Internet Appendix

This appendix provides a summary of the missing data problem and discusses several popular econometric approaches to handling missing data that are considered in this paper. With partially observed data, we can rarely be sure of the mechanism leading to such missing data. Therefore, we highlight some approaches to analyzing missing data under different mechanisms, which helps to establish inference robustness in the face of uncertainty about the missingness mechanism. In particular, we consider listwise deletion, deterministic imputation, inverse probability weighting, Heckman selection, and multiple imputation. For exposition simplicity (as in the main body of the paper), we consider the case where only one explanatory variable contains missing observations. Let y_i be the dependent variable and z_i be the explanatory variables with missingness. We have the linear relation:

$$y_i = \alpha + \theta z_i + \varepsilon_i, \quad i = 1, \dots, N. \tag{IA1}$$

Let s_i be a selection indicator where $s_i = 1$ when z_i is not missing and firm i is included in the regression. Otherwise, when $s_i = 0$ firm i is deleted from the data. The validity of solutions to this problem depends on the missing mechanism, thus we first present the three missing mechanisms.

1. Missing completely at random (MCAR):

$$P(s = 0|y, z, x) = P(s = 0).$$

   This means that the missing probability does not depend on any random variables.

2. Missing at random (MAR): The probability of missing can be formulated by:

$$P(s = 0|y, z, x) = P(s = 0|x).$$

   In other words, the probability of missingness only depends on the set of *observed* variables x, but not on the missing variable itself nor on unobservables.

3. Missing not at random (MNAR): the missing mechanism is neither MAR nor MCAR. For example, the missing mechanism depends on the value of z itself, or on unobserved variables, e.g., high-income individuals do not participate in surveys related to income.

Effects of Listwise Deletion

Listwise deletion only uses a subsample of observations, deleting those that contain missing values in the z-variable.[9] This leads to estimating the following regression using the subsample of the data:

$$y_i = s_i\alpha + \theta s_i z_i + s_i\varepsilon_i, \tag{IA2}$$

[9] We consider the univariate setup for simplicity. There might be other covariates of interest that drive the outcome variable but including them in the regression does not change the problem of deletion.

where $s_i z_i$ is now the explanatory variable and $s_i \varepsilon_i$ is the error term. The OLS (ordinary least squares) estimator is unbiased if $E(s_i \varepsilon_i z_i) = 0$, which can be implied by $E(\varepsilon_i | z_i, s_i) = 0$. If MCAR holds and z_i is exogenous, then $E(\varepsilon_i | z_i, s_i) = E(\varepsilon_i | z_i) = 0$. Thus, deletion can lead to consistent estimates in the case of MCAR. However, if selection is driven by observed or even unobserved variables as in MAR and MNAR cases, $E(\varepsilon_i | z_i, s_i) \neq 0$ in general because ε_i can be correlated with s_i even if one controls for z_i, leading to biased estimates produced by deletion.

Deterministic Imputation

Another popular approach used in empirical studies is to impute the missing observations using various methods, and then treat the resulting data as given for further analysis. Frequently used deterministic imputation employs, e.g., zero, overall average, average from "similar" observations, or fitted values based on some pre-specified models. The validity of this method depends on whether the specified imputation models are correct. If the imputation model perfectly coincides with the missing mechanism, then the resulting estimate using the imputed sample is consistent. On the contrary, misspecification of the imputation models can lead to potentially biased estimates because of the distortion of the variance-covariance matrices.

Inverse Probability Weighting

Inverse probability weighting assigns different weights to observed data points depending on their probability of being observed. Thus, the computation of IPW requires researchers to know the probability of being observed. Consider the case of MAR, where the probability of missing (or equivalently being observed) only depends on a set of observed variables x. Denote $p(x) \equiv P(s = 1|x) = P(s = 1|y, x, z)$, then we can solve the missing data problem by:

$$\min_{\alpha,\theta} \sum_{i=1}^{N} \left(\frac{s_i}{p(x_i)}\right) (y_i - \alpha - \theta z_i)^2.$$

In practice, $p(x)$ is often unknown except in some special cases, and thus we need to estimate it. To this end, we can regress the selection indicator s on x using flexible binary choice models, such as logit or probit, or even nonparametric models, and obtain the estimated selection probability (or alternatively called the propensity score) $\hat{p}(x)$.

Heckman's Correction for Selection Bias

We know from (IA2) that the OLS estimator $\hat{\theta}$ is biased because $E(y_i | s_i = 1, z_i) = \alpha_i + \theta z_i + E(\varepsilon_i | z_i, s_i = 1)$, and $E(\varepsilon_i | z_i, s_i = 1) \neq 0$ in general. Heckman's method assumes that the missing mechanism is determined by the following model:

$$s_i^* = \beta x_i + \eta_i, \quad i = 1, \ldots, N, \tag{IA3}$$

where s_i^* is the latent variable associated with s_i, i.e., $s_i = 1$ if $s_i^* > 0$ and $s_i = 0$ if $s_i^* \leq 0$. Further, assume that the error terms in (IA3) is normally distributed with variance σ_η^2 and

correlated with ε_i in (IA1), and their covariance is ρ; x and z are both exogeneous. The Heckman selection procedure approximates the "omitted variable" $(\varepsilon_i|z_i, s_i = 1)$ by its consistent estimate and includes this proxy in the regression to correct for the bias. In particular, based on the joint distribution of η_i and ε_i, one could write $E(\varepsilon_i|z_i, s_i = 1) = \sigma_\eta \rho \lambda(x_i\beta) = \gamma\lambda(x_i\beta)$, where $\lambda(x_i\beta)$ is the inverse Mills ratio defined by:

$$\lambda(x_i\beta) = \frac{\phi(-x_i\beta)}{1-\Phi(-x_i\beta)} = \frac{\phi(x_i\beta)}{\Phi(x_i\beta)}.$$

Then we can rewrite the conditional expectation of y_i given x_i and selection into the sample as:

$$E(y_i|x_i, s_i = 1) = \alpha + \theta z_i + \gamma\lambda(x_i\beta).$$

This leads to Heckman's two-step procedure.

Step 1: Estimate a probit regression $\mathrm{P}(s_i = 1|x_i) = \Phi(x_i\beta)$ using all N observations and obtain the estimate $\hat{\beta}$. Then compute the inverse Mills ratio $\lambda(x_i\hat{\beta})$.
Step 2: Estimate the regression $y_i = \alpha + \theta z_i + \gamma\lambda(x_i\hat{\beta})$ using OLS.

The estimates $\hat{\alpha}$, $\hat{\theta}$, and $\hat{\gamma}$ are consistent when x correctly includes all of the selection variables. The validity of Heckman's procedure also heavily relies on the distributional assumptions of the two errors, η_i and ε_i. For example, the deviation from the normality assumption of η_i may negatively affect the performance of the Heckman's procedure. Since γ captures the covariance between η_i and ε_i and a nonzero correlation implies selection bias, we can test whether selection is exogenous (or equivalently MCAR) by testing whether $\hat{\gamma} = 0$. For more extensions of Heckman's procedure, see Wooldridge (2002, Chapter 17).

Multiple Imputation

Multiple imputation (MI) is essentially an iterative version of stochastic imputation, which aims at explicitly modeling the uncertainty/variability ignored by the deterministic imputation procedures. Instead of imputing in a single value, multiple imputation uses the (joint) distribution of the observed data to estimate the parameters of interest multiple times to capture the uncertainty/variability in this imputation procedure. A general multiple imputation procedure consists of three steps:

Step 1. Imputation: Impute the missing data with their estimates and create a complete sample. Repeat this process multiple times.

Step 2. Estimation: For each complete sample, estimate the parameters of interest.

Step 3. Pooling: Combine the parameter estimates obtained from each completed data set.

The imputation method should be chosen depending on the type of variables with missing observations and the pattern of missingness. For example, MI with multivariate normal regressions can be applied to impute one or more continuous variables of arbitrary missing-value patterns; MI with chained equations employs a separate conditional distribution for each imputed variable and

is often used to impute a variable with finite and discrete support (e.g., binary, multinomial, or count variable). We illustrate the MI with multivariate normal regressions (MI_MVN). As all MI methods, MI with multivariate normal regressions analyses the data in three steps: imputation, estimation, and pooling. We discuss the three steps in turn.

First, MI_MVN imputes the missing observations using data augmentation. In this case, we assume that the variable containing missing observations z is related with a set of (completely) observed variables x by:

$$z_i = \delta' x_i + v_i, \quad i = 1, \dots, N,$$

where $v_i \sim N(0, \sigma_v^2)$. Denote $w_i = (z_i, x_i)$. Data augmentation in this case is essentially an iterative Markov chain Monte Carlo (MCMC) procedure that iterates between two (sub-)steps, a replacement step and posterior step.

- Replacement step: We replace the missing values of z_i with draws from the conditional posterior distribution of z_i given observed variables and the values of model parameters in this iteration. Particularly, for each iteration t, we can replace the missing observations by:

$$z_i^{(t)} \sim P\left(z_i \middle| x_i, \delta^{(t-1)}, \sigma_v^{(t-1)}\right), \quad \text{for} \quad i \in \{i | s_i = 1\}.$$

- Posterior step: We draw the new values of model parameters from their conditional posterior distribution given the observed data and imputed data from the previous replacement step.

$$\sigma_v^{(t)} \sim P\left(\sigma_v \middle| x_i, z_i^{(t)}\right), \text{ and } \quad \delta^{(t)} \sim P\left(\delta \middle| x_i, z_i^{(t)}, \sigma_v^{(t)}\right),$$

  where $z_i^{(t)}$ is the imputed value from iteration t if it is missing and the original value if non-missing.

The conditional posterior distributions above are jointly determined from the prior distribution for the model parameter $P(\delta, \sigma_v)$, e.g., uniform, Jeffreys, or ridge, and the assumed normal distribution of the data. These two steps (replacement and posterior) are iterated until a specified number of iterations or there is numerical convergence.

Second, we estimate the regression of interest (IA1) with the imputed (pseudo-complete) data set using various approaches, e.g., OLS, HS. Since the imputation is conducted for multiple times, D times, we obtain multiple estimates for the same regression parameter θ.

Third, we combine/pool the estimates (coefficients and standard errors) across all imputed datasets and obtain a single statistic for each parameter. The final estimated slope coefficient $\hat{\theta}$ is simply an arithmetic mean of the corresponding estimate obtained from each of the imputed data. The variance of $\hat{\theta}$ is obtained by the total variance formula and is written by the average estimated variance of coefficient estimates across D imputed datasets plus the sample variance of coefficient estimates based on D imputations.

A major advantage of multiple imputation over deterministic imputation is that the final statistics appropriately reflect the uncertainty caused by imputation. If the joint normality is a reasonable assumption and the specification of x is correct (i.e., MAR), MI_MVN produces consistent estimates. In practice, a safe strategy is to include all observables in x to better approximate the posterior distribution.

Table IA1
Predictability of Unreported Innovation with Lasso

The table presents the OLS regression results for predictability of unreported innovation using only Lasso variables. Columns (1)-(4) *World* present the results for all countries in the sample. Columns (5)-(7) present the results for the *US* only. Panel A presents the results for unreported R&D, and Panel B presents the results for non-USPTO patent seeking firms. Standard errors are double clustered at firm and time level. T-statistics are presented in brackets. Variable definitions are presented in Table A1. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. Adj. R^2 is the adjusted R^2.

Panel A. Unreported R&D

	World				US		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln(Total Assets)	-0.020***	-0.017***	-0.012***	-0.009***	0.040***	0.004*	-0.014***
	(-10.20)	(-8.31)	(-7.86)	(-3.54)	(13.28)	(1.73)	(-4.40)
Stock Liquidity	-0.007***	-0.007***	-0.005***	-0.001***	-0.008***	-0.003***	-0.001***
	(-8.92)	(-9.99)	(-14.33)	(-3.47)	(-13.50)	(-6.23)	(-3.27)
Patent Intensity	-604.300***	-1.310	13.370	38.122**	-700.176***	-19.498***	-17.045*
	(-17.42)	(-0.11)	(1.06)	(2.02)	(-21.33)	(-3.08)	(-1.77)
Country FE			Yes				
Industry FE		Yes	Yes			Yes	
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE				Yes			Yes
N	300,634	300,634	300,634	328,734.0	77,982	77,982	76,944
Adj. R^2	0.13	0.23	0.38	0.80	0.23	0.53	0.93

Panel B. Non-USPTO Patent Seeking Firms

Ln(Total Assets)	-0.012***	-0.009***	-0.020***	-0.006***	-0.000	-0.023***	-0.013***
	(-11.62)	(-9.41)	(-19.28)	(-6.88)	(-0.08)	(-9.67)	(-4.03)
Stock Liquidity	-0.006***	-0.007***	-0.004***	-0.001***	-0.007***	-0.005***	-0.001***
	(-20.77)	(-21.86)	(-17.74)	(-5.11)	(-14.73)	(-12.24)	(-3.55)
Patent Intensity	-0.000**	-0.000**	-0.000**	-0.000	-0.000***	-0.000**	0.000
	(-2.09)	(-2.10)	(-2.22)	(-1.14)	(-3.22)	(-2.60)	(1.01)
R&D Stock	-368.647***	-30.191**	-25.181***	-4.871	-554.522***	-47.131***	-0.239
	(-17.34)	(-2.57)	(-3.23)	(-0.90)	(-18.37)	(-3.72)	(-0.02)
Country FE			Yes				
Industry FE		Yes	Yes			Yes	
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE				Yes			Yes
N	327,997	327,997	327,997	326,067	77,958	77,958	76,926
R^2	0.09	0.15	0.24	0.76	0.15	0.32	0.78

.

Table IA2
Simulation Based on the Empirical Distribution from Compustat Data

This table provides the evaluation statistics, bias (relative bias over true parameter) and root mean squared error (RMSE) for the simulation based on the empirical distribution from Compustat (US) data, as described in section 6.1 of the paper. Bias presents the average of the absolute bias across all five variables and RMSE presents the average RMSE across the five variables. The empirical distribution is from the panel of 783 firms with non-missing information for all variables except R&D. The methods evaluated are listwise deletion (LD), imputation with zero (ImpZero), imputation with industry mean, two-digit SIC code (ImpMean), inverse probability weighting (IPW), Heckman selection procedure (HS), and multiple imputation (MI). The regressions for imputation with zero and industry mean include a dummy variable for the imputed observations. MI is spec uses all the variables in the regression and is estimated using MCMC with 200 iterations for convergence. We present results for three missingness mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Panel A presents the results for the missingness regression which includes the lasso variables. Panel B presents the results with the MI specification in Panel A as well as includes the Lasso variables in the Sales growth regression. Panel C presents the Double Lasso results. Variable definitions are presented in Table A1. We generate missingness R&D for 50 and 70% of the sample. We conduct 500 simulations.

		Missing 70%						Missing 50%					
		LD	Imp. Zero	Imp. Mean	IPW	HS	MI	LD	Imp. Zero	Imp. Mean	IPW	HS	MI
Panel A. Missingess Regression with Q, A, V, and PI													
MCAR	Bias	0.80	0.24	0.22	3.69	3.67	0.11	0.63	0.16	0.13	3.42	3.36	0.05
	RMSE	0.05	0.02	0.02	0.09	0.09	0.02	0.03	0.02	0.02	0.07	0.07	0.02
MAR	Bias	1.02	0.13	0.12	20.22	118.27	0.10	0.63	0.17	0.15	17.43	100.29	0.07
	RMSE	0.07	0.02	0.02	0.53	2.94	0.02	0.04	0.02	0.02	0.45	2.87	0.02
MNAR	Bias	0.98	0.13	0.12	11.96	67.10	0.11	0.58	0.18	0.15	10.18	59.42	0.08
	RMSE	0.07	0.02	0.02	0.26	2.27	0.02	0.03	0.02	0.02	0.25	2.05	0.02
Panel B. Missingess Regression with Q, A, V, and PI and Sales growth regression with V and PI													
MCAR	Bias	0.86	0.26	0.24	3.74	3.75	0.17	0.60	0.23	0.15	3.52	3.48	0.08
	RMSE	0.05	0.02	0.02	0.09	0.09	0.02	0.04	0.02	0.02	0.07	0.08	0.02
MAR	Bias	1.10	0.25	0.23	18.96	100.69	0.09	0.61	0.18	0.17	16.26	98.45	0.06
	RMSE	0.08	0.02	0.02	0.48	2.73	0.02	0.04	0.02	0.02	0.39	2.50	0.02
MNAR	Bias	0.99	0.13	0.11	11.49	61.56	0.14	0.60	0.17	0.14	9.93	52.77	0.05
	RMSE	0.07	0.02	0.02	0.29	1.86	0.02	0.03	0.02	0.02	0.24	1.76	0.02

		Missing 70%						Missing 50%					
		LD	Imp. Zero	Imp. Mean	IPW	HS	MI	LD	Imp. Zero	Imp. Mean	IPW	HS	MI
						Panel C. Double Lasso							
MCAR	Bias	0.77	0.24	0.22	3.77	3.74	0.08	0.60	0.21	0.19	3.59	3.47	0.09
	RMSE	0.05	0.02	0.02	0.09	0.09	0.02	0.04	0.02	0.02	0.07	0.07	0.02
MAR	Bias	0.66	0.18	0.15	15.51	5.76	0.05	0.48	0.19	0.17	16.46	4.50	0.09
	RMSE	0.03	0.02	0.02	0.45	0.40	0.02	0.02	0.02	0.02	0.43	0.35	0.02
MNAR	Bias	0.63	0.24	0.20	10.00	3.58	0.07	0.49	0.20	0.18	10.08	3.34	0.06
	RMSE	0.03	0.02	0.02	0.29	0.24	0.02	0.02	0.02	0.02	0.25	0.18	0.02

Table IA3
Simulation Based on Simulated Data

This table provides the evaluation statistics, bias (relative bias over true parameter) and root mean squared error (RMSE) for the simulation based on simulated data, as described in section 6.3 of the paper. The methods evaluated are listwise deletion (LD), imputation with zero (ImpZero), imputation with industry mean, two-digit SIC code (ImpMean), inverse probability weighting (IPW), Heckman selection (HS), and multiple imputation (MI). MI is estimated using MCMC with 200 iterations for convergence. The regressions for imputation with zero and industry mean include a dummy variable for the imputed observations. We present results for three missingness mechanisms: missing completely at random (MCAR) in Panel A, missing at random (MAR) in Panel B, and missing not at random (MNAR) in Panel C. We generate missingness in x1 for 50 and 70% of the sample. We conduct 500 simulations.

		Missing 70%						Missing 50%					
		LD	Imp Zero	Imp Mean	IPW	HS	MI	LD	Imp Zero	Imp Mean	IPW	HS	MI
						Panel A. MCAR							
Bias	θ_1	0.00	-0.19	-0.19	0.00	0.00	-0.01	0.00	-0.13	-0.13	0.00	0.00	-0.01
	θ_2	0.01	0.28	0.28	0.01	0.01	0.01	0.00	0.19	0.19	0.00	0.00	0.00
RMSE	θ_1	0.11	0.21	0.21	0.08	0.11	0.09	0.06	0.07	0.07	0.06	0.06	0.05
	θ_2	0.11	0.29	0.29	0.08	0.11	0.09	0.06	0.10	0.10	0.07	0.06	0.05
						Panel B. MAR							
Bias	θ_1	-0.15	-0.23	-0.23	-0.16	-0.08	-0.08	-0.11	-0.16	-0.16	-0.11	-0.09	-0.05
	θ_2	-0.12	0.12	0.12	-0.12	-0.08	-0.05	-0.08	0.04	0.04	-0.07	-0.06	-0.04
RMSE	θ_1	0.17	0.24	0.24	0.18	0.17	0.10	0.13	0.17	0.17	0.13	0.12	0.08
	θ_2	0.15	0.13	0.13	0.15	0.16	0.09	0.07	0.06	0.06	0.07	0.07	0.05
						Panel C. MNAR							
Bias	θ_1	-0.17	-0.28	-0.28	-0.19	-0.13	-0.10	-0.13	-0.19	-0.19	-0.13	-0.11	-0.05
	θ_2	-0.16	0.14	0.14	-0.15	-0.13	-0.08	-0.11	0.04	0.04	-0.11	-0.10	-0.07
RMSE	θ_1	0.19	0.29	0.29	0.20	0.17	0.12	0.14	0.20	0.20	0.14	0.13	0.07
	θ_2	0.17	0.12	0.12	0.17	0.16	0.10	0.13	0.06	0.06	0.13	0.12	0.08

Table IA4
Imputation Effect on Empirical Inference

This table replicates the results in Fama and French (2002) using different imputation methods and two-way fixed effects. We present the results of a contemporaneous regression with two-way fixed effects: $\frac{L_t}{A_t} = \beta_0 + \beta_1 \frac{V_t}{A_t} + \beta_2 \frac{ET_t}{A_t} + \beta_3 \frac{Dp_t}{A_t} + \beta_4 RDD_t + \beta_5 \frac{RD_t}{A_t} + \beta_6 \ln(A_t) + e_t$. "ImpZero" presents the result for the sample with imputation with zero and an indicator variable, "LD" presents the results for listwise deletion, "MI" presents the results for multiple imputation implemented using all the variables in the regression in the imputation, "MI Lasso" presents the results for multiple imputation implemented using all the variables in the regression and the Lasso variables stock liquidity and industry patent intensity in the imputation, "Pseudo RD" presents the result using pseudo R&D as an explanatory variable, and "Text-based Innov." presents the results for the analyst coverage based innovation variable (Bellstam et al., 2020). The dependent variable is book leverage $\frac{L_t}{A_t}$ at time T. $\frac{V_t}{A_t}$ is the market to book ratio, $\frac{ET_t}{A_t}$ is earnings before interest and taxes as a proportion of total assets, $\frac{Dp_t}{A_t}$ is depreciation as a proportion of total assets, $\frac{RD_t}{A_t}$ is the R&D expenses as a proportion of total assets, RDD_t is an indicator variable equal to 1 if R&D expenditure is missing and has been imputed with zero, and zero otherwise, *Pseudo R&D*$_t$ is an indicator variable equal to 1 if a firm applies for a patent in PATSTAT and has no reported R&D, and zero otherwise, *Text-based Innov.*$_t$ is the firm analyst-based innovation measure from (Bellstam et al., 2020), and $\ln(A_t)$ is the natural logarithm of total assets. Non-dividend payers include firms that do not pay dividend in year *T-1*. Panel A presents the results for the dividend paying firms and Panel B for the non-dividend paying firms. The sample period is 1965-1999. Standard errors are double clustered.

Panel A. Dividend Payer Firms

Variable	Imp Zero (1)	LD (2)	MI (3)	MI Lasso (4)	Pseudo R&D (5)	Text-based Innovation (6)
Intercept	0.305***	0.344***	0.366***	0.368***	0.300***	0.246***
	(22.62)	(19.83)	(56.52)	(55.24)	(22.13)	(3.94)
$\frac{V_t}{A_t}$	-0.001	-0.001	0.001	0.001	0.000	-0.006
	(-0.15)	(-0.47)	(0.40)	(0.60)	(-0.10)	(-1.29)
$\frac{ET_t}{A_t}$	-0.158**	-0.215**	-0.184**	-0.192	-0.157	-0.628
	(-1.99)	(-2.66)	(-2.07)	(-2.17)	(-1.99)	(-3.91)
$\frac{Dp_t}{A_t}$	-1.076***	-0.059	-1.057***	-1.048***	-1.049***	-0.797***
	(-6.12)	(-0.30)	(-10.67)	(-10.56)	(-6.05)	(-2.77)
RDD$_t$	0.070***				0.075***	
	(11.96)				(12.35)	
$\frac{RD_t}{A_t}$	-0.290***	-0.435***	0.081***	0.033	-0.290***	
	(-2.71)	(-4.54)	(3.91)	(1.48)	(-2.72)	
Pseudo R&D$_t$					-0.098***	
					(-12.43)	
Text-based Innovation$_t$						-0.009*
						(-1.78)
$\ln(A_t)$	0.041***	0.029***	0.038***	0.038***	0.042***	0.048***
	(29.95)	(13.61)	(96.36)	(95.10)	(30.14)	(6.89)

Panel B. Non-dividend Payer Firms

Variable	Imp Zero	LD	MI	MI Lasso	Pseudo R&D	Text-based Innovation
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.325***	0.394***	0.376***	0.381***	0.323***	0.242
	(4.70)	(20.07)	(6.74)	(7.05)	(4.66)	(1.11)
$\frac{V_t}{A_t}$	0.027	-0.004***	0.028**	0.029***	0.027	-0.008
	(1.32)	(-3.24)	(2.24)	(2.30)	(1.32)	-1.40
$\frac{ET_t}{A_t}$	-0.517***	-0.301***	-0.139	-0.136	-0.517***	-0.404
	(-3.15)	(-4.77)	(-0.54)	(-0.52)	(-3.14)	-1.56
$\frac{Dp_t}{A_t}$	0.691	1.984***	0.636*	0.651*	0.692	1.725*
	(1.29)	(7.96)	(1.66)	(1.70)	(1.29)	1.84
RDD	0.079***				0.082***	
	(4.61)				(4.73)	
$\frac{RD_t}{A_t}$	-0.702***	-0.335***	0.955***	0.962***	-0.701***	
	(-2.83)	(-3.33)	(3.45)	(3.43)	(-2.83)	
Pseudo R&D					-0.134***	
					(-6.03)	
Text-based Innovation						-0.095***
						(-5.50)
$\ln(A_t)$	0.032***	0.013**	0.022***	0.024***	0.033***	0.042
	(4.54)	(2.60)	(4.98)	(5.35)	(4.62)	(1.60)