{f(X|\mathcal{Z}, \alpha)g(\alpha)}{\int f(X|\mathcal{Z}, \alpha)g(\alpha)d\alpha}, \tag{A-7}
\]

Hence,

\[
\int \alpha f(\alpha|X, \mathcal{Z})d(\alpha) = \int \alpha f(X|\mathcal{Z}, \alpha)g(\alpha)d\alpha \frac{g(\alpha)}{\int f(X|\mathcal{Z}, \alpha)g(\alpha)d\alpha}
= \frac{\int \alpha \prod_{t=1}^T f(x_t|\mathcal{Z}, \alpha)g(\alpha)d\alpha}{\int \prod_{t=1}^T f(x_t|\mathcal{Z}, \alpha)g(\alpha)d\alpha}
= \hat{\alpha}(X, \mathcal{Z}_i, \Theta_1) \tag{A-8}
\]

where the second equality follow from the fact that conditional on \( \mathcal{Z} \) and \( \alpha \), each of the \( x_t, x_t \in \{x_1, \ldots, x_T\} \) are independently normally distributed with mean \( \mathcal{Z}_t^{\prime}\delta + (\mathcal{Z}_t^{\prime}\delta + \alpha)\kappa \) and standard deviation \( \Sigma_{\epsilon t} \). \( g(\alpha) \) by our assumption is normally distributed with mean zero and variance \( \sigma_{\alpha}^2 \) and \( \alpha = \frac{\delta}{\sigma_{\alpha}} \) follows a standard normal distribution. The functional form of \( \hat{\alpha}_i(x_i, \mathcal{Z}_i, \Theta_1) \) is given by:

\[
\hat{\alpha}_i(x_i, \mathcal{Z}_i, \Theta_1) = \frac{\int \sigma_{\alpha} \alpha \prod_{t=1}^T \exp(-\frac{1}{2}(x_t - \mathcal{Z}_t^{\prime}\delta - (\mathcal{Z}_t^{\prime}\delta + \sigma_{\alpha}a)\kappa)^T\Sigma_{\epsilon t}^{-1}(x_t - \mathcal{Z}_t^{\prime}\delta - (\mathcal{Z}_t^{\prime}\delta + \sigma_{\alpha}a)\kappa))\phi(a)d\alpha}{\int \prod_{t=1}^T \exp(-\frac{1}{2}(x_t - \mathcal{Z}_t^{\prime}\delta - (\mathcal{Z}_t^{\prime}\delta + \sigma_{\alpha}a)\kappa)^T\Sigma_{\epsilon t}^{-1}(x_t - \mathcal{Z}_t^{\prime}\delta - (\mathcal{Z}_t^{\prime}\delta + \sigma_{\alpha}a)\kappa))\phi(a)d\alpha} \tag{A-9}
\]

The right hand side of (A-9) is the expected a posteriori value of \( \alpha_i \). \( \hat{\alpha}_i(x_i, \mathcal{Z}_i, \Theta_1) \) is the estimated expected a posteriori value of \( \alpha_i \), which can be estimated by employing numerical integration techniques, such as Gauss-Hermite quadratures, with respect to \( \alpha \) at the estimated \( \Theta_1 \) from the first stage. Also, it can be shown that

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

**Proof of Lemma 1** Given in Section A.1.
Lemma 1 implies that
\[
X_{it}'B + (Z_i'\delta + \hat{\alpha}_i)k + \bar{\Sigma}_{\zeta}^{-1}\hat{\epsilon}_{it} \xrightarrow{a.s.} E(y^*_i|X, Z) = \int E(y^*_i|X, Z, \alpha)f(\alpha|X, Z)d(\alpha).
\]

Therefore, if the reduced form population parameters, \(\Theta_1\), are known, the above implies that we could write the linear predictor of \(y^*_i\), given \(X_i\) and \(Z_i\) in error form as
\[
y^*_i = X_{it}'B + (Z_i'\delta + \hat{\alpha}_i)k + \bar{\Sigma}_{\zeta}^{-1}\hat{\epsilon}_{it} + \tilde{\gamma}_{it}, \tag{A-10}
\]
where conditional of \(X_i\) and \(Z_i\), \(\tilde{\gamma}_{it}\) is i.i.d. with mean 0. For linear models, say if all the variables in \(y^*_i\) were continuous and observed, with estimates of expected a posteriori values, \(\hat{\Sigma}_{\zeta}^{-1}\hat{\epsilon}_{it}\), and the estimates of \(\bar{\Sigma}_{\zeta}^{-1}\hat{\epsilon}_{it}\) the parameters of interest, \(B\), can be consistently estimated by running a seemingly unrelated regression (SUR) or a panel version of SUR to gain efficiency. We note here that for any \(k, k \in \{1, \ldots, n\}, n = 4\) in our model, \(\bar{\Sigma}_{\zeta}^{-1}\hat{\epsilon}_{it}\) take the form
\[
\rho T_{k't}(\Sigma_{\zeta}, \hat{\epsilon}_{i1t}, \ldots, \hat{\epsilon}_{mit}) + \ldots + \rho T_{k'm}(\Sigma_{\zeta}, \hat{\epsilon}_{i1t}, \ldots, \hat{\epsilon}_{mit})
\]
where each of the \(f\)’s above are linear in \(\hat{\epsilon}_{it}\). The estimates \(\rho T_{k't}, \rho T_{k'm}, l \in \{1, \ldots, m\}\), provides us with a test of exogeneity of the regressor \(x_l\) with respect to \(\gamma_k\).

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 \(f^*_i\) of \(y^*_i\) where \(f^*_i\) is the latent variable underlying \(f_i\), the binary variable that takes value 1 when the firm is financially constrained and 0 otherwise. Also, let, given the first stage population parameters \(\Theta_1\), the linear predictor of \(f^*_i\), given \(X_i\) and \(Z_i\), in error form given by (A-10) be
\[
f^*_i = \lambda_{it}' \varphi + \lambda(Z_i'\delta + \hat{\alpha}_i) + \bar{\Sigma}_{\zeta}^{-1}\hat{\epsilon}_{it} + \tilde{\zeta}_{it},
\]
where \(\tilde{\zeta}_{it}\) is i.i.d. and normally distributed with mean 0 and variance \(\sigma^2_{\tilde{\zeta}}\). Since in probit models the parameters are identified only up to a scale, for an individual \(i\), the probability of \(f_i = 1\) given \(X_i\) and \(Z_i\) is given by
\[
\Pr(f_i = 1|X, Z) = \Pr(f_i = 1|X, Z, \hat{\epsilon}_i, \hat{\alpha}) = \Pr(y^*_i > 0|X, Z, \hat{\epsilon}_i, \hat{\alpha}) = \Phi\left(\lambda_{it}' \varphi + \lambda(Z_i'\delta + \hat{\alpha}_i) + \bar{\Sigma}_{\zeta}^{-1}\hat{\epsilon}_{it}\right) \left(\frac{1}{\sigma_{\tilde{\zeta}}}\right), \tag{A-11}
\]
where the first equality follows from the fact that \(\hat{\epsilon}_i\) and \(\hat{\alpha}_i\) is a function of \(X\) and \(Z\) and \(\Phi()\) is the cumulative standard normal density function. However, \(\Pr(f_i = 1|X, Z, \hat{\epsilon}_i, \hat{\alpha}) = \Pr(f_i = 1|X, Z, E(\epsilon_i|X, Z), E(\tilde{\alpha}|X, Z))\) is generally not equal to \(\Pr(f_i = 1|X, Z, \epsilon_i, \tilde{\alpha})\). Our measure of interest, however, is \(\int \Pr(f_i = 1|X, Z, \epsilon_i, \tilde{\alpha})d\epsilon_i, d\tilde{\alpha}\), 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
\[
\frac{\Delta \Pr(f_i = 1)}{\Delta z_k} = \int \Pr(f_i = 1|X, Z_{-z_k}, z_{kt - z_k}, (z_{kt} + \Delta z_k), \epsilon_i, \tilde{\alpha})d\epsilon_i, d\tilde{\alpha} - \int \Pr(f_i = 1|X, Z_{-z_k}, z_{kt - z_k}, z_{kt}, \epsilon_i, \tilde{\alpha})d\epsilon_i, d\tilde{\alpha} / \Delta z_k, \tag{A-12}
\]
where \( g(\epsilon_t, \hat{\alpha}) \) is the joint distribution function of \( \epsilon_t \) and \( \hat{\alpha} \). To recover the above measure in (A-15), like Chamberlain (1984), we make an assumption about the conditional distribution for \( \lambda \hat{\alpha}_i \) and \( \tilde{\Sigma}_{\tilde{t}t} \tilde{\Sigma}_{\tilde{t}t}^{-1} \epsilon_t \), where the conditioning variable are \( X_i \) and \( Z_i \). We assume that

\[
\lambda \hat{\alpha}_i = E(\lambda \hat{\alpha}_i | X_i, Z_i) + \tilde{\alpha}_{\tilde{t}} \text{ and } \hat{\Sigma}_{\tilde{t}t} \hat{\Sigma}_{\tilde{t}t}^{-1} \epsilon_{it} = E(\hat{\Sigma}_{\tilde{t}t} \hat{\Sigma}_{\tilde{t}t}^{-1} \epsilon_{it} | X_i, Z_i) + \hat{\epsilon}_{\hat{t}t} \quad (A-13)
\]

where \( \tilde{\alpha}_{\tilde{t}} \) and \( \hat{\epsilon}_{\hat{t}t} \), a scalar, are normally distributed with mean 0, and are independent of everything else. Now, for any individual \( i \), we have shown that

\[
E(\hat{\alpha} | X, Z) = \tilde{Z}' \delta + \int \alpha f(\alpha | X, Z) d(\alpha) = \tilde{Z}' \delta + \hat{\alpha}(X, Z, \Theta_1) \text{ and } E(\hat{\Sigma}_{\tilde{t}t} \hat{\Sigma}_{\tilde{t}t}^{-1} \epsilon_{it} | X_i, Z_i) = \hat{\Sigma}_{\tilde{t}t} \hat{\Sigma}_{\tilde{t}t}^{-1} \hat{\epsilon}_{it}.
\]

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

\[
f_t^* = X_t' \varphi + \lambda(\tilde{Z}' \delta + \hat{\alpha}) + \hat{\Sigma}_{\tilde{t}t} \hat{\Sigma}_{\tilde{t}t}^{-1} \hat{\epsilon}_{it} + \omega_{\tilde{t}t} + \tilde{\epsilon}_{\tilde{t}}, \quad (A-14)
\]

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

Now, having assumed the conditional distribution of \( \lambda \hat{\alpha}_i \) and \( \hat{\Sigma}_{\tilde{t}t} \hat{\Sigma}_{\tilde{t}t}^{-1} \epsilon_{it} \), for any individual \( i \), we now have

\[
\Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}) = \Pr(f_t = 1 | X, Z, \hat{\epsilon}_t, \hat{\alpha}, \omega_{\tilde{t}t})
\]

and

\[
\int \Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}) d(\epsilon_t, \hat{\alpha}) = \int \Pr(f_t = 1 | X, Z, \hat{\epsilon}_t, \hat{\alpha}, \omega_{\tilde{t}t}) dF(\hat{\epsilon}_t, \hat{\alpha}, \omega_{\tilde{t}t}),
\]

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

\[
\int \Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}, \omega_{\tilde{t}t}) dF(\hat{\epsilon}_t, \hat{\alpha}, \omega_{\tilde{t}t}) = \int \Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}, \omega_{\tilde{t}t}) d(h(\omega_{\tilde{t}t}, | \hat{\alpha}, \epsilon_t)) dG(\hat{\epsilon}_t, \hat{\alpha})
\]

\[
= \int \Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}, \omega_{\tilde{t}t}) d(h(\omega_{\tilde{t}t})) dG(\hat{\epsilon}_t, \hat{\alpha})
\]

\[
= \int \Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}) dG(\hat{\epsilon}_t, \hat{\alpha}), \quad (A-15)
\]

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

\[
\int \Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}) d(\epsilon_t, \hat{\alpha}) = \int \Pr(f_t = 1 | X, Z, \epsilon_t, \hat{\alpha}) dG(\epsilon_t, \hat{\alpha})
\]

\[
= \Phi \left( \left\{ \lambda \tilde{Z}' \delta + \hat{\alpha} + \frac{1}{\sigma_{\omega_{\tilde{t}t}}^2 + \sigma_\zeta^2} \right\} \right) dG(\hat{\epsilon}_t, \hat{\alpha}), \quad (A-16)
\]

where \( \sigma_{\omega_{\tilde{t}t}}^2 \) is the variance of \( \omega_{\tilde{t}t} \) and \( \sigma_\zeta^2 \) the variance of \( \zeta \). To obtain the sample analog of RHS of (A-16) for any fixed \( X_t' = \lambda \tilde{Z}' \) we can compute

\[
\frac{1}{\sum_{i=1}^N \sum_{i=1}^T_{T_i} \Pr(f_{it} = 1 | \lambda \tilde{X}_t', \hat{\epsilon}_{it}, \hat{\alpha}_i)}.
\]

(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 \( w_i = \bar{w} \) to \( \bar{w} + \Delta_w \). In the limit when \( \Delta_w \) 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

\[
\frac{\partial \Pr(f_{it} = 1)}{\partial w} = \int \varphi_w \varphi \left( \bar{X}_{it}^f \varphi + \lambda(\bar{Z}_{it}^f + \hat{\alpha}) + \bar{\Sigma}_{\epsilon \epsilon} \bar{\Sigma}_{\epsilon \epsilon}^{-1} \epsilon_{it} \right) dG(\epsilon_{it}, \hat{\alpha}),
\]

(A-18)

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

\[
\frac{\partial \hat{\Pr}(f_{it} = 1)}{\partial w} = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \varphi_w \varphi \left( \bar{X}_{it}^f \varphi + \lambda(\bar{Z}_{it}^f + \hat{\alpha}_i) + \bar{\Sigma}_{\epsilon \epsilon} \bar{\Sigma}_{\epsilon \epsilon}^{-1} \epsilon_{it} \right),
\]

(A-19)

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

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

\[
\int \left[ \Phi \left( \bar{X}_{it}^f \varphi_{-w} + \varphi_w + \lambda(\bar{Z}_{it}^f + \hat{\alpha}) + \bar{\Sigma}_{\epsilon \epsilon} \bar{\Sigma}_{\epsilon \epsilon}^{-1} \epsilon_{it} \right) 
- \Phi \left( \bar{X}_{it}^f \varphi_{-w} + \lambda(\bar{Z}_{it}^f + \hat{\alpha}) + \bar{\Sigma}_{\epsilon \epsilon} \bar{\Sigma}_{\epsilon \epsilon}^{-1} \epsilon_{it} \right) \right] dG(\epsilon_{it}, \hat{\alpha}),
\]

(A-20)

whose sample analog, given \( X_{it-w}^f = \bar{X}_{it-w}^f \), the estimated first and second stage coefficients, \( \hat{\epsilon}_{it} \), and \( \hat{\epsilon}_{it} \), 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 \( \hat{\epsilon}_{it} \) in equations (13) to (15) in the main text is \( \hat{\epsilon}_{it} = \{\hat{\eta}_{it}, \hat{\eta}_{ht}, \hat{\zeta}_{it}, \hat{\nu}_{it}\}' = \{\omega_{\eta_{it}}, \omega_{\eta_{ht}}, \omega_{\hat{\eta}_{ht}}, \omega_{\hat{\eta}_{ht}}, \omega_{\hat{\zeta}_{it}}, \omega_{\hat{\zeta}_{it}}, \omega_{\hat{\nu}_{it}}, \omega_{\hat{\nu}_{it}}\}' \), where \( \omega_{\eta_{it}}, \omega_{\eta_{ht}} \), and \( \omega_{\hat{\eta}_{ht}} \) are defined in the same way as the \( \omega_{\hat{\eta}_{ht}} \) is defined in (A-14) and \( \hat{\epsilon}_{it} = \{\hat{\eta}_{it}, \hat{\eta}_{ht}, \hat{\zeta}_{it}, \hat{\nu}_{it}\}' \) in equations (8) and (9) in the main text.

A.1 Proof of Lemma 1

**Proof:** Let \( \Theta^*_1 \) be true value of first stage reduced form parameters. Now, for an individual \( i \)

\[
\hat{\alpha}(X, Z, \Theta_1) = \int \alpha \exp\left(-\frac{1}{2} r(\Theta_1, \alpha)\right) \phi(\alpha) d\alpha,
\]

where \( r(\Theta_1, \alpha) = \sum_{t=1}^{T} (x_i - Z_i^f + (\bar{Z}_i^f + \alpha) \kappa)^t \Sigma_{\epsilon \epsilon}^{-1} (x_i - Z_i^f + (\bar{Z}_i^f + \alpha)) \kappa) \).

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

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

32
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

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

Given that both the numerator and the denominator in (A-9) defined at $\hat{\Theta}_1$ converge almost surly to the same defined at $\Theta_1^*$, it can now be easily shown that

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

<table>
<thead>
<tr>
<th>CIS2.5</th>
<th>CS13</th>
<th>CIS3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of firms for which</td>
<td>No. of firms for which</td>
<td>No. of firms for which</td>
</tr>
<tr>
<td>( N_f )</td>
<td>( N_f = n_f )</td>
<td>( N_f &gt; n_f )</td>
</tr>
<tr>
<td>1</td>
<td>9400</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>151</td>
<td>1255</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>608</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>316</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>247</td>
</tr>
<tr>
<td>6</td>
<td>149</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>107</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>93</td>
</tr>
<tr>
<td>10</td>
<td>83</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>106</td>
</tr>
<tr>
<td>12</td>
<td>49</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>43</td>
</tr>
<tr>
<td>14</td>
<td>59</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>46</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>62</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>36</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>15</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>34</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>46</td>
<td>26</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>29</td>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td>30</td>
<td>14</td>
<td>29</td>
</tr>
<tr>
<td>31</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>32</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>33</td>
<td>11</td>
<td>32</td>
</tr>
<tr>
<td>34</td>
<td>18</td>
<td>34</td>
</tr>
<tr>
<td>37</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>38</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>43</td>
<td>15</td>
<td>48</td>
</tr>
<tr>
<td>44</td>
<td>17</td>
<td>50</td>
</tr>
<tr>
<td>45</td>
<td>14</td>
<td>57</td>
</tr>
<tr>
<td>48</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>49</td>
<td>22</td>
<td>40</td>
</tr>
<tr>
<td>51</td>
<td>28</td>
<td>53</td>
</tr>
<tr>
<td>56</td>
<td>19</td>
<td>55</td>
</tr>
<tr>
<td>66</td>
<td>33</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>CIS2.5 Innovator</td>
<td>CIS2.5 Non-Innovator</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>R&amp;D*</td>
<td>0.506</td>
<td>0.338</td>
</tr>
<tr>
<td>Share of Innovative Sales in Total Sales (%)</td>
<td>8.532</td>
<td>10.944</td>
</tr>
<tr>
<td>Long-term Debt*</td>
<td>0.789</td>
<td>0.834</td>
</tr>
<tr>
<td>Cashflow*</td>
<td>0.869</td>
<td>0.841</td>
</tr>
<tr>
<td>Dummy for Multiple Enterprises</td>
<td>0.369</td>
<td>0.019</td>
</tr>
<tr>
<td>Liquidity Reserve*</td>
<td>0.913</td>
<td>1.837</td>
</tr>
<tr>
<td>Dividends*</td>
<td>0.082</td>
<td>0.133</td>
</tr>
<tr>
<td>Market Share (%)</td>
<td>0.926</td>
<td>0.067</td>
</tr>
<tr>
<td>Size (Log of Employed)</td>
<td>5.038</td>
<td>4.007</td>
</tr>
<tr>
<td>Ratio of Intangible to Total Assets (%)</td>
<td>4.284</td>
<td>2.771</td>
</tr>
<tr>
<td>Dummy for Negative Cashflow</td>
<td>0.069</td>
<td>0.110</td>
</tr>
<tr>
<td>No. of Observations</td>
<td>2.947</td>
<td>2.416</td>
</tr>
</tbody>
</table>

* Variables normalized by total capital assets
<table>
<thead>
<tr>
<th>Variables of interest</th>
<th>Coefficient Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specification 1</td>
</tr>
<tr>
<td></td>
<td>Eq. (14)</td>
</tr>
<tr>
<td>Share of Innovative Sales</td>
<td>0.201***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
</tr>
<tr>
<td>Long Term Debt</td>
<td>0.781***</td>
</tr>
<tr>
<td></td>
<td>(0.247)</td>
</tr>
<tr>
<td>Cashflows</td>
<td>0.313***</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
</tr>
<tr>
<td>Dummy for Negative Cashflows</td>
<td>0.99***</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
</tr>
<tr>
<td>Liquidity Reserve</td>
<td>-0.26***</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
</tr>
<tr>
<td>Dividends</td>
<td>-3.624***</td>
</tr>
<tr>
<td></td>
<td>(0.454)</td>
</tr>
<tr>
<td>Size</td>
<td>-0.49***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
</tr>
<tr>
<td>Market Share</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.011***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>Ratio of Intangible to Assets</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
</tr>
<tr>
<td>Dummy for Multiple</td>
<td>0.682</td>
</tr>
<tr>
<td>Enterprise Firms</td>
<td>(0.162)</td>
</tr>
<tr>
<td>Control Functions† for</td>
<td></td>
</tr>
<tr>
<td>Share of Innovative Sales</td>
<td>-1.328***</td>
</tr>
<tr>
<td></td>
<td>(0.184)</td>
</tr>
<tr>
<td>Sales</td>
<td>-6.209***</td>
</tr>
<tr>
<td></td>
<td>(2.198)</td>
</tr>
<tr>
<td></td>
<td>(2.058)</td>
</tr>
<tr>
<td></td>
<td>(1.089)</td>
</tr>
<tr>
<td>Ratio of Intangible to Assets</td>
<td>-1.290***</td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
</tr>
<tr>
<td>Size</td>
<td>-0.871***</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
</tr>
<tr>
<td>Individual Effects</td>
<td>-0.729***</td>
</tr>
<tr>
<td></td>
<td>(0.187)</td>
</tr>
<tr>
<td>(\rho_\iota)</td>
<td>0.589***</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
</tr>
</tbody>
</table>

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 \(\tilde{\Sigma}_\iota = \{\rho_{\zeta_1 \sigma_1}, \ldots, \rho_{\zeta_m \sigma_m}\}\) of equation (14) and \(\tilde{\Sigma}_\nu = \{\rho_{\nu_1 \sigma_1}, \ldots, \rho_{\nu_m \sigma_m}\}\) of equation (15).
Table 5: Average Partial Effects of Second Stage Estimates

<table>
<thead>
<tr>
<th></th>
<th>Specification 1</th>
<th>Specification 2</th>
<th>Specification 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eq. (14)</td>
<td>Eq. (15)</td>
<td>Eq. (14)</td>
</tr>
<tr>
<td>Financial Constraints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of Innovative Sales</td>
<td>0.028***</td>
<td>0.028***</td>
<td>0.028***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Long Term debt</td>
<td>0.107***</td>
<td>-0.091***</td>
<td>0.108***</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.025)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Cashflows</td>
<td>0.043***</td>
<td>0.043***</td>
<td>0.043***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Dummy for Negative Cashflows</td>
<td>0.163***</td>
<td>0.166***</td>
<td>0.166***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Liquidity Reserve</td>
<td>-0.036***</td>
<td>0.127***</td>
<td>-0.041***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.023)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Dividends</td>
<td>-0.497***</td>
<td>0.005</td>
<td>-0.502***</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Size</td>
<td>-0.067***</td>
<td>0.072***</td>
<td>-0.066***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Market Share</td>
<td>0.001</td>
<td>0.032***</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.005)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.001***</td>
<td>-0.003***</td>
<td>-0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Ratio of Intangible Assets</td>
<td>0.006</td>
<td>-0.064***</td>
<td>0.008***</td>
</tr>
<tr>
<td>to Total Assets</td>
<td>(0.004)</td>
<td>(0.008)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Dummy for Multiple</td>
<td>0.011</td>
<td>0.555***</td>
<td>0.621***</td>
</tr>
<tr>
<td>Enterprise Firms</td>
<td>(0.023)</td>
<td>(0.097)</td>
<td>(0.013)</td>
</tr>
</tbody>
</table>

Significance levels: *: 10%  **: 5%  ***: 1%
Figure 1: Plot of APE of Long-term Debt on the Probability of Innovation conditional on being Financially Constrained, $\frac{\partial \text{Pr}(s_{it} = 1|f_{it} = 1)}{\partial \text{DEBT}_{it}}$, or not Financially Constrained, $\frac{\partial \text{Pr}(s_{it} = 1|f_{it} = 0)}{\partial \text{DEBT}_{it}}$, against Age, Size, and Leverage.

(a) Age

(b) Log of Employed

(c) Long-Term Debt

$\frac{\partial \text{Pr}(s_{it} = 1|f_{it} = 1)}{\partial \text{DEBT}_{it}}$, $\frac{\partial \text{Pr}(s_{it} = 1|f_{it} = 0)}{\partial \text{DEBT}_{it}}$
Table 6: Third Stage Estimates: R&D Switching Regression Model

<table>
<thead>
<tr>
<th>Variables of Interest</th>
<th>Specification 1</th>
<th>Specification 2</th>
<th>Control Functions</th>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Control Function for Size</td>
<td>No Control Function for Size</td>
<td>For Financially Constrained Firms</td>
<td>No Control Function for Size</td>
<td>No Control Function for Size</td>
</tr>
<tr>
<td>$f$, Binary variable for Financial Constraint</td>
<td>-1.049</td>
<td>-0.84**</td>
<td>Share of Innovative Sales</td>
<td>-1.559***</td>
<td>-1.597***</td>
</tr>
<tr>
<td>$f$ * Share of Innovative Sales</td>
<td>0.217***</td>
<td>0.219***</td>
<td>Sales</td>
<td>(0.159)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>$(1 - f)$ * Share of Innovative Sales</td>
<td>0.201***</td>
<td>0.205***</td>
<td>Long-term Debt</td>
<td>0.525**</td>
<td>0.511**</td>
</tr>
<tr>
<td>$f$ * Cashflow</td>
<td>0.07</td>
<td>0.071*</td>
<td>Dividends</td>
<td>-1.296***</td>
<td>-1.232***</td>
</tr>
<tr>
<td>$(1 - f)$ * Cashflow</td>
<td>0.005</td>
<td>0.003</td>
<td>Liquidity Reserve</td>
<td>-0.395</td>
<td>-0.332</td>
</tr>
<tr>
<td>$f$ * Dummy for Multiple Enterprise</td>
<td>0.799***</td>
<td>0.682***</td>
<td>Ratio of Intangible to Total Assets</td>
<td>-0.034</td>
<td>-0.036</td>
</tr>
<tr>
<td>$(1 - f)$ * Dummy for Multiple Enterprise</td>
<td>0.514***</td>
<td>0.429***</td>
<td>Size</td>
<td>0.067</td>
<td>(0.106)</td>
</tr>
<tr>
<td>$f$ * Market Share</td>
<td>0.027*</td>
<td>0.019**</td>
<td>Financial Constraint</td>
<td>0.967***</td>
<td>0.83***</td>
</tr>
<tr>
<td>$(1 - f)$ * Market Share</td>
<td>0.011</td>
<td>0.005</td>
<td>Selection</td>
<td>0.636*</td>
<td>0.589*</td>
</tr>
<tr>
<td>$f$ * Size</td>
<td>-0.494***</td>
<td>-0.431***</td>
<td>Individual effects</td>
<td>-0.413*</td>
<td>-0.297***</td>
</tr>
<tr>
<td>$(1 - f)$ * Size</td>
<td>-0.364***</td>
<td>-0.318***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f$ * Age</td>
<td>-0.012***</td>
<td>-0.012***</td>
<td>Share of Innovative Sales</td>
<td>-1.52***</td>
<td>-1.57***</td>
</tr>
<tr>
<td>$(1 - f)$ * Age</td>
<td>-0.002</td>
<td>-0.003**</td>
<td>Long-term Debt</td>
<td>-0.029</td>
<td>-0.034</td>
</tr>
<tr>
<td>Dividends</td>
<td>0.022</td>
<td>0.027</td>
<td>Sales</td>
<td>(0.040)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Liquidity Reserve</td>
<td>0.018***</td>
<td>0.189***</td>
<td></td>
<td>(0.063)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Ratio of Intangible to Total Assets</td>
<td>-0.089***</td>
<td>-0.092***</td>
<td></td>
<td>(0.013)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Size</td>
<td>0.034</td>
<td>(0.074)</td>
<td>Financial Constraint</td>
<td>-0.277</td>
<td>-0.186**</td>
</tr>
<tr>
<td>$(C_{01}(t))$</td>
<td>(0.198)</td>
<td>(0.065)</td>
<td>Selection</td>
<td>-0.883***</td>
<td>-0.745***</td>
</tr>
<tr>
<td>$(C_{02}(t))$</td>
<td>(0.324)</td>
<td>(0.114)</td>
<td>Individual effects</td>
<td>0.346***</td>
<td>0.312***</td>
</tr>
</tbody>
</table>

Average Partial Effect | -0.241 | -0.175 | Long-term Debt | -0.029 | -0.034 |
| of Financial Constraint | (0.7) | (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 $\tilde{\Sigma}_{\eta_{1}} = \{\rho_{n_{1}e_{1}}\sigma_{n_{1}} \ldots , \rho_{n_{1}e_{m}}\sigma_{n_{1}}\}$ for the firms that are financially constrained and $\tilde{\Sigma}_{\eta_{0}} = \{\rho_{n_{0}e_{1}}\sigma_{n_{0}} \ldots , \rho_{n_{0}e_{m}}\sigma_{n_{0}}\}$ for the firms that are not financially constrained of the R&D equation (13).
Supplement to “Microeconometric 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 \( N' \) be the total number of observations, i.e., \( N' = \sum_{p=1}^{P} N_pp \). 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 \( I(p) = (M_{p-1}+1, \ldots, M_p)(p = 1, \ldots, P; M_0 = 0) \). We may, formally, consider \( I_1 \) as a cross section and \( I_p(p = 2, \ldots, P) \) as a balanced panel with \( p \) observations of each individual.

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

\[
\begin{align*}
x_{it} &= Z_{it}'\delta + \tilde{Z}_{it}'\kappa + \alpha_i + \epsilon_{it} = Z_{it}'\delta + \tilde{Z}_{it}'\kappa + u_{it},
\end{align*}
\]

where \( x_{it} = (x_{1it}, \ldots, x_{mit})' \) and \( Z_{it} = \text{diag}(z_{1it}, \ldots, z_{mit}) \) is the matrix of exogenous variables appearing in each of the \( m \) reduced form equation in (C-1), \( \delta = (\delta_1', \ldots, \delta_m')', \kappa = (\kappa_1, \ldots, \kappa_m)' \), and \( \epsilon_{it} = (\epsilon_{1it}, \ldots, \epsilon_{mit})' \). \( \sigma^2_{\alpha} \) is the variance of \( \alpha_i \), 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, \( \delta, \Sigma_{\epsilon}, \kappa, \) and \( \sigma^2_{\alpha} \). Given the distribution of \( \alpha_i \), \( \kappa\alpha_i \) is normally distributed with mean zero and variance \( \Sigma_{\alpha} \), given by:

\[
\Sigma_{\alpha} = \sigma^2_{\alpha}\Sigma_{\kappa} = \sigma^2_{\alpha} \begin{pmatrix}
\kappa_1^2 & \kappa_1\kappa_2 & \kappa_2^2 \\
\vdots & \vdots & \vdots \\
\kappa_1\kappa_m & \kappa_2\kappa_m & \cdots & \kappa_m^2
\end{pmatrix}.
\]

∗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
\( \epsilon_{it} \) is normally distributed with mean zero and variance \( \Sigma_{\epsilon\epsilon} \). We assume that \( \alpha_i \) and \( \epsilon_{it} \) are mutually uncorrelated and given that \( Z'_i t \) is exogenous, \( \alpha_i \) and \( \epsilon_{it} \) are uncorrelated with \( Z'_i t \).

Let \( x_{i(p)} = (x'_{i1}, \ldots, x'_{ip})' \), \( Z_i(p) = (Z'_{i1}, \ldots, Z'_{ip})' \) and \( e_{i(p)} = (e'_{i1}, \ldots, e'_{ip})' \) and write the model as

\[
x_{i(p)} = Z'_{i(p)} \delta + (e_p \otimes Z'_i \delta \kappa) + (e_p \otimes \alpha_i \kappa) + e_{i(p)} = Z'_{i(p)} \delta + (e_p \otimes Z'_i \delta \kappa) + u_{i(p)},
\]

(C-2)

\[
E(u_{i(p)} u'_{i(p)}) = I_p \otimes \Sigma_{\epsilon\epsilon} + E_p \otimes \Sigma_{\alpha} = K_p \otimes \Sigma_{\epsilon\epsilon} + J_p \otimes \Sigma_{(p)} = \Omega_u(p)
\]

(C-3)

where

\[
\Sigma_{(p)} = \Sigma_{\epsilon\epsilon} + p \Sigma_{\alpha}, \ p = 1, \ldots, \ P
\]

(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 \( \hat{\delta} = \{\hat{\delta}', \hat{\delta}'\}' \) when \( \kappa, \sigma_{\alpha} \) (hence \( \Sigma_{\alpha} \)), and \( \Sigma_{\epsilon\epsilon} \) are known. Define \( Q_{i(p)} = u'_{i(p)} \Omega^{-1}_{u(p)} u_{i(p)} \), then GMM estimation is the problem of minimizing \( Q = \sum_{p=1}^{P} \sum_{i \in I(p)} Q_{i(p)} \) with respect to \( \hat{\delta} \).

Since \( \Omega^{-1}_{u(p)} = K_p \otimes \Sigma_{\epsilon\epsilon}^{-1} + J_p \otimes (\Sigma_{\epsilon\epsilon} + p \Sigma_{\alpha})^{-1} \), we can rewrite \( Q \) as

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

(C-5)

with \( u_{i(p)} = x_{i(p)} - Z'_{i(p)} \delta - (e_p \otimes Z'_i \delta \kappa) \). Had we not imposed the restriction that \( \hat{\delta} \) be the same for each of the \( m \) equations we could have estimated \( \delta \) and \( \hat{\delta} \) by employing GLS estimation as in Björn (2004).

### C.2 Maximum Likelihood Estimation

We now consider ML estimation of \( \Theta_1 = \{\delta, \Sigma_{\epsilon\epsilon}, \kappa, \sigma_{\alpha}^2\} \). Assuming normality of \( \alpha_i \) and the disturbances \( \epsilon_{it} \), i.e., \( \alpha_i \kappa \sim \text{INN}(0, \sigma_{\alpha}^2 \Sigma_\alpha) \) and \( \epsilon_{it} \sim \text{INN}(0, \Sigma_{\epsilon\epsilon}) \), then \( u_{i(p)} = (e_p \otimes \alpha_i \kappa) + e_{i(p)} \sim \text{INN}(0_{mp}, I_{mp}) \). The log-likelihood function of all \( x \)’s conditional on all \( Z \)’s for an individual in group \( p \) and for all individuals then become, respectively,

\[
L_{i(p)}(\Theta_1) = \frac{-mp}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega_u(p)| - \frac{1}{2} Q_{i(p)}(\hat{\delta}, \Sigma_{\epsilon\epsilon}, \kappa, \sigma_{\alpha}^2)
\]

(C-6)

\[
L_1(\Theta_1) = \sum_{p=1}^{P} \sum_{i \in I(p)} L_{i(p)} = \frac{-mN}{2} \ln(2\pi) - \frac{1}{2} \sum_{p=1}^{P} N_{p} \ln |\Omega_u(p)| - \frac{1}{2} \sum_{p=1}^{P} \sum_{i \in I(p)} Q_{i(p)}(\hat{\delta}, \Sigma_{\epsilon\epsilon}, \kappa, \sigma_{\alpha}^2),
\]

(C-7)
where $|\Omega_u(p)| = |\Sigma_{ie}|^{p-1}$.

We split the problem into: (A) Maximization of $L$ with respect to $\hat{\delta}$ for given $(\Sigma_{ie}, \kappa, \sigma^2_\alpha)$ and (B) Maximization of $L_1(\Theta_1)$ with respect to $(\Sigma_{ie}, \kappa, \sigma^2_\alpha)$ for given $\hat{\delta}$. Subproblem (A) is identical with the GMM problem, since maximization of $L_1(\Theta_1)$ with respect to $\hat{\delta}$ for given $(\Sigma_{ie}, \kappa, \sigma^2_\alpha)$ is equivalent to minimization of $\sum_{p=1}^P \sum_{i \in I(p)} Q_i(p)(\hat{\delta}, \Sigma_{ie}, \kappa, \sigma^2_\alpha)$.

The first order conditions with respect to $\Sigma_{ie}, \kappa,$ and $\sigma^2_\alpha$, which we derive in Appendix E does not have a closed form solution. To obtain estimates of $\Sigma_{ie}, \kappa,$ and $\sigma^2_\alpha$, we numerically maximize $L_1(\Theta_1)$ with respect to $\Sigma_{ie}, \kappa,$ and $\sigma^2_\alpha$ for a given $\hat{\delta}$ 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 $\hat{\delta}$ and (C-7) with respect to $\Sigma_{ie}, \kappa,$ and $\sigma^2_\alpha$ and iterating until convergence. Biørn (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 $\eta_1, \eta_0, \zeta$ and $v$. We know that the conditional expectation of $\eta$, where $\eta$ is either $\eta_1$ or $\eta_0$, given $\zeta$ and $v$, $E[\eta|\zeta, v]$, is given by

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

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

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

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

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

which can be written as

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

(D-1)
Hence,

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

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

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

where, \( \phi_2 \) and \( \Phi_2 \) denote respectively the density and cumulative density function of a standard bivariate normal. Now, consider the expression \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\bar{\zeta} - \rho_{\zeta \upsilon} \bar{\upsilon}) \phi_2(\bar{\zeta}, \bar{\upsilon}, \rho_{\zeta \upsilon}) \, d\bar{\zeta} \, d\bar{\upsilon} \), of the RHS in (D-2). Given that \( \phi_2(\bar{\zeta}, \bar{\upsilon}, \rho_{\zeta \upsilon}) = \phi(\bar{\zeta}) \frac{1}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}} \phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right) \), the concerned expression can be written as

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\bar{\zeta} - \rho_{\zeta \upsilon} \bar{\upsilon}) \phi(\bar{\zeta}) \frac{1}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}} \phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right) \, d\bar{\zeta} \, d\bar{\upsilon} = \int_{-\infty}^{\infty} \zeta \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right)\right) \, d\bar{\zeta} - \rho_{\zeta \upsilon} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\upsilon} \phi(\bar{\zeta}) \frac{1}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}} \phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right) \, d\bar{\zeta} \, d\bar{\upsilon}.
\]

(D-3)

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

\[
\int_{-\infty}^{\infty} \zeta \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right)\right) \, d\bar{\zeta} - \rho_{\zeta \upsilon} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\upsilon} \phi(\bar{\zeta}) \frac{1}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}} \phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right) \, d\bar{\zeta} \, d\bar{\upsilon} = \int_{-\infty}^{\infty} \zeta \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right)\right) \, d\bar{\zeta} - \rho_{\zeta \upsilon} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\upsilon} \phi(\bar{\zeta}) \frac{1}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}} \phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right) \, d\bar{\zeta} \, d\bar{\upsilon}.
\]

(D-4)

\[
= \left(1 - \rho^2_{\zeta \upsilon}\right) \int_{-\infty}^{\infty} \zeta \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right)\right) \, d\bar{\zeta} - \rho_{\zeta \upsilon} \sqrt{(1 - \rho^2_{\zeta \upsilon})} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\upsilon} \phi(\bar{\zeta}) \phi(y) \, d\bar{\zeta} \, dy
\]

\[
= \left(1 - \rho^2_{\zeta \upsilon}\right) \int_{-\infty}^{\infty} \zeta \phi(\bar{\zeta}) \left(1 - \Phi\left(\frac{\bar{\upsilon} - \rho_{\zeta \upsilon} \bar{\zeta}}{\sqrt{(1 - \rho^2_{\zeta \upsilon})}}\right)\right) \, d\bar{\zeta} - \rho_{\zeta \upsilon} \sqrt{(1 - \rho^2_{\zeta \upsilon})} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\upsilon} \phi(\bar{\zeta}) \phi(y) \, d\bar{\zeta} \, dy
\]

(D-4)
\[\begin{align*}
&= (1 - \rho_{\zeta \nu}^2) \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \hat{\zeta} \phi(\hat{\zeta}) \Phi\left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right) d\hat{\zeta} - \rho_{\zeta \nu} \sqrt{1 - \rho_{\zeta \nu}^2} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} y \phi(\hat{\zeta}) \phi(y) d\hat{\zeta} dy. \\
&\quad \text{(D-5)}
\end{align*}\]

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

\[\begin{align*}
&= -(1 - \rho_{\zeta \nu}^2) \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \phi(\hat{\zeta}) \Phi\left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right) d\hat{\zeta} + \rho_{\zeta \nu} \sqrt{1 - \rho_{\zeta \nu}^2} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \phi(\hat{\zeta}) \Phi\left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right) d\hat{\zeta} \\
&\quad \text{(D-6)}
\end{align*}\]

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

\[\begin{align*}
&-\rho_{\zeta \nu} \sqrt{1 - \rho_{\zeta \nu}^2} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} y \phi(\hat{\zeta}) \phi(y) d\hat{\zeta} dy = \rho_{\zeta \nu} \sqrt{1 - \rho_{\zeta \nu}^2} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \phi(\hat{\zeta}) \phi(y) d\hat{\zeta} dy \\
&= \rho_{\zeta \nu} \sqrt{1 - \rho_{\zeta \nu}^2} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \phi(y) \left|\int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \phi(\hat{\zeta}) d\hat{\zeta}\right| = \rho_{\zeta \nu} \sqrt{1 - \rho_{\zeta \nu}^2} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \phi(y) \left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right) d\hat{\zeta}. \\
&\quad \text{(D-7)}
\end{align*}\]

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

\[\begin{align*}
\int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} (\hat{\zeta} - \rho_{\zeta \nu} \hat{\nu}) \phi_2(\hat{\zeta}, \hat{\nu}, \rho_{\zeta \nu}) d\hat{\zeta} d\hat{\nu} &= (1 - \rho_{\zeta \nu}^2) \phi\left(\frac{\frac{a}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right) \Phi\left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right).
\end{align*}\]

Similarly, it can be shown that

\[\begin{align*}
\int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} \int_{\frac{b}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}^{\infty} (\hat{\nu} - \rho_{\zeta \nu} \hat{\zeta}) \phi_2(\hat{\zeta}, \hat{\nu}, \rho_{\zeta \nu}) d\hat{\zeta} d\hat{\nu} &= (1 - \rho_{\zeta \nu}^2) \phi\left(\frac{\frac{a}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right) \Phi\left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right).
\end{align*}\]

Hence,

\[\begin{align*}
E[\eta | \hat{\zeta}] &= \frac{a}{\sigma_{\zeta}} - \frac{b}{\sigma_{\zeta}} = \frac{\sigma_{\eta} \rho_{\eta \zeta} \phi\left(\frac{\frac{a}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right)}{\Phi_2\left(\frac{\frac{a}{\sigma_{\zeta}} - \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right)} \Phi\left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right) + \frac{\sigma_{\eta} \rho_{\eta \zeta} \phi\left(\frac{\frac{b}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right)}{\Phi_2\left(\frac{\frac{a}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right)} \Phi\left(\frac{\frac{b}{\sigma_{\zeta}} + \rho_{\zeta \nu} \frac{\zeta}{\sigma_{\zeta}}}{\sqrt{1 - \rho_{\zeta \nu}^2}}\right).
\end{align*}\]

\[\text{(D-8)}\]
Now, consider

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

$$= \frac{1}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_\upsilon}, -\rho_{\zeta \upsilon}\right)} \Phi\left(\frac{b}{\sigma_\upsilon} - \rho_{\zeta \upsilon} \frac{\bar{\upsilon}}{\sigma_\upsilon}\right) + \frac{1}{\Phi_2\left(\frac{-a}{\sigma_\zeta}, \frac{b}{\sigma_\upsilon}, -\rho_{\zeta \upsilon}\right)} \Phi\left(\frac{\bar{\upsilon}}{\sigma_\upsilon} - \rho_{\zeta \upsilon} \frac{\bar{\zeta}}{\sigma_\zeta}\right).$$

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

$$E[\eta|\zeta \leq -a, \upsilon > -b] = \Phi\left(\frac{b}{\sigma_\upsilon} + \rho_{\zeta \upsilon} \frac{\bar{\upsilon}}{\sigma_\upsilon}\right) + \Phi\left(\frac{\bar{\upsilon}}{\sigma_\upsilon} + \rho_{\zeta \upsilon} \frac{\bar{\zeta}}{\sigma_\zeta}\right).$$

**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,

$$f^*_{it} = X^\prime_{it} \phi + (\bar{Z}^\prime \delta + \bar{\alpha}_i) \lambda + \tilde{\Sigma}_{\zeta \zeta}^{-1} \tilde{\epsilon}_{it} + \tilde{\zeta}_{it},$$

$$s^*_{it} = X^\prime_{it} \gamma + (\bar{Z}^\prime \theta + \bar{\alpha}_i) \theta + \tilde{\Sigma}_{\upsilon \upsilon}^{-1} \tilde{\epsilon}_{it} + \tilde{\upsilon}_{it},$$

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 \( i \)

\[
\Pr(s_{it} = 1| f_{it} = 1, X, Z, \hat{\alpha}, \hat{\epsilon}_t) = \frac{\Pr(s_{it} = 1, f_{it} = 1, X, Z, \hat{\alpha}, \hat{\epsilon}_t)}{\Pr(f_{it} = 1, X, Z, \hat{\alpha})} = \Phi_2(\varphi_t, \gamma_t, \rho_{\xi\xi_t}),
\]

\[
\Pr(s_{it} = 1| f_{it} = 0, X, Z, \hat{\alpha}, \hat{\epsilon}_t) = \frac{\Pr(s_{it} = 1, f_{it} = 0, X, Z, \hat{\alpha}, \hat{\epsilon}_t)}{\Pr(f_{it} = 0, X, Z, \hat{\alpha})} = \frac{\Phi_2(\varphi_t, -\gamma_t, \rho_{\xi\xi_t})}{1 - \Phi(\varphi_t)},
\]

where \( \Phi_2 \) is the cumulative distribution function of a standard bivariate normal and

\[
\varphi_{it} = \left( X_{it}^\prime \varphi + \lambda(\delta_i' + \hat{\alpha}_i) + \Sigma_{ic} \hat{\Sigma}_{ic}^{-1} \hat{\epsilon}_{it} \right) \frac{1}{\sigma_{\xi}}, \quad \text{and} \quad \gamma_{it} = \left( X_{it}^\prime \gamma + \theta(\delta_i' + \hat{\alpha}_i) + \Sigma_{ic} \hat{\Sigma}_{ic}^{-1} \hat{\epsilon}_{it} \right) \frac{1}{\sigma_{\xi}}.
\]

Hence, for an individual \( i \) we have

\[
\frac{\partial \Pr(s_{it} = 1| f_{it} = 1)}{\partial w} = \int \frac{\partial}{\partial w} \left( \frac{\Phi_2(\varphi_t, \gamma_t, \rho_{\xi\xi_t})}{\Phi(\varphi_t)} \right) dG(\hat{\alpha}, \hat{\epsilon}).
\]

(E-1)

If \( w \) belongs to both the specifications, \( \varphi_t \) and \( \gamma_t \), then the above involves taking derivative of CDF of a standard bivariate normal with respect to \( \varphi_t \) and \( \gamma_t \). It can be shown that in (E-1)

\[
\frac{\partial}{\partial w} \left( \frac{\Phi_2(\varphi_t, \gamma_t, \rho_{\xi\xi_t})}{\Phi(\varphi_t)} \right) = \frac{1}{\Phi(\varphi_t)} \left[ g_s \gamma_w + \left( g_f - \phi_2(\varphi_t) \right) \frac{\phi(\varphi_t)}{\Phi(\varphi_t)} \varphi_w \right],
\]

(E-2)

where

\[
g_f = \phi(\varphi_t) \Phi \left( \frac{\gamma_t - \rho_{\xi\xi_t} \varphi_t}{\sqrt{1 - \rho_{\xi\xi_t}^2}} \right) \quad \text{and} \quad g_s = \phi(\gamma_t) \Phi \left( \frac{\varphi_t - \rho_{\xi\xi_t} \gamma_t}{\sqrt{1 - \rho_{\xi\xi_t}^2}} \right).
\]

(E-3)

The derivatives of the other conditional probabilities with respect to \( \varphi_t \) and \( \gamma_t \) can be found in Greene (2002). Once the integrand in (E-2) is estimated at \( X_{it}^\prime = \hat{X}_i^\prime \) and \( X_{it}^\prime = \hat{X}_i^\prime \), given the estimates \( \hat{\alpha}_i \) and \( \hat{\epsilon}_{it} \), 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

\[
\frac{\partial \Pr(f_{it} = 1)}{\partial w} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \varphi_w(\hat{X}_i t' \varphi),
\]

where \( \hat{X}_i t' = \{ \hat{X}_i t', \hat{\alpha}_i, (\hat{\Sigma}_{ic} \hat{\Sigma}_{ic}^{-1} \hat{\epsilon}_{it}) \} \) and \( \varphi = \{ \varphi', \hat{\lambda}, \hat{\Sigma}_{ic} \} \). Since each of the \( \varphi_w(\hat{X}_i t' \varphi) \) is a function of \( \varphi \) the variance of \( \frac{\partial \Pr(f_{it} = 1)}{\partial w} \) will be a function of the variance of the estimate of \( \varphi \). Now, we know that by the linear approximation approach (delta method), the asymptotic covariance matrix of \( \frac{\partial \Pr(f_{it} = 1)}{\partial w} \) is given by

\[
\text{Asy. Var} \left[ \frac{\partial \Pr(f_{it} = 1)}{\partial w} \right] = \left[ \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial \varphi_w(\hat{X}_i t' \varphi)}{\partial \varphi} \right] V_{\varphi} \left[ \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial \varphi_w(\hat{X}_i t' \varphi)}{\partial \varphi} \right] \varphi, \quad (E-4)
\]
where \( V_{2f}^* \) is the second stage error adjusted covariance matrix, shown in Appendix E, of \( \hat{\varphi} \).

The RHS of (E-4) turns out to be

\[
\left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \phi(\tilde{X}_{it}' \tilde{\varphi}) [e_w - (\varphi\tilde{X}_{it}') \tilde{\varphi}_w \tilde{X}_{it}'] \right] V_{2f}^* \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \phi(\tilde{X}_{it}' \tilde{\varphi}) [e_w - (\varphi\tilde{X}_{it}') \tilde{\varphi}_w \tilde{X}_{it}'] \right]',
\]

(E-5)

where and \( e_w \) is a row vector having the dimension of \( \varphi' \) and with 1 at the position of \( \varphi_w \) in \( \varphi \) and zeros elsewhere. The estimated asymptotic covariance matrix of the APE of all the continuous variables in \( \tilde{X}' \) 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

\[
\Delta_w \Pr(f_{it} = 1) = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Phi_{it}(\cdot) \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Phi_{it}(\cdot) \right]',
\]

(E-6)

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

\[
\text{Asy. Var}_w \Pr(f_{it} = 1) = \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Phi_{it}(\cdot) \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Phi_{it}(\cdot) \right] \right] V_{2f}^* \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Phi_{it}(\cdot) \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Phi_{it}(\cdot) \right] \right]',
\]

where

\[
\frac{\partial \Delta \Phi_{it}(\cdot)}{\partial \varphi} = \frac{\partial \Phi_{it}(\cdot, w = 1)}{\partial \varphi} - \frac{\partial \Phi_{it}(\cdot, w = 0)}{\partial \varphi} = \phi_{it}(\cdot, w = 1) \left[ \tilde{X}_{it-1}' \right] - \phi_{it}(\cdot, w = 0) \left[ \tilde{X}_{it-1}' \right].
\]

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 \( \tilde{X}_{2it} = \{\tilde{X}_{it}', \tilde{X}_{it}^s\}' \) and \( \Theta_2 = \{\varphi', \gamma', \rho_{\gamma \varphi} \}' \), where \( \tilde{X}_{it}^s = \{\tilde{X}_{it}', \tilde{\alpha}_i, (\tilde{\Sigma}^{-1} \tilde{\varphi} \tilde{\varphi}^T \tilde{\alpha}_i, (\tilde{\Sigma}^{-1} \tilde{\varphi} \tilde{\varphi}^T \tilde{\alpha}_i) \}' \) and \( \gamma = \{\gamma', \theta, \tilde{\Sigma}^T \}' \) and denote the right hand side of (E-2) as \( \Lambda(s=1|f=1), \varphi(\tilde{X}_{2it}, \Theta_2) \). Then the APE of \( w \) on the conditional probability of being an innovator given that the firm is financially constrained is given by

\[
\frac{\partial \text{Pr}(s_{it} = 1|f_{it} = 1)}{\partial w} = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \Lambda(s=1|f=1), \varphi(\tilde{X}_{2it}, \Theta_2)
\]

By the delta method we know that the asymptotic variance of \( \frac{\partial \text{Pr}(s_{it} = 1|f_{it} = 1)}{\partial w} \) is given by

\[
\left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Lambda(s=1|f=1), \varphi(\tilde{X}_{2it}, \Theta_2) \right] V_{2f}^* \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Lambda(s=1|f=1), \varphi(\tilde{X}_{2it}, \Theta_2) \right] ^{T}, \tag{E-7}
\]
where \( V_2^* \) is second stage error corrected covariance matrix of \( \hat{\Theta}_2 \). The derivative of \( \Lambda (s=1|f=1), w(\hat{X}_{2it}, \hat{\Theta}_2) \) with respect to the second stage parameters, \( \Theta_2 \), can easily 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 \( X_{it} = \bar{X} \), where \( X_{it} \) is the union of elements appearing in \( X_{it}^f, X_{it}^s \), and \( X_{it}^s \), 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

\[
\Delta_f E(r_{it}|\bar{X}) = \int E(r_{1it}|\bar{X}, f_{it} = 1, s_{it} = 1, \hat{\alpha}, \hat{\epsilon})dG(\hat{\alpha}, \hat{\epsilon})
\]

\[ - \int E(r_{0it}|\bar{X}, f_{it} = 0, s_{it} = 1, \hat{\alpha}, \hat{\epsilon})dG(\hat{\alpha}, \hat{\epsilon}). \tag{E-8} \]

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

\[
E(r_{1it}|\bar{X}, f_t = 1, s_t = 1, \hat{\alpha}, \hat{\epsilon}_t) = \beta_f + \tilde{X}^r \beta_1 + \mu_1(\tilde{Z}\delta + \hat{\alpha}) + \tilde{\Sigma}_0 \tilde{\Sigma}^{-1} \hat{\epsilon}_t + \sigma_{\hat{\eta}_{0i}} C_{11}(\hat{\alpha}, \hat{\epsilon}_t) + \sigma_{\hat{\eta}_{0i}} C_{12}(\hat{\alpha}, \hat{\epsilon}_t) \tag{E-9} \]

if \( f_{it}^* > 0 \), and

\[
E(r_{0it}|\bar{X}, f_t = 0, s_t = 1, \hat{\alpha}, \hat{\epsilon}_t) = \tilde{X}_t^r \beta_0 + \mu_0(\tilde{Z}\delta + \hat{\alpha}) + \tilde{\Sigma}_0 \tilde{\Sigma}^{-1} \hat{\epsilon}_t + \sigma_{\hat{\eta}_{0i}} C_{01}(\hat{\alpha}, \hat{\epsilon}_t) + \sigma_{\hat{\eta}_{0i}} C_{02}(\hat{\alpha}, \hat{\epsilon}_t) \tag{E-10} \]

if \( f_{it}^* \leq 0 \), and where the correction terms – \( C_{11}(\tilde{X}^f, \tilde{X}^s, \hat{\alpha}, \hat{\epsilon}_t), C_{12}(\tilde{X}^f, \tilde{X}^s, \hat{\alpha}, \hat{\epsilon}_t), C_{01}(\tilde{X}^f, \tilde{X}^s, \hat{\alpha}, \hat{\epsilon}_t), \) and \( C_{02}(\tilde{X}^f, \tilde{X}^s, \hat{\alpha}, \hat{\epsilon}_t) \) – are defined at the given \( X_{it}^f = \tilde{X}^f \) and \( X_{it}^s = \tilde{X}^s \). 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 seconds stage, the variance of the APE will be a function of the variances of the correction terms. Since the correction terms are in turn functions of the estimated coefficients in the second stage, the variance of the estimated APE be a function of the variance of the estimated second stage coefficients.

To see this, consider the the conditional APE of the financial constraint on the R&D...
expenditure, which is given by
\[
\Delta f \hat{E}(r_t|\bar{X}) = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \left[ s_{it} \left( \hat{\beta}_f + \hat{\mathcal{X}}''(\hat{\beta}_f - \hat{\beta}_0) + (\hat{\mu}_1 - \hat{\mu}_0)(\hat{\delta} + \hat{\alpha}) + (\hat{\Sigma} - \hat{\Sigma}_0) \hat{\epsilon}_t \right) \\
+ \sigma_{\tilde{\alpha}} \rho_{\tilde{\alpha}i'} C_{11}(\hat{\alpha}, \hat{\epsilon}_i) + \sigma_{\tilde{\alpha}} \rho_{\tilde{\alpha}i'} C_{12}(\hat{\alpha}, \hat{\epsilon}_i) - \sigma_{\tilde{\alpha}} \rho_{\tilde{\alpha}i'} C_{01}(\hat{\alpha}, \hat{\epsilon}_i) - \sigma_{\tilde{\alpha}} \rho_{\tilde{\alpha}i'} C_{02}(\hat{\alpha}, \hat{\epsilon}_i) \right]
\]  
(E-11)

Let us denote the structural coefficients of our model as \( \Theta_s = \{ \Theta'_2, \Theta'_3 \}' \) where \( \Theta'_2 \) and \( \Theta'_3 \) are the vector of structural coefficients estimated in the third stage respectively. Again, by the application of the delta method we know that

\[
\text{Asy. Var}[\Delta f \hat{E}(r_t|\bar{X})] = \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Delta f \hat{E}(r_t|\bar{X}) \right]' \left[ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} T_i \sum_{t=1}^{T_i} \partial \Delta f \hat{E}(r_t|\bar{X}) \right],
\]

(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 \( \Theta_2 \), the above involves taking the derivative of the correction terms with respect to the second stage parameters \( \Theta_2 \).

**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 \( \Theta_1, \Theta_2, \) and \( \Theta_3 \) 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

\[
E(\mathcal{L}_{i(p)3}\Theta_3) = E \left[ \sum_{t=1}^{P} s_{it} X_{it}' (r_{it} - X_{it}' \Theta_3) \right] = 0 \quad \text{(F-3)}
\]

where \( L_{i(p)}(\Theta_1) \) is the likelihood function for individual \( i \) belonging to the group \( p \), \( p \in \{1, \ldots, P\} \), for the first step system of reduced form equations. The notation \( p \) was introduced
in Appendix C. \( p \) is the number of time period an individual is observed in an unbalanced panel; the minimum being 1 and maximum \( P \). Hence \( \sum_{i=1}^{N} \sum_{t=1}^{T_i} = \sum_{p=1}^{P} \sum_{i \in I(p)} \sum_{t=1}^{T_i} \), 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 \( X_{it}^* \) on \( r_{it} \) for those firms, that have been selected, \( s_{it} = 1 \), where

\[
\begin{align*}
    r_{it} &= f_{it}r_{it} \\
    X_{it}^* &= f_{it}\{f_{it}, \alpha_{it}^*, \bar{Z}_t\delta + \hat{\alpha}_i(\Theta_1), (\Sigma_{\epsilon}^{-1}\epsilon_{it}(\Theta_1))^\prime, C_{11it}(\Theta_1, \Theta_2), C_{12it}(\Theta_1, \Theta_2)\}^\prime
\end{align*}
\]

if \( f_{it} = 1 \), else

\[
\begin{align*}
    r_{it} &= (1 - f_{it})r_{it} \\
    X_{it}^* &= (1 - f_{it})\{f_{it}, \alpha_{it}^*, \bar{Z}_t\delta + \hat{\alpha}_i(\Theta_1), (\Sigma_{\epsilon}^{-1}\epsilon_{it}(\Theta_1))^\prime, C_{01it}(\Theta_1, \Theta_2), C_{02it}(\Theta_1, \Theta_2)\}^\prime
\end{align*}
\]

if \( f_{it} = 0 \).

The estimates for \( \Theta_1 \), \( \Theta_2 \), and \( \Theta_3 \) 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

\[
\frac{1}{N}L_{1\Theta_1}(\hat{\Theta}_1) = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in I(p)} \frac{\partial \ln L_{i\Theta_1}(\hat{\Theta}_1)}{\partial \Theta_1} = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in I(p)} \sum_{t=1}^{T_i} \frac{\partial L_{i\theta_1}(\hat{\Theta}_1, \hat{\Theta}_2)}{\partial \Theta_1}
\]

where \( L_{i\Theta_1} = \ln L_{i\Theta_1}(\hat{\Theta}_1) \) is given by equation (C-6) in Appendix C. \( \Theta_1 = \{\bar{\delta}^\prime, \hat{\delta}^\prime \text{vech}(\Sigma_{\epsilon})^\prime, \kappa', \sigma_{\epsilon}^2\}^\prime \) 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

\[
\frac{1}{N}L_{2\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in I(p)} \frac{\partial L_{i\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2)}{\partial \Theta_2} = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in I(p)} \sum_{t=1}^{T_i} \frac{\partial L_{it2}(\hat{\Theta}_1, \hat{\Theta}_2)}{\partial \Theta_2}
\]

where \( L_{it2}(\Theta_1, \Theta_2) \) is given by equations (19) and (20) in the main text and \( \Theta_2 = \{\bar{\gamma}^\prime, \hat{\gamma}^\prime, \rho_{\epsilon\delta}\}^\prime \) was defined in Appendix E. Finally, the sample analog of the population for the third step estimation is given by

\[
\frac{1}{N}L_{3\Theta_3}(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3) = \frac{1}{N} \sum_{p=1}^{P} \sum_{i \in I(p)} \sum_{t=1}^{T_i} s_{it} X_{it}^*(r_{it} - X_{it}^*\hat{\Theta}_3)
\]

In Appendix A, we had shown that with \( \alpha_i \) substituted by their EAP values \( \hat{\alpha}_i(X_i, Z_i, \Theta_1) \) still leads to the identification of \( \Theta_2 \) and \( \Theta_3 \). Let \( \Theta_1^*, \Theta_2^* \), and \( \Theta_3^* \) respectively be the true values of \( \Theta_1, \Theta_2 \) and \( \Theta_3 \). Under the assumptions we make, maximizing \( L_{i\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) \) is asymptotically equivalent to maximizing \( L_{i\Theta_2}(\Theta_1^*, \Theta_2) \), where \( \hat{\Theta}_1 \) is a consistent first step estimate of \( \Theta_1 \). Hence \( \hat{\Theta}_2 \) obtained by solving \( \frac{1}{N}L_{2\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) = 0 \) is a consistent estimate of
By the same logic \( \hat{\Theta}_3 \) obtained by solving \( \frac{1}{N} L_{3\Theta_3}(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3) = 0 \) 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

\[
\frac{1}{N} \begin{bmatrix}
    L_{1\Theta_1}(\hat{\Theta}_1) \\
    L_{2\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) \\
    L_{3\Theta_3}(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3)
\end{bmatrix} = 0.
\]  

(F-7)

A series of Taylor’s expansion of \( L_{1\Theta_1}(\hat{\Theta}_1), L_{2\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) \) and \( L_{3\Theta_3}(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3) \) around \( \Theta^* \) gives

\[
\frac{1}{N} \begin{bmatrix}
    L_{1\Theta_1}(\hat{\Theta}_1) \\
    L_{2\Theta_2}(\hat{\Theta}_1, \hat{\Theta}_2) \\
    L_{3\Theta_3}(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3)
\end{bmatrix} \approx \begin{bmatrix}
    0 \\
    0 \\
    0
\end{bmatrix} + \begin{bmatrix}
    \sqrt{N}(\hat{\Theta}_1 - \Theta_1^*) \\
    \sqrt{N}(\hat{\Theta}_2 - \Theta_2^*) \\
    \sqrt{N}(\hat{\Theta}_3 - \Theta_3^*)
\end{bmatrix} = - \frac{1}{\sqrt{N}} \begin{bmatrix}
    L_{1\Theta_1} \\
    L_{2\Theta_2} \\
    L_{3\Theta_3}
\end{bmatrix}.
\]

In matrix notation the above can be written as

\[
B_{\Theta N} \sqrt{N}(\hat{\Theta} - \Theta) = - \frac{1}{\sqrt{N}} \Lambda_{\Theta N},
\]  

(F-8)

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

\[
B_s = \begin{bmatrix}
    L_{1\Theta_1} \\
    L_{2\Theta_1} \\
    L_{3\Theta_1}
\end{bmatrix},
\]

where a typical element, say, \( L_{2\Theta_1} \) is \( E(L_{i(\theta)\Theta_1}) \). \( \frac{\sqrt{N}}{\sqrt{N}} \Lambda_N \) in (F-8) converges in distribution to an asymptotically normal random variable with mean zero and a covariance matrix \( A_s = \lim E \frac{1}{N} \Lambda_N A_N' \), where \( A_s \) is given by

\[
A_s = \begin{bmatrix}
    V_{L_1L_1} & V_{L_1L_2} & V_{L_1L_3} \\
    V_{L_2L_1} & V_{L_2L_2} & V_{L_2L_3} \\
    V_{L_3L_1} & V_{L_3L_2} & V_{L_3L_3}
\end{bmatrix},
\]

where a typical element of \( A_s \), say \( V_{L_1L_2} \) is given by \( V_{L_1L_2} = E[L_{i(\theta)\Theta_1}(\Theta_1) L_{i(\theta)\Theta_2}(\Theta_1, \Theta_2)] \).

Under the regularity conditions \( \sqrt{N}(\hat{\Theta} - \Theta^*) \) is asymptotically normal with zero mean and covariance matrix\(^1\) given by \( B_s^{-1} A_s B_s^{-1}' \).

\[
\sqrt{N}(\hat{\Theta} - \Theta^*) \overset{d}{=} N[(0), (B_s^{-1} A_s B_s^{-1}')] \quad \text{(F-9)}
\]

\(^1\)The covariance matrices \( V_{ij}^* \) in equation (E-5), \( V_{i}^* \) in equation (E-7), and \( V_{i}^* \) in equation (E-12) can obtained by selecting the appropriate submatrix of \( \frac{1}{N} B_s^{-1} A_s B_s^{-1}' \).
The moment conditions for every individual, at the estimates of $\Theta_1$, $\Theta_2$, and $\Theta_3$, of the three stages can be employed to obtain the sample analog of every element in $A_s$. For example, to get an estimate of $V_{L_1L_2}$ we have to estimate $\frac{1}{N}\sum_{p=1}^{P} \sum_{i \in I_p} [L_{i(p)}1\theta_1(\hat{\Theta}_1)L_{i(p)2\theta_2}(\hat{\Theta}_1, \hat{\Theta}_2)]$.

Consider now the elements of $B_s$. Since in the first and the second stage we employ MLE, at $\Theta_1^*$ and $\Theta_2^*$ to which $\hat{\Theta}_1$ and $\hat{\Theta}_2$ converge, we have

$$
L_{1\theta_1\theta_1} = E \left[ \frac{\partial L_{i(p)1}(\Theta_1)}{\partial \Theta_1} \right] = -E \left[ \frac{\partial L_{i(p)1}(\Theta_1)}{\partial \Theta_1} \right] = -V_{L_1L_1} \text{ and }
$$
$$
L_{2\theta_2\theta_2} = E \left[ \frac{\partial L_{i(p)2}(\Theta_2)}{\partial \Theta_2} \right] = -E \left[ \frac{\partial L_{i(p)2}(\Theta_2)}{\partial \Theta_2} \right] = -V_{L_2L_2}.
$$

We can employ the derivative $L_{i(p)1}(\Theta_1)$ of with respect to $\Theta_1$ and of $L_{i(p)2}(\Theta_1, \Theta_2)$ with respect to $\Theta_2$ to compute $L_{i(p)1\theta_1\theta_1}$ and $L_{i(p)2\theta_2\theta_2}$ for all individuals, which can then be used to compute the sample analog of $L_{1\theta_1\theta_1}$ and $L_{2\theta_2\theta_2}$. This leaves us with the problem of constructing sample analogs of $L_{2\theta_2\theta_2}$, $L_{3\theta_1\theta_1}$, $L_{3\theta_1\theta_1}$, and $L_{3\theta_2\theta_2}$. While it is straightforward to compute sample analog of $L_{3\theta_1\theta_1}$, computation of sample analogs of $L_{2\theta_2\theta_2}$, $L_{3\theta_1\theta_1}$, and $L_{3\theta_2\theta_2}$ can be challenging. In the next subsections we derive the derivative of $L_{i(p)2\theta_2}(\Theta_1, \Theta_2)$ and $L_{i(p)3\theta_2}(\Theta_1, \Theta_1, \Theta_2)$ with respect to $\Theta_1$ and the derivative of $L_{i(p)3\theta_1}(\Theta_1, \Theta_2, \Theta_2)$ with respect to $\Theta_2$. 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, $u_{it}$, given in (C-1), for an individual $i$, $i \in I_p$, in the $(m \times p)$ matrix $E_{i(p)} = [u_{i1}, \ldots, u_{ip}]$, write $u_{i(p)} = vec(E_{i(p)})$, where ‘vec()’ is the vectorization operator, and define

$$
W_{ui(p)} = E_{i(p)}K_pE_{i(p)}' \text{ and } B_{ui(p)} = E_{i(p)}J_pE_{i(p)}', \quad \text{(F-10)}
$$

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, $E_p = e_pe_p'$.

Below we show that

$$
\frac{\partial L_{i(p)}}{\partial \delta} = 2Z_{i(p)}Q_{ui(p)}^{-1}u_{i(p)},
$$
$$
\frac{\partial L_{i(p)}}{\partial \theta} = -2Z_i\kappa \left[ \Sigma_{i(p)}^{-1}E_{i(p)}J_p + \Sigma_{\epsilon\epsilon}^{-1}E_{i(p)}K_p \right] e_p,
$$
$$
\frac{\partial L_{i(p)}}{\partial vec(\Sigma_{\epsilon\epsilon})} = -\frac{1}{2} vec \left( \Sigma_{i(p)}^{-1} + (p - 1)\Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{i(p)}^{-1}B_{ui(p)}\Sigma_{i(p)}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}W_{ui(p)}\Sigma_{\epsilon\epsilon}^{-1} \right),
$$

$$
\frac{\partial L_{i(p)}}{\partial \kappa} = -p\sigma_\alpha^2 \left[ \Sigma_{i(p)}^{-1} - \Sigma_{i(p)}^{-1}B_{ui(p)}\Sigma_{i(p)}^{-1} \right] \kappa + Z_i\delta \left[ \Sigma_{i(p)}^{-1}E_{i(p)}J_p + \Sigma_{\epsilon\epsilon}^{-1}E_{i(p)}K_p \right] e_p,
$$

and

$$
\frac{\partial L_{i(p)}}{\partial \sigma_\alpha} = -\frac{1}{2} [ vec(\Sigma_{i(p)}') - vec(\Sigma_{i(p)}^{-1}B_{ui(p)}\Sigma_{i(p)}^{-1}) ] vec(\kappa), \quad \text{(F-11)}
$$

13
where ‘vech(·)’ operator is column-wise vectorization of the lower triangle of the symmetric matrix Σ

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) $\text{tr}(ABCD) = \text{tr}(CDAB) = \text{vech}(A')' (D' \otimes B) \text{vech}(C) = \text{vech}(C')' (B' \otimes D) \text{vech}(A)$,

(d) $\frac{\partial \text{tr}(CB^{-1})}{\partial B} = -(B^{-1}CB^{-1})'$,

(e) $\frac{\partial x'}{\partial x} = x \otimes I_n + I_n \otimes x$, 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

$$L_{i(p)} = \frac{-mp}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega_{u(p)}| - \frac{1}{2} Q_{i(p)}(\delta, \tilde{\delta}, \Sigma_{ee}, \kappa, \sigma^2)$$

Then

$$\frac{\partial L_{i(p)}}{\partial \Sigma_{ee}} = \frac{1}{2} \frac{\partial \ln |\Omega_{u(p)}|}{\partial \Sigma_{ee}} - \frac{1}{2} \frac{\partial Q_{i(p)}(\delta, \tilde{\delta}, \Sigma_{ee}, \kappa, \sigma^2)}{\partial \Sigma_{ee}}.$$

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

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

For any given $\delta$ and $\tilde{\delta}$ we have

$$Q_{i(p)}() = u_{i(p)}' [K_p \otimes \Sigma_{ee}^{-1}] u_{i(p)} + u_{i(p)}' [J_p \otimes \Sigma_{(p)}^{-1}] u_{i(p)}$$

$$= \text{vech}(E_{i(p)})' [K_p \otimes \Sigma_{ee}^{-1}] \text{vech}(E_{i(p)}) + \text{vech}(E_{i(p)})' [J_p \otimes \Sigma_{(p)}^{-1}] \text{vech}(E_{i(p)})$$

From (c) we get

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

Using (F-10) we obtain

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

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

14
and from (d) we get

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

(F-13)

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

$$\frac{\partial L_{i(p)}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} = -\frac{1}{2}\text{vech}\left(\Sigma_{i(p)}^{-1} + (p - 1)\Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{i(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}W_{ui(p)}\Sigma_{\epsilon\epsilon}^{-1}\right).$$

(F-14)

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

$$d(\ln |\Omega_{u(p)}|) = d(\ln(|\Sigma_{\epsilon\epsilon}|^{-1}|\Sigma_{(p)}|)) = d(\ln(|\Sigma_{(p)}|)) = \text{vec}[\Sigma_{(p)}^{-1}\text{vech}(d(\Sigma_{(p)}))]$$

$$= \text{vec}[\Sigma_{(p)}^{-1}]\text{vec}[pd(\sigma^2_{\alpha})\Sigma_{\kappa} + p\sigma^2_{\delta}d(\Sigma_{\kappa})]$$

$$= \text{vec}[\Sigma_{(p)}^{-1}]\text{vec}[pd(\sigma^2_{\alpha})\Sigma_{\kappa} + p\sigma^2_{\delta}(\kappa \otimes I_m + I_m \otimes \kappa)d(\kappa)],$$

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

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

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

$$= -\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})\text{vec}(pd(\sigma^2_{\alpha})\Sigma_{\kappa} + p\sigma^2_{\delta}d(\Sigma_{\kappa}))$$

$$+ \text{vec}(\Sigma_{\epsilon\epsilon}^{-1})\text{vec}(d(B_{ui(p)})) + \text{vec}(\Sigma_{\epsilon\epsilon}^{-1})\text{vec}(d(W_{ui(p)}))$$

$$= -\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})\text{vec}\left(pd(\sigma^2_{\alpha})\Sigma_{\kappa} + p\sigma^2_{\delta}(\kappa \otimes I_m + I_m \otimes \kappa)d(\kappa)\right)$$

$$+ \text{vec}(\Sigma_{\epsilon\epsilon}^{-1})\left[ (I_m \otimes \hat{E}_{i(p)}J_p)\text{vec}(d(\hat{E}_{i(p)}')) + (\hat{E}_{i(p)}J_p' \otimes I_m)\text{vec}(d(\hat{E}_{i(p)}')) \right]$$

$$+ \text{vec}(\Sigma_{\epsilon\epsilon}^{-1})\left[ (I_m \otimes \hat{E}_{i(p)}K_p)\text{vec}(d(\hat{E}_{i(p)}')) + (\hat{E}_{i(p)}K_p' \otimes I_m)\text{vec}(d(\hat{E}_{i(p)}')) \right]$$

$$= -\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})\text{vec}(pd(\sigma^2_{\alpha})\Sigma_{\kappa} + p\sigma^2_{\delta}(\kappa \otimes I_m + I_m \otimes \kappa)d(\kappa))$$

$$- \text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})\text{vec}(\Sigma_{\kappa}d(\sigma^2_{\alpha}) - p\sigma^2_{\delta}\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})(\kappa \otimes I_m + I_m \otimes \kappa)d(\kappa))$$

$$- \text{vec}(\Sigma_{\epsilon\epsilon}^{-1})\left[ (I_m \otimes \hat{E}_{i(p)}J_p)(\hat{Z}'_{\delta}\delta d\kappa \otimes e_p) + (\hat{E}_{i(p)}J_p' \otimes I_m)(e_p \otimes \hat{Z}'_{\delta}\delta d\kappa) \right]$$

$$- \text{vec}(\Sigma_{\epsilon\epsilon}^{-1})\left[ (I_m \otimes \hat{E}_{i(p)}K_p)(\hat{Z}'_{\delta}\delta d\kappa \otimes e_p) + (\hat{E}_{i(p)}K_p' \otimes I_m)(e_p \otimes \hat{Z}'_{\delta}\delta d\kappa) \right]$$

$$= -p\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})\text{vec}(\Sigma_{\kappa}d(\sigma^2_{\alpha}) - p\sigma^2_{\delta}\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})(\kappa \otimes I_m + I_m \otimes \kappa)d(\kappa))$$

$$- \text{vec}(J_p'\hat{E}_{i(p)}\Sigma_{(p)}^{-1})(\hat{Z}'_{\delta}\delta d\kappa \otimes e_p) + \text{vec}(\Sigma_{(p)}^{-1}\hat{E}_{i(p)}J_p')(e_p \otimes \hat{Z}'_{\delta}\delta d\kappa)$$

$$- \text{vec}(K_p'\hat{E}_{i(p)}\Sigma_{\epsilon\epsilon}^{-1})(\hat{Z}'_{\delta}\delta d\kappa \otimes e_p) + \text{vec}(\Sigma_{\epsilon\epsilon}^{-1}\hat{E}_{i(p)}K_p')(e_p \otimes \hat{Z}'_{\delta}\delta d\kappa)$$

$$= -p\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})\text{vec}(\Sigma_{\kappa}d(\sigma^2_{\alpha}) - 2p\sigma^2_{\delta}\text{vec}(\Sigma_{(p)}^{-1}B_{ui(p)}\Sigma_{(p)}^{-1})(\kappa \otimes I_m + I_m \otimes \kappa)d(\kappa))$$

$$- 2p\sigma^2_{\delta}\text{vec}(J_p'\hat{E}_{i(p)}\Sigma_{(p)}^{-1} + K_p'\hat{E}_{i(p)}\Sigma_{\epsilon\epsilon}^{-1})\hat{Z}'_{\delta}\delta d\kappa.$$

(F-16)
where the second equality follows from employing (d), and the fact that $\Sigma_{ee}$ being given, $d\Sigma(p) = d(\Sigma_{ee} + p\sigma_2^2\Sigma_e) = pd(\sigma_2^2\Sigma_e)$. Since $\Sigma_{k} = \kappa \kappa'$, 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

$$
\frac{\partial Q_{i(p)}(\cdot)}{\partial \kappa} = -2p\sigma_{\alpha}^2[\Sigma_{-1}(p)B_{ui(p)}\Sigma_{-1}(p)k] - 2\bar{Z}_{i}\bar{\delta} \left[ \Sigma_{-1}(p)\tilde{E}_{i(p)}J_\rho + \Sigma_{ee}^{-1}\tilde{E}_{i(p)}K_\rho \right] e_p. \quad (F-17)
$$

Similarly, from (F-15) we obtain

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

Combining (F-17) and (F-18) we get the expression in (F-11) for $\frac{\partial \mathcal{L}_{i(p)}}{\partial \sigma_{\alpha}^2}$.

Again, from (F-15) and (F-16) respectively we obtain $\frac{\partial \ln |\Omega_{ui(p)}|}{\partial \sigma_{\alpha}^2} = p\text{vec}[\Sigma_{-1}(p)']\text{vec}(\Sigma_{k})$ and $\frac{\partial Q_{i(p)}(\cdot)}{\partial \sigma_{\alpha}^2} = -p\text{vec}[\Sigma_{-1}(p)B_{ui(p)}\Sigma_{-1}(p)']\text{vec}(\Sigma_{k})$, 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

$$
\frac{\partial \mathcal{L}_{i(p)}}{\partial \delta} = -2\bar{Z}_{i}\kappa' \left[ \Sigma_{-1}(p)\tilde{E}_{i(p)}J_\rho + \Sigma_{ee}^{-1}\tilde{E}_{i(p)}K_\rho \right] e_p. \quad (F-19)
$$

### F.2 Derivative of $\mathcal{L}_{i(p)2\Theta_2}$ with respect to $\Theta_1$

Let us begin by deriving the derivative of score functions, $\mathcal{L}_{i(p)2\Theta_2}$, of second stage likelihood with respect to $\Theta_1$. 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 $\Theta_1$. Now, we know that $\Theta_1$ enters the second stage of the sequential estimator through $\bar{Z}_{i}\delta + \alpha_1(\Theta_1)$ and $\Sigma_{ee}^{-1}\tilde{E}_{it}(\Theta_1)$, 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 $\Theta_1$ 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, $\Theta_{2f}$, coefficients of the selection equation $\Theta_{2s}$ and $\rho_{\tilde{E}_i\tilde{E}_i}$, the correlation between the idiosyncratic components of the financial constraint and the selection equation. In matrix form we can write

$$
\mathcal{L}_{i(p)2\Theta_2}\Theta_1 = \frac{\partial \mathcal{L}_{i(p)2\Theta_2}}{\partial \Theta_1} = \sum_{t=1}^{p} \frac{\partial \mathcal{L}_{it2\Theta_2}}{\partial \Theta_1} = \sum_{t=1}^{p} \begin{bmatrix}
\frac{\partial \mathcal{L}_{it2\Theta_2f}}{\partial \Theta_1} \\
\frac{\partial \mathcal{L}_{it2\Theta_2s}}{\partial \Theta_1} \\
\frac{\partial \mathcal{L}_{it2\rho_{\tilde{E}_i\tilde{E}_i}}}{\partial \Theta_1}
\end{bmatrix},
$$

where the score functions, $\mathcal{L}_{it2\Theta_2f}$, $\mathcal{L}_{it2\Theta_2s}$, and $\mathcal{L}_{it2\rho_{\tilde{E}_i\tilde{E}_i}}$, above are the score functions of the log likelihood function for bivariate probit when it belongs to CIS3 and CIS3.5, and are given by

$$
\mathcal{L}_{it2\Theta_2f}(\Theta_1, \Theta_2) = \frac{q_{it}f_{it}g_{it}x_{it}f}{\Phi_2}, \quad \mathcal{L}_{it2\Theta_2s}(\Theta_1, \Theta_2) = \frac{q_{its}g_{its}x_{its}}{\Phi_2}, \quad \text{and} \quad \mathcal{L}_{it2\rho_{\tilde{E}_i\tilde{E}_i}}(\Theta_1, \Theta_2) = \frac{q_{itf}q_{its}\phi_2}{\Phi_2}
$$

(F-20)
Given the above, we have \( \Theta^\ast \), what is challenging to obtain is the derivative of \( s_i \) when \( s_i \). When \( \rho \) = 2, \( \hat{\rho} \) = 0, and \( \rho \) = 1. When \( s_i = 0 \) for CIS2.5, we have \( \gamma_i = \Theta^\ast \) when \( s_i = 0 \), the score functions remain the same as in (F-20) when \( s_i = 1 \), but are

\[
\mathcal{L}_{it2\Theta_2} (\Theta_1, \Theta_2) = 0_{\Theta_2}, \quad \mathcal{L}_{it2\Theta_1} (\Theta_1, \Theta_2) = - \frac{\phi(-\Theta^\ast)}{\Phi(-\Theta^\ast)} \Theta^\ast, \text{ and } \mathcal{L}_{it2\zeta_0} (\Theta_1, \Theta_2) = 0
\]

when \( s_i = 0 \), and where \( 0_{\Theta_2} \) is a vector of zeros.

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

\[
\frac{\partial \mathcal{L}_{2\Theta_2}}{\partial \Theta_1} = q_j \left\{ \frac{\partial}{\partial \Theta_1} \left( \frac{g_j}{\Phi_2} \right) \right\} + \frac{\partial}{\partial \Theta_1} \left( \frac{\phi_2}{\Phi_2} \right)
\]

where \( j \in \{ f, s \} \) and

\[
\frac{\partial \mathcal{L}_{2\zeta_0}}{\partial \Theta_1} = q_j q_s \left\{ \frac{\partial}{\partial \Theta_1} \left( \frac{\phi_2}{\Phi_2} \right) \right\}
\]

when the firm year observation, \( it \), is such that it belongs to CIS3 and CIS3.5, and CIS2.5 when \( s_i = 1 \). When \( s_i = 0 \) for CIS2.5 we have \( \frac{\partial \mathcal{L}_{2\Theta_2}}{\partial \Theta_1} = 0_{\Theta_2} \), \( \frac{\partial \mathcal{L}_{2\zeta_0}}{\partial \Theta_1} = 0 \), and

\[
\frac{\partial \mathcal{L}_{2\Theta_2}}{\partial \Theta_1} = - \left\{ \frac{\partial}{\partial \Theta_1} \left( \frac{\phi(-\gamma_i)}{\Phi(-\gamma_i)} \right) \right\} + \frac{\phi(-\gamma_i)}{\Phi(-\gamma_i)} \frac{\partial \Theta^\ast}{\partial \Theta_1}
\]

To obtain expressions for (F-22), (F-23), and (F-24) we need the derivative of \( \frac{\partial}{\partial \Theta_1} \), \( j \in \{ f, s \} \), with respect to \( \phi \) and \( \gamma_i \), the derivative of \( \frac{\partial}{\partial \Theta_1} \) with respect to \( \varphi \) and \( \gamma_i \), and the derivative of \( \frac{\partial g_i}{\partial \Theta_1} \) with respect to \( \gamma_i \). While these can be easily obtained and can be found in Greene (2002), what is challenging to obtain is the derivative of \( X_i^f \) and \( X_i^s \) with respect to \( \Theta_1 \).
\[
\frac{\partial X_{it}^j}{\partial \theta'} = \begin{bmatrix}
\frac{\partial X_{it}^j}{\partial (Z_{it}^j + \alpha_i)} & \frac{\partial X_{it}^j}{\partial (Z_{it}^j + \alpha_i)} & \frac{\partial X_{it}^j}{\partial (Z_{it}^j + \alpha_i)} & \frac{\partial X_{it}^j}{\partial \text{vech}(\Sigma_{ee})} & \frac{\partial X_{it}^j}{\partial \sigma^2_{\alpha}}
\end{bmatrix},
\]

where \( j \in \{f, s\} \). While \( \frac{\partial X_{it}^j}{\partial \sigma^2_{\alpha}} = 0 \), below we show that

\[
\frac{\partial (Z_{it}^j + \alpha_i)}{\partial \kappa'} = \frac{1}{U_{dr}^2} \sum_{t=1}^{p} \left\{ \left[ U_{nr}^2 - U_{dr} F_{dr} \right] (Z_{it}^j \delta \kappa' - r_{it}' \Sigma_{ee}^{-1} + U_{nr} F_{dr} - U_{dr} F_{nr}) \right\} \kappa' \Sigma_{ee}^{-1} Z_{it}^j.
\]

\[
\frac{\partial (Z_{it}^j + \alpha_i)}{\partial \delta'} = \frac{1}{U_{dr}^2} \sum_{t=1}^{p} \left\{ \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \text{vech}(\Sigma_{ee})' \left[ \text{vech}(\Sigma_{ee})' \right] (Z_{it}^j \delta \kappa' - r_{it}' \Sigma_{ee}^{-1} + U_{nr} F_{dr} - U_{dr} F_{nr}) \right\} (\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1})' L_m.
\]

\[
\frac{\partial (Z_{it}^j + \alpha_i)}{\partial \sigma^2_{\alpha}} = \frac{1}{2 \sigma^4_{\alpha} U_{dr}^2} (U_{dr} F_{nr} - U_{nr} F_{dr}).
\]

\[
\frac{\partial \Sigma_{ee}^{-1} e_{it}}{\partial \delta'} = -\tilde{\Sigma}_{ee}^{-1} Z_{it}^j + \frac{\tilde{\Sigma}_{ee}^{-1} \kappa}{U_{dr}^2} \sum_{t=1}^{p} \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \kappa' \Sigma_{ee}^{-1} Z_{it}^j.
\]

\[
\frac{\partial \Sigma_{ee}^{-1} e_{it}}{\partial \kappa'} = -\tilde{\Sigma}_{ee}^{-1} \kappa Z_{it}^j + \frac{\tilde{\Sigma}_{ee}^{-1} \kappa}{U_{dr}^2} \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \kappa' \Sigma_{ee}^{-1} Z_{it}^j.
\]

\[
\frac{\partial \Sigma_{ee}^{-1} e_{it}}{\partial \sigma^2_{\alpha}} = -\frac{\Sigma_{ee}^{-1} \kappa}{2 U_{dr}^2} \sum_{t=1}^{p} \left[ (U_{nr}^2 - U_{dr} F_{dr}) \text{vech}(\Sigma_{ee})' \left[ \text{vech}(\Sigma_{ee})' \right] (U_{nr} F_{dr} - U_{dr} F_{nr}) \right\} (\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1})' L_m.
\]

\[
\frac{\partial \Sigma_{ee}^{-1} e_{it}}{\partial \sigma^2_{\alpha}} = \frac{\Sigma_{ee}^{-1} \kappa}{2 \sigma^4_{\alpha} U_{dr}^2} (U_{dr} F_{nr} - U_{nr} F_{dr}),
\]

where

\[
U_{nr} = \int \alpha \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}' \Sigma_{ee}^{-1} \epsilon_{it}\right) \phi(\alpha) d\alpha, \quad F_{nr} = \int \alpha^3 \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}' \Sigma_{ee}^{-1} \epsilon_{it}\right) \phi(\alpha) d\alpha,
\]

\[
U_{dr} = \int \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}' \Sigma_{ee}^{-1} \epsilon_{it}\right) \phi(\alpha) d\alpha, \quad F_{dr} = \int \alpha^2 \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}' \Sigma_{ee}^{-1} \epsilon_{it}\right) \phi(\alpha) d\alpha,
\]

18
and \( \mathbf{r}_{it} = \mathbf{x}_{it} - \mathbf{Z}_{it}\delta - \kappa \bar{Z}_i\delta \). 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 \( \bar{Z}_i^t\delta + \hat{\alpha}_i \) and \( \bar{\Sigma}^{-1}_{ee}\hat{\epsilon}_{it} \) and with respect to \( \Theta_1 \)

Let us first consider the derivative of \( \bar{Z}_i^t\delta + \alpha_i \) and \( \bar{\Sigma}^{-1}_{ee}\hat{\epsilon}_{it} \) with respect to \( \delta' \). We have

\[
\frac{\partial (\bar{Z}_i^t\delta + \hat{\alpha}_i)}{\partial \delta'} = \frac{\partial \bar{Z}_i^t\delta}{\partial \delta'} + \frac{\partial \hat{\alpha}_i}{\partial \delta'} = 0 + \frac{\partial}{\partial \delta'} \left[ \int\alpha \exp(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}^t \bar{\Sigma}^{-1}_{ee} \epsilon_{it}) \phi(\alpha) d\alpha \right]
\]

\[
= 0 - \int \alpha \exp(.) \phi(\alpha) d\alpha \int \exp(.) \epsilon_{it}^t \bar{\Sigma}^{-1}_{ee} \epsilon_{it} \phi(\alpha) d\alpha - \int \alpha \exp(.) \phi(\alpha) d\alpha \int \exp(.) \epsilon_{it}^t \bar{\Sigma}^{-1}_{ee} \epsilon_{it} \phi(\alpha) d\alpha,
\]

\[
\text{To derive the above result in (F-27) we used the fact that } \frac{\partial (\epsilon_{it}^t \Sigma^{-1}_{ee} \epsilon_{it})}{\partial \delta'} = \frac{\partial (\epsilon_{it}^t \Sigma^{-1}_{ee} \epsilon_{it})}{\partial \delta'} = -2\epsilon_{it}^t \Sigma^{-1}_{ee} \bar{Z}_i^t.
\]

Taking into account the fact that \( \epsilon_{it} = \mathbf{x}_{it} - \mathbf{Z}_{it}\delta - (\bar{Z}_i^t\delta + \alpha_i)\kappa \), after some rearrangements it can be shown that

\[
\frac{\partial \hat{\alpha}_i}{\partial \delta'} = -\frac{1}{U_{dr}^2} \sum_{t=1}^{T} \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \kappa' \Sigma^{-1}_{ee} \bar{Z}_i^t,
\]

where

\[
U_{nr} = \int \alpha \exp(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}^t \Sigma^{-1}_{ee} \epsilon_{it}) \phi(\alpha) d\alpha, \quad F_{nr} = \int \alpha^3 \exp(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}^t \Sigma^{-1}_{ee} \epsilon_{it}) \phi(\alpha) d\alpha
\]

\[
U_{dr} = \int \exp(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}^t \Sigma^{-1}_{ee} \epsilon_{it}) \phi(\alpha) d\alpha, \quad F_{dr} = \int \alpha^2 \exp(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}^t \Sigma^{-1}_{ee} \epsilon_{it}) \phi(\alpha) d\alpha.
\]

Hence we have

\[
\frac{\partial (\bar{Z}_i^t\delta + \hat{\alpha}_i)}{\partial \delta'} = -\frac{1}{U_{dr}^2} \sum_{t=1}^{T} \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \kappa' \Sigma^{-1}_{ee} \bar{Z}_i^t, \quad (F-29)
\]

and

\[
\frac{\partial \Sigma^{-1}_{ee} \epsilon_{it}}{\partial \delta'} = \frac{\partial \Sigma^{-1}_{ee} (\mathbf{x}_{it} - \mathbf{Z}_{it}\delta - \bar{Z}_i^t\kappa)}{\partial \delta'} - \frac{\partial \Sigma^{-1}_{ee} \hat{\alpha}_i \kappa}{\partial \delta'} = -\bar{\Sigma}^{-1}_{ee} \bar{Z}_i^t + \frac{\bar{\Sigma}^{-1}_{ee} \kappa}{U_{dr}^2} \sum_{t=1}^{T} \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \kappa' \Sigma^{-1}_{ee} \bar{Z}_i^t, \quad (F-30)
\]

From (F-29) and (F-30) we can see that while \( \frac{\partial (\bar{Z}_i^t\delta + \hat{\alpha}_i)}{\partial \delta'} \) for an individual \( i \) remains the same for all time periods, \( \frac{\partial \Sigma^{-1}_{ee} \epsilon_{it}}{\partial \delta'} \) varies with time. Similarly it can be shown that

\[
\frac{\partial (\bar{Z}_i^t\delta + \hat{\alpha}_i)}{\partial \delta'} = \bar{Z}_i^t - \frac{1}{U_{dr}^2} \sum_{t=1}^{T} \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \kappa' \Sigma^{-1}_{ee} \kappa \bar{Z}_i^t = \bar{Z}_i^t - \frac{p}{U_{dr}^2} \left[ U_{nr}^2 - U_{dr} F_{dr} \right] \kappa' \Sigma^{-1}_{ee} \kappa \bar{Z}_i^t, \quad (F-31)
\]
and

\[
\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\tilde{\epsilon}_{i\epsilon}}{\partial \delta'} = \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}(x_{i\epsilon} - Z_{i\epsilon}'\delta - Z_{i\epsilon}'\delta\kappa)}{\partial \delta'} - \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\alpha}_i\kappa}{\partial \delta'} = -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\kappa\tilde{\epsilon}_{i\epsilon}' + \frac{p\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\kappa}{U_{dr}'^2}
\]

\[U_{nr}' - U_{dr}F_{dr}\] \[\kappa'\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\kappa\tilde{\epsilon}_{i\epsilon}'.
\]

(F-32)

Let us now consider the derivative of \(\tilde{Z}_{i\delta}' + \hat{\alpha}_{i\epsilon}\) with respect to \(\kappa\). We have

\[
\frac{\partial (\tilde{Z}_{i\delta}' + \hat{\alpha}_{i\epsilon})}{\partial \kappa'} = \frac{\partial}{\partial \kappa'} \left[ \int \alpha \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}\epsilon_{it}\right)\phi(\alpha)d\alpha \right]
\]

\[= -\frac{1}{(\int \exp(\cdot)\phi(\alpha)d\alpha)^2} \sum_{t=1}^{p} \left[ \int \alpha \epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}(\tilde{Z}_{i\delta}' + \hat{\alpha}_{i\epsilon})\phi(\alpha)d\alpha \int \exp(\cdot)\phi(\alpha)d\alpha \right]
\]

\[= -\frac{1}{(\int \exp(\cdot)\phi(\alpha)d\alpha)^2} \sum_{t=1}^{p} \left[ \int \alpha \exp(\cdot)\epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}(\tilde{Z}_{i\delta}' + \hat{\alpha}_{i\epsilon})\phi(\alpha)d\alpha \int \exp(\cdot)\phi(\alpha)d\alpha \right],
\]

(F-33)

which after simplification can be written as

\[
\frac{\partial (\tilde{Z}_{i\delta}' + \hat{\alpha}_{i\epsilon})}{\partial \kappa'} = -\frac{1}{U_{dr}'^2} \sum_{t=1}^{p} \left\{ \left[ U_{nr}'^2 - U_{dr}F_{dr} \right](\tilde{Z}_{i\delta}'\kappa' - r_{it}'\kappa)\gamma_{\epsilon\epsilon}^{-1} + \left[ U_{nr}'F_{dr} - U_{dr}'F_{nr} \right] \kappa'\gamma_{\epsilon\epsilon}^{-1} \right\},
\]

(F-34)

where \(r_{it} = x_{i\epsilon} - Z_{i\epsilon}'\delta - Z_{i\epsilon}'\delta\kappa\) and \(U_{nr}', U_{dr}', F_{nr}', \) and \(F_{dr}'\) are given in (F-28). Also, it can be shown that

\[
\frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\tilde{\epsilon}_{i\epsilon}}{\partial \kappa'} = \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}(x_{i\epsilon} - Z_{i\epsilon}'\delta - Z_{i\epsilon}'\delta\kappa)}{\partial \kappa'} - \frac{\partial \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\alpha}_i\kappa}{\partial \kappa'}
\]

\[= -\tilde{\Sigma}_{\epsilon\epsilon}^{-1}\frac{Z_{i\delta}'}{U_{dr}'^2} - \tilde{\Sigma}_{\epsilon\epsilon}^{-1}\frac{\kappa}{U_{dr}'^2} \sum_{t=1}^{p} \left\{ \left[ U_{nr}'^2 - U_{dr}F_{dr} \right](\tilde{Z}_{i\delta}'\kappa' - r_{it}'\kappa)\gamma_{\epsilon\epsilon}^{-1} + \left[ U_{nr}'F_{dr} - U_{dr}'F_{nr} \right] \kappa'\gamma_{\epsilon\epsilon}^{-1} \right\},
\]

(F-35)

Now consider the derivative of \(\tilde{Z}_{i\delta}' + \hat{\alpha}_{i\epsilon}\) with respect to \(\text{vech}(\Sigma_{\epsilon\epsilon})\). We have

\[
\frac{\partial (\tilde{Z}_{i\delta}' + \hat{\alpha}_{i\epsilon})}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} = \frac{\partial \hat{\alpha}_{i\epsilon}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} = \frac{\partial}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} \left[ \int \alpha \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}\epsilon_{it}\right)\phi(\alpha)d\alpha \right]
\]

\[= \frac{-1}{2} \left[ \int \alpha \psi(\alpha) \frac{\partial \sum_{t=1}^{p} \epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}\epsilon_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} d\alpha \int \psi(\alpha)d\alpha - \int \alpha \psi(\alpha)d\alpha \int \psi(\alpha) \frac{\partial \sum_{t=1}^{p} \epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}\epsilon_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} d\alpha \right],
\]

where \(\psi(\alpha) = \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}\epsilon_{it}\right)\phi(\alpha)\). With \(\frac{\partial \sum_{t=1}^{p} \epsilon_{it}'\gamma_{\epsilon\epsilon}^{-1}\epsilon_{it}}{\partial \text{vech}(\Sigma_{\epsilon\epsilon})} = \sum_{t=1}^{p} \text{vec}(\Sigma_{\epsilon\epsilon})' \epsilon_{it}'\epsilon_{it}(\Sigma_{\epsilon\epsilon})'L_{it}^m\)
the above can be written as

\[
\frac{\partial \hat{\alpha}_i}{\partial \text{vech}(\Sigma_{ee})} = \frac{1}{2(\int \psi(\alpha) d\alpha)^2} \sum_{t=1}^{T} \left[ \int \alpha \psi(\alpha) \text{vec}((\Sigma_{ee}^{-1})' \epsilon_i \epsilon_i' (\Sigma_{ee}^{-1})')' L_m' d\alpha \int \psi(\alpha) d\alpha \right. \\
- \left. \int \psi(\alpha) \text{vec}((\Sigma_{ee}^{-1})' \epsilon_i \epsilon_i' (\Sigma_{ee}^{-1})')' L_m' d\alpha \int \alpha \psi(\alpha) d\alpha \right]
\]

\[
= \frac{1}{2U_{dr}^2} \sum_{t=1}^{T} \left[ \int \alpha \psi(\alpha) \text{vec}(\epsilon_i \epsilon_i' (\Sigma_{ee}^{-1})')' L_m' d\alpha U_{dr} \\
- U_{nr} \left( \int \psi(\alpha) \text{vec}(\epsilon_i \epsilon_i' (\Sigma_{ee}^{-1})')' L_m' d\alpha \right) \right]
\]

\[
= \frac{1}{2U_{dr}^2} \sum_{t=1}^{T} \left[ \left( U_{dr} \alpha \text{vec}(\epsilon_i \epsilon_i')' - U_{nr} \text{vec}(\epsilon_i \epsilon_i')' \psi(\alpha) d\alpha \right) (\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1})' L_m' \right],
\]

(F-36)

where \( L_m \) is an elimination matrix. To simply further, write \( \epsilon_i \) as \( \epsilon_i = x_i - Z_i \delta - k \hat{Z}_i \delta - \kappa \alpha = r_{it} - \kappa \alpha \), where \( r_{it} = x_i - Z_i \delta - k \hat{Z}_i \delta \). Then \( \epsilon_i \epsilon_i' = r_{it} r_{it}' - k r_{it} \alpha - r_{it} \kappa \alpha + \kappa \kappa' \alpha^2 \), then (F-36) after some simplification can be written as

\[
\frac{\partial \hat{\alpha}_i}{\partial \text{vech}(\Sigma_{ee})} = \frac{1}{2U_{dr}^2} \sum_{t=1}^{T} \left[ \left( U_{dr}^2 - U_{dr} F_{dr} \right) \text{vec}(k r_{it} + r_{it} \kappa)' + \left( U_{dr} F_{nr} - U_{nr} F_{dr} \right) \text{vec}(\Sigma_{ee})' \right] (\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1})' L_m'.
\]

(F-37)

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

\[
d(\Sigma_{ee}^{-1} \kappa \hat{\alpha}_i) = d(\Sigma_{ee}) \Sigma_{ee}^{-1} \kappa \hat{\alpha}_i + \Sigma_{ee} d(\Sigma_{ee}^{-1}) \kappa \hat{\alpha}_i + \Sigma_{ee} \Sigma_{ee}^{-1} \kappa d(\hat{\alpha}_i). \tag{F-38}
\]

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

\[
\frac{\partial (\Sigma_{ee}) \Sigma_{ee}^{-1} \kappa \hat{\alpha}_i}{\partial \text{vech}(\Sigma_{ee})} = \frac{1}{2} (\kappa \hat{\alpha}_i \Sigma_{ee}^{-1} \otimes I_m) \text{vec}((\text{dg}(\Sigma_{ee}))^{-1/2}) \frac{\partial \text{vec}(\Sigma_{ee})}{\partial \text{vech}(\Sigma_{ee})}
\]

\[
= \frac{1}{2} (\kappa \hat{\alpha}_i \Sigma_{ee}^{-1} \otimes I_m) \text{vec}((\text{dg}(\Sigma_{ee}))^{-1/2})' L_m'. \tag{F-39}
\]

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

\[
\frac{\Sigma_{ee} \partial (\Sigma_{ee}^{-1}) \kappa \hat{\alpha}_i}{\partial \text{vech}(\Sigma_{ee})} = - (\kappa \hat{\alpha}_i \otimes \Sigma_{ee})' (\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1}) \frac{\partial \text{vec}(\Sigma_{ee})}{\partial \text{vech}(\Sigma_{ee})} = - (\kappa \hat{\alpha}_i \otimes \Sigma_{ee})' (\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1}) L_m'. \tag{F-40}
\]

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

\[
\frac{\Sigma_{ee}^{-1} \kappa \partial \hat{\alpha}_i}{\partial \text{vech}(\Sigma_{ee})} = \frac{\Sigma_{ee}^{-1} \kappa}{2U_{dr}^2} \sum_{t=1}^{T} \left[ \left( U_{dr}^2 - U_{dr} F_{dr} \right) \text{vec}((k r_{it} + r_{it} \kappa)') \\
+ \left( U_{dr} F_{nr} - U_{nr} F_{dr} \right) \text{vec}(\Sigma_{ee})' \right] (\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1})' L_m'. \tag{F-41}
\]

21
Combining (F-39), (F-40), and (F-41) we obtain
\[
\frac{\partial \hat{\Sigma}_e^{-1} \epsilon_{it}}{\partial \text{vech}(\Sigma_{ee})'} = \left[ (\epsilon_{it}' \Sigma_{ee}^{-1} \otimes I_m) \text{vec}((\text{dg}(\Sigma_{ee}))^{-1/2})' - (\epsilon_{it} \otimes \Sigma_e')'(\Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1})' \right] L'_m \\
- \frac{\hat{\Sigma}_e^{-1} \kappa}{2(U_{dr})^2} \sum_{t=1}^{p} \left( (U_{nr}^2 - U_{dr} F_{dr}) \text{vec}(\kappa r_{it}') + (U_{dr} F_{nr} - U_{nr} F_{dr}) \text{vec}(\kappa)'ight) \left( \Sigma_{ee}^{-1} \otimes \Sigma_{ee}^{-1} \right)' L'_m.
\]

Finally, let consider the derivative of \( \hat{Z}' \hat{\delta} + \hat{\alpha} \) with respect to \( \sigma^2_\alpha \). We have
\[
\frac{\partial (\hat{Z}' \hat{\delta} + \hat{\alpha})}{\partial \sigma^2_\alpha} = \frac{\partial \hat{\alpha}}{\partial \sigma^2_\alpha} = \frac{\partial}{\partial \sigma^2_\alpha} \left[ \int \alpha \exp(.) \phi(\alpha) d\alpha \right] = \frac{\int \alpha \exp(.) \frac{\partial \phi(\alpha)}{\partial \sigma^2_\alpha} d\alpha}{\int \exp(.) \phi(\alpha) d\alpha}.
\]
Given that \( \frac{\partial \phi(\alpha)}{\partial \sigma^2_\alpha} = -\frac{1}{2\sigma^4_\alpha} \phi(\alpha) + \frac{\sigma^2_\alpha}{2\sigma^4_\alpha} \phi(\alpha) \), the above after simplification reduces to
\[
\frac{\partial (\hat{Z}' \hat{\delta} + \hat{\alpha})}{\partial \sigma^2_\alpha} = \frac{1}{2\sigma^4_\alpha U_{dr}^2}(U_{dr} F_{nr} - U_{nr} F_{dr}),
\]
and we can write \( \frac{\partial \hat{\Sigma}_e^{-1} \epsilon_{it}}{\partial \text{vech}(\Sigma_{ee})'} \) as
\[
\frac{\partial \hat{\Sigma}_e^{-1} \epsilon_{it}}{\partial \text{vech}(\Sigma_{ee})'} = -\frac{\hat{\Sigma}_e^{-1} \kappa \partial (\hat{\alpha})}{\partial \sigma^2_\alpha} = -\frac{\hat{\Sigma}_e^{-1} \kappa}{2\sigma^4_\alpha U_{dr}^2}(U_{dr} F_{nr} - U_{nr} F_{dr}).
\]

**F.3 Derivative of \( L_{i(p)\Theta_3} \) with respect to \( \Theta_1 \) and \( \Theta_2 \)**

As stated earlier in order to construct error corrected standard errors of the structural parameters we also need sample analogs of \( L_{3\Theta_3, \Theta_1}, L_{3\Theta_3, \Theta_2}, \) and \( L_{3\Theta_3, \Theta_3} \) to construct \( B_s \) in (F-9). While it is straightforward to compute sample analog of \( L_{3\Theta_3, \Theta_1} \), computation of sample analogs of \( L_{3\Theta_3, \Theta_2} \) and \( L_{3\Theta_3, \Theta_3} \) needs some work. Here we derive the derivative of \( L_{i(p)\Theta_3}(\Theta_1, \Theta_2, \Theta_3) \) with respect to \( \Theta_1 \) and \( \Theta_2 \). Now, we know that
\[
\frac{\partial L_{i(p)\Theta_3}}{\partial \Theta'_j} = \sum_{t=1}^{p} \frac{\partial L_{i(p)\Theta_3}}{\partial \Theta'_j} = \sum_{t=1}^{p} \frac{\partial}{\partial \Theta'_j} \left[ X_{it}(\Theta_1, \Theta_2)(r_{it} - X_{it}'(.)'\Theta_3) \right] = \sum_{t=1}^{p} s_{it} \left[ \frac{X_{it}'}{\partial \Theta'_j}(r_{it} - X_{it}'(.)'\Theta_3) + X_{it}'(\hat{\Theta}_1, \hat{\Theta}_2) \frac{X_{it}'}{\partial \Theta'_j}(\hat{\Theta}_3) \right] j \in \{1, 2\},
\]
where

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

And \(X^{r}_{1it} = \{X^{r}_{11it}, X^{r}_{10it}\}'\) where \(X^{r}_{1it}\) and \(X^{r}_{0it}\) have been defined in equation (6) in the main text.

We know that \(\frac{\partial X^{r}_{it}}{\partial \Theta_1} = \frac{\partial X^{r}_{it}}{\partial \Theta_2} = 0\), that \(\frac{\partial X^{r}_{it}}{\partial \varphi_{it}} = 0\) and \(\frac{\partial X^{r}_{it}}{\partial \gamma_{it}} = 0\) 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 \(\Theta_1\) and \(\Theta_2\). We have

\[
\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \Theta'_1} = \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \varphi_{it}_j} \frac{\partial X^{r}_{it}_f}{\partial \Theta'_1} + \frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \gamma_{it}_j} \frac{\partial X^{r}_{it}_s}{\partial \Theta'_1}, \quad j \in \{0, 1\}, k \in \{1, 2\}.
\]

(F-46)

Given the functional form of \(C_{jk}(\Theta_1, \Theta_2)\) in equations (24) and (25), its derivative with respect to \(\varphi_{it}\) and \(\gamma_{it}\) can be easily obtained. The partial derivatives \(\frac{\partial X^{r}_{it}_f}{\partial \Theta'_1}\) and \(\frac{\partial X^{r}_{it}_s}{\partial \Theta'_1}\) have been worked out above. Now consider the derivative of \(C_{jk}(\Theta_1, \Theta_2)\) with respect to \(\Theta_2 = \{\Theta'_{2f}, \Theta'_{2s}, \rho_{\tilde{\xi}_{it}}\}'\).

\[
\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \Theta'_2} = \begin{bmatrix}
\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \varphi_{it}_j} \frac{\partial X^{r}_{it}_f}{\partial \Theta'_2} \\
\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \gamma_{it}_j} \frac{\partial X^{r}_{it}_s}{\partial \Theta'_2} \\
\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \rho_{\tilde{\xi}_{it}}} \frac{\partial X^{r}_{it}_s}{\partial \Theta'_2}
\end{bmatrix}'
\]

(F-47)

Again, given the functional form of \(C_{jk}(\Theta_1, \Theta_2)\), \(\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \varphi_{it}_j}\) and \(\frac{\partial C_{jk}(\Theta_1, \Theta_2)_{it}}{\partial \gamma_{it}_j}\) 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 \(\varphi_{it}, \gamma_{it}\) and \(\rho_{\tilde{\xi}_{it}}\), and these are stated in Greene (2002).

References


The UNU-MERIT WORKING Paper Series

2012-01 Maastricht reflections on innovation by Luc Soete
2012-02A methodological survey of dynamic microsimulation models by Jinjing Li and Cathal O'Donoghue
2012-03 Evaluating binary alignment methods in microsimulation models by Jinjing Li and Cathal O'Donoghue
2012-04 Estimates of the value of patent rights in China by Can Huang
2012-05 The impact of malnutrition and post traumatic stress disorder on the performance of working memory in children by Elise de Neubourg and Chris de Neubourg
2012-06 Cross-national trends in permanent earnings inequality and earnings instability in Europe 1994-2001 by Denisa Maria Sologon and Cathal O'Donoghue
2012-07 Foreign aid transaction costs by Frieda Vandeninden
2012-08A simulation of social pensions in Europe by Frieda Vandeninden
2012-09 The informal ICT sector and innovation processes in Senegal by Almamy Konté and Mariama Ndong
2012-10 The monkey on your back?! Hierarchical positions and their influence on participants' behaviour within communities of learning by Martin Rehm, Wim Gijseelaers and Mien Segers
2012-11 Do Ak models really lack transitional dynamics? by Yoseph Yilma Getachew
2012-12 The co-evolution of organizational performance and emotional contagion by R. Cowan, N. Jonard, and R. Weehuizen
2012-13 "Surfeiting, the appetite may sicken": Entrepreneurship and the happiness of nations by Wim Naudé, José Ernesto Amorós and Oscar Cristi
2012-14 Social interactions and complex networks by Daniel C. Opolot
2012-15 New firm creation and failure: A matching approach by Thomas Gries, Stefan Jungblut and Wim Naudé
2012-16 Gains from child-centred Early Childhood Education: Evidence from a Dutch pilot programme by Robert Bauchmüller
2012-17 Highly skilled temporary return, technological change and Innovation: The Case of the TRQN Project in Afghanistan by Melissa Siegel and Katie Kuschminder
2012-18 New Technologies in remittances sending: Opportunities for mobile remittances in Africa by Melissa Siegel and Sonja Fransen
2012-19 Implementation of cross-country migration surveys in conflict-affected settings: Lessons from the IS Academy survey in Burundi and Ethiopia by Sonja Fransen, Katie Kuschminder and Melissa Siegel
2012-20 International entrepreneurship and technological capabilities in the Middle East and North Africa by Juliane Brach and Wim Naudé
2012-21 Entrepreneurship, stages of development, and industrialization by Zoltan J. Ács and Wim Naudé
2012-22 Innovation strategies and employment in Latin American firms by Gustavo Crespi and Pluvia Zuniga
2012-23 An exploration of agricultural grassroots innovation in South Africa and implications for innovation indicator development by Brigid Letty, Zanele Shezi and Maxwell Mudhara
2012-24 Employment effect of innovation: microdata evidence from Bangladesh and Pakistan by Abdul Waheed
2012-25 Open innovation, contracts, and intellectual property rights: an exploratory empirical study by John Hagedoorn and Ann-Kristin Ridder
2012-26 Remittances provide resilience against disasters in Africa by Wim Naudé and Henri Bezuidenhout
2012-27 Entrepreneurship and economic development: Theory, evidence and policy by Wim Naudé
2012-28 Whom to target - an obvious choice? by Esther Schüring and Franziska Gassmann
2012-29 Sunk costs, extensive R&D subsidies and permanent inducement effects by Pere Arqué-Castells and Pierre Mohnen
2012-30 Assessing contingent liabilities in public-private partnerships (PPPs) by Emmanouil Sfakianakis and Mindel van de Laar
2012-31 Informal knowledge exchanges under complex social relations: A network study of handloom clusters in Kerala, India by Robin Cowan and Anant Kamath
2012-32 Proximate, intermediate and ultimate causality: Theories and experiences of growth and development by Adam Szirmai
2012-33 Institutions and long-run growth performance: An analytic literature review of the institutional determinants of economic growth by Richard Bluhm and Adam Szirmai
2012-34 Techniques for dealing with reverse causality between institutions and economic performance by Luciana Cingolani and Denis de Crombrugghe
2012-35 Preliminary conclusions on institutions and economic performance by Denis de Crombrugghe and Kristine Farla
2012-36 Stylized facts of governance, institutions and economic development. Exploring the institutional profiles database by Bart Verspagen
2012-37 Exploring the Panel Components of the Institutional Profiles Database (IPD) by Luciana Cingolani and Denis de Crombrugghe
2012-38 Institutions and credit by Kristine Farla
2012-39 Industrial policy for growth by Kristine Farla
2012-40 Explaining the dynamics of stagnation: An empirical examination of the North, Wallis and Weingast approach by Richard Bluhm, Denis de Crombrugghe and Adam Szirmai
2012-41 The importance of manufacturing in economic development: Past, present and future perspectives by Wim Naudé and Adam Szirmai
2012-42 Lords of Uhuru: the political economy of elite competition and institutional change in post-independence Kenya by Biniam Bedasso
2012-43 Employment and wages of people living with HIV/AIDS by Pilar García-Gómez, José M. Labeaga and Juan Oliva
2012-44 Prescriptions for network strategy: Does evidence of network effects in cross-section support them? by Joel A.C. Baum, Robin Cowan and Nicolas Jonard
2012-45 Perspectives on human development theory in democracy promotion: A comparison of democracy promotion programmes in Egypt through the lenses of classical and revised modernisation theory by Inger Karin Moen Dyrmes
2012-46 Nonlinearities in productivity growth: A semi-parametric panel analysis by Théophile T. Azomahou, Bity Diene and Mbaye Diene
2012-47 Optimal health investment with separable and non-separable preferences by Théophile T. Azomahou, Bity Diene, Mbaye Diene and Luc Soete
2012-48 Income polarization and innovation: Evidence from African economies by Théophile T. Azomahou and Mbaye Dien
2012-49 Technological capabilities and cost efficiency as antecedents of foreign market entry by Fabrizio Cesaroni, Marco S. Giarratana and Ester Martinez-Ros

2012-50 Does the internet generate economic growth, international trade, or both? by Huub Meijers

2012-51 Process innovation objectives and management complementarities: patterns, drivers, co-adoption and performance effects by Jose-Luis Hervas-Oliver, Francisca Sempere-Ripoll and Carles Boronat-Moll

2012-52 A systemic perspective in understanding the successful emergence of non-traditional exports: two cases from Africa and Latin America by Michiko Iizuka and Mulu Gebreeyesus

2012-53 Determinants of quadic patenting: Market access, imitative threat, competition and strength of intellectual property rights by Can Huang and Jojo Jacob

2012-54 Envy and habits: Panel data estimates of interdependent preferences by Jose Maria Casado, Francisco Alvarez-Cuadrado, Jose Maria Labeaga and Dhanoos Sutthiphisal

2012-55 The impact of Medium-Skilled immigration: A general equilibrium approach by Joan Muysken, Ehsan Vallizadeh and Thomas Ziesemer

2012-56 Tax incentives or subsidies for R&D? by Isabel Busom, Beatriz Corchuelo and Ester Martinez Ros

2012-57 The impact of development aid on education and health: Survey and new evidence from dynamic models by Thomas Ziesemer

2012-58 Do R&D tax incentives lead to higher wages for R&D workers? Evidence from the Netherlands by Boris Lokshin and Pierre Mohnen

2012-59 Determinants of the prevalence of diarrhoea in adolescents attending school: A case study of an Indian village school by Shyama V. Ramani, Timothée Frühauf, Arijita Dutta and Huub Meijers

2012-60 Communication costs and trade in Sub-Saharan Africa by Evans Mupela and Adam Szirmai

2012-61 Differential welfare state impacts for frontier working age families by Irina S. Burlacu and Cathal O'Donoghue

2012-62 Microeconometric evidence of financing frictions and innovative activity by Amaresh K Tiwari, Pierre Mohnen, Franz C. Palm, Sybrand Schim van der Loeff