Innovation and institutional ownership revisited: an empirical investigation with count data models

Susanne Berger[1] · Herbert Stocker[2] ·
Achim Zeileis[1] (iD)

Abstract By discriminating between a lazy manager and a career concerns hypothesis, Aghion et al. (Am Econ Rev 103(1):277–304, 2013. doi:10.1257/aer.103.1.277) try to disentangle the link between innovation and institutional ownership. Citation-weighted patent counts are used as a proxy for innovation, which motivates the use of count data models. A replication in a narrow sense confirms their empirical results which are mainly based on Poisson models (i.e., with a single set of regression coefficients). However, when extending the model framework by count data hurdle models, it is shown that the two hurdle parts do not coincide—as they should under the Poisson model—but lead to different results. Nevertheless, a remarkably stable positive correlation of citation-weighted patents and institutional ownership across all model specifications can be shown.

Keywords Innovation · Institutional ownership · Count data · Hurdle model · Replication

JEL Classification C25 · G32 · O31

1 Introduction

What sets the wheels of innovation in motion? And what keeps them running? With a focus on the impact of the owners of publicly listed companies on innovation, Aghion

✉ Susanne Berger
susanne.berger@uibk.ac.at

Achim Zeileis
Achim.Zeileis@R-project.org

[1] Department of Statistics, University of Innsbruck, Universitätsstr. 15, 6020 Innsbruck, Austria

[2] Department of Economics, University of Innsbruck, Universitätsstr. 15, 6020 Innsbruck, Austria

et al. (2013) attract attention to institutional ownership that comprises '…all investors in financial markets which are neither private households nor public institutions' (Menkhoff 2002, p. 907).

Aghion et al. (2013) show empirically that institutional ownership is positively correlated with innovation. Additionally, the authors aim to disentangle the link between innovation, institutional ownership, and product market competition. This is achieved by an empirical discrimination between a lazy manager and a career concerns hypothesis. As the dependent variable innovation is mapped out through the use of future citation-weighted patent counts, the methodological approach employed by Aghion et al. (2013) is based on count data models. The authors estimate a Poisson model for the mean equation and employ standard errors clustered at either the firm level or a three-digit industry level.

Replication in a narrow sense—i.e., using the same data and model specification, but a different software package—largely confirms the empirical results obtained by Aghion et al. (2013). The software in use for the replication as well as for an extended analysis is R (R Core Team 2013), instead of Stata (StataCorp. 2011), which is used by Aghion et al. (2013). Only in the negative binomial specification, some differences had to be resolved regarding standard errors and starting values.

We argue that the relevance of factors that contribute to the occurrence of innovative outcome differs between two distinct processes. Two aspects are of main interest in this context: the determinants that are necessary for a firm to be able to innovate and the factors that keep an already innovating firm carrying out further innovations. Thus, we go one step further and extend the analysis of Aghion et al. (2013) with hurdle models, which are useful to shed some light on potentially existing structural differences. Hurdle models are two-part models with a binary part that models the decision to innovate at all, and a count part that models ongoing innovation, respectively. This approach can be subsumed as a 'replication in a wider sense'—i.e., critically assessing the analysis by altering the econometric model—we present a robustness check that reveals on the one hand a stable positive correlation of innovation with institutional ownership, but on the other hand, the results are inconclusive regarding the career concerns and the lazy manager hypothesis. Indeed, there is some indication for a departure from the single-equation Poisson model, both from a statistical as well as from an economic perspective, as the Poisson model implicitly assumes one underlying process, for which the factors that drive innovation are the same. However, it might be the case that the first innovation is more difficult to achieve, especially compared to succeeding innovations. From a statistical perspective, both hurdle parts should produce similar parameter estimates to those obtained by a Poisson specification if the assumption of one underlying process is correct. Due to differing parameter estimates, there is some indication that the innovation process is not adequately reflected by a single-equation Poisson model. Thus, our contribution is the use of a more flexible modeling approach achieved through the utilization of count data hurdle models, instead of Poisson models. We argue that two-part hurdle models are an interesting alternative that can provide us with new insights into the determinants that trigger innovative outcome. Moreover, altering the econometric model from Poisson to the more general hurdle model enables us to uncover the instability of effects which Aghion et al. (2013) place great emphasis on.

In line with Aghion et al. (2013), most studies in the empirical literature that consider the relationship between (institutional) ownership structure and innovation (or R&D spendings) find a positive correlation.[1] To name a few examples, Baysinger et al. (1991) find a positive effect of institutional investors on R&D spendings (see also Hansen and Hill 1991). Kochhar and David (1996) state that institutions are long-term oriented, and thus have a positive influence on firm innovation. Bushee (1998) finds that high institutional ownership is accompanied by a reluctance from managers to cut R&D after a decline in earnings. David et al. (2001) state that it is rather institutional investor activism that has a positive influence on R&D input. Summing up, a variety of components appear in the recent literature that can explain a positive effect of institutional investors on innovation. However, in this context the discussion only rarely addresses to what extent hurdle models can lead to new insights into the emergence of innovation. Nevertheless, some scholars have stated that there are two phases of the innovation process that are subject to different constraints (König and Licht 1995; Licht and Zoz 1996; Crepon and Duguet 1997; Brouwer and Kleinknecht 1999; Mohnen and Röller 2005; Mäkinen 2007). Moreover, Xu et al. (2015) find that the influence of institutional owners is not the same for these differing innovation processes, thus justifying the use of hurdle models. Another part of the literature has found that the propensity to patent—the fraction of innovations for which at least one patent application is filed—differs across industries, firms, and kinds of innovations (Comanor and Scherer 1969; Basberg 1987; Griliches 1990; Archibugi and Planta 1996; Kleinknecht et al. 2002; Van Der Panne and Kleinknecht 2005). To model these nonlinearities in the innovation process, hurdle models are, though not very frequently, applied in the literature (König and Licht 1995; Licht and Zoz 1996; Crepon and Duguet 1997; Brouwer and Kleinknecht 1999). For example, Licht and Zoz (1996) state that the decision to apply for a first patent and the decision to apply for succeeding patents are ruled by different processes. The protection of intellectual property rights and how to deal with patentable inventions are basic issues that require reflection of the firms, especially when the yield of holding a patent is unknown. While this basic decision is often based on the context of the first invention, the decision to apply for patents for further inventions rests on this first principle decision (Licht and Zoz 1996).

The rest of the paper is structured as follows. Section 2 comprises the replication of the basic results obtained by Aghion et al. (2013) as well as an extension of their analysis by means of hurdle models, and Sect. 3 contains the replication and extension of models with competition included as a further explanatory variable. Section 4 concludes.

2 Basic models: innovation and institutional ownership

2.1 Replication (in a narrow sense)

Future citation-weighted patent counts are used as a proxy for the dependent variable innovation, which motivates the use of count data models. As a starting point (and

---

[1] An exception that 'confirms' the rule is Graves (1988), who finds a significant negative relationship between institutional ownership and R&D expenditures in the computer industry.

following Aghion et al. 2013), a log link is implemented. In the spirit of a quasi-maximum likelihood approach, the Poisson model is employed for the mean equation along with clustered standard errors. The conditional expectation function is

$$\mathrm{E}(\mathrm{Cites}_{it} \mid \boldsymbol{x}_{it}, \gamma_i, \delta_t) = \exp\left(\boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta} + \gamma_i + \delta_t\right), \tag{1}$$

where Cites_{it} is the number of citation-weighted patent counts for company i in year t, the vector $\boldsymbol{x}_{it}$ contains all explanatory variables for firm i in year t in the model, γ_i is firm-specific fixed effects controls, and δ_t is time dummy variable (Aghion et al. 2013, pp. 280–281). All models include institutional ownership, which is measured as the percentage of outstanding shares held by institutions, the capital to labor ratio as well as sales (in logs), time dummies and four-digit industry dummies as explanatory variables (Aghion et al. 2013). Some models additionally include the stock of R&D expenditures (in logs) and the presample mean-scaling estimator[2] developed by Blundell and Powell (2004).

Table 1 comprises the (successful) replication of Table 1 in Aghion et al. (2013, p. 283) , where the outcomes show that the coefficient of institutional ownership is consistently positive and significant.

In the negative binomial specification, some additional efforts are necessary to resolve differences between Stata (StataCorp. 2011) and R (R Core Team 2013) regarding starting values and standard errors. As for the former, the function glm.nb() (Venables and Ripley 2002), used for negative binomial regression in R, sometimes has problems to find starting values for very high-dimensional regressions. Hence, we estimate the negative binomial parameter estimates using a quasi-Newton optimization based on analytical gradients and a numerical approximation of the Hessian starting from Poisson coefficient estimates. Concerning the clustered sandwich standard errors, Stata and R give different results by default. The root of this difference is that the standard errors in Stata are based on the observed information matrix while R employs the expected information by default. When employing the numerical approximation of the observed information in R, Stata's results can be replicated except for small numeric differences that are likely due to the great amount of industry dummies as well as the model specification.

2.2 Extended analysis

However, the data show overdispersion as well as excess zeros. There are about 35.2 % (accounting for 2183 out of 6208) firm–year observations with zero citation-weighted patents in the data. On one hand, the zeros can come from either the decision to keep potentially patentable discoveries in secrecy, or from the lack of any patentable finding (see Crepon and Duguet 1997), both of which resulting in a lack of any patents. On the other hand, the zeros can come from holding patents, but without citations. In the data analyzed, there are about 32.6 % firm–year observations with zero patents in the

---

[2] To account for firm-specific fixed effects, the presample average of citation-weighted patents is included into the model (Aghion et al. 2013, p. 281).

Table 1 Basic models—replication

| Model | OLS | OLS | Poisson | Poisson | Poisson | NegBin | NegBin | NegBin |
Dependent variable	ln(Cites)	ln(Cites)	Cites	Cites	Cites	Cites	Cites	Cites
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Institutional owners	0.006**	0.005**	0.010***	0.008***	0.007**	0.009***	0.008***	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ln(capital/labor)	0.433***	0.261**	0.483***	0.346*	0.440***	0.613***	0.343***	0.264***
	(0.094)	(0.085)	(0.136)	(0.165)	(0.132)	(0.106)	(0.087)	(0.076)
ln(sales)	0.568***	0.310***	0.820***	0.349**	0.184**	0.493***	0.229***	0.127***
	(0.037)	(0.045)	(0.042)	(0.117)	(0.063)	(0.047)	(0.058)	(0.037)
ln(R&D) stock		0.337***		0.493***	0.009		0.448***	0.178***
		(0.040)		(0.140)	(0.107)		(0.039)	(0.029)
Fixed effects controls	No	No	No	No	Yes	No	No	Yes
Observations	4025	4025	6208	6208	6208	6208	6208	6208
AIC			744,624.834	667,665.617	446,655.330	48,159.657	47,345.215	46,058.249
BIC			745,614.672	668,668.922	447,672.102	49,156.229	48,355.254	47,081.756

This table is a replication of Table 1 in Aghion et al. (2013). Standard errors (in parentheses) are clustered at firm level. There are 803 different firms. Additionally, all regressions include time dummies for each year (with reference category 1991) as well as four-digit industry dummies. Fixed effects controls are included using the presample mean-scaling estimator developed by Blundell and Powell (2004). Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

data and 2.6 % with patents but without citations. In other words, there are only a few patents in our data which do not get cited. Summarizing the above, the amount of zeros in the dependent variable is higher than expected by the Poisson distribution, which casts doubt on the distributional assumption and suggests potentially different determinants driving the zero and nonzero citations.

Furthermore, overdispersion is a common characteristic of count data (in the field of economics), meaning that the conditional variance is higher than the conditional mean. The fraction $\frac{\text{Var(Cites)}}{\text{Mean(Cites)}} = 4836.2$ reveals a substantial amount of overdispersion for citation-weighted patents (note that covariates are not taken into account here). A negative binomial model offers some remedy in such a situation (see, e.g., Hausman et al. 1984). As a likelihood model, it does explicitly account for dispersion. Aghion et al. (2013) consider negative binomial models only in their basic models in Table 1 (columns 6, 7, and 8). However, it is worth pointing out that the negative binomial model does also not explain the high proportion of zero citations discussed above.

Finally, the Poisson model assumes independent occurrences over time (see, e.g., Cameron and Trivedi 1998) and it may also be the case that the first innovation (the first citation-weighted patent count) is especially hard to obtain in comparison with succeeding innovations, such that '…the innovation process is characterized by non-linearities' (Crepon and Duguet 1997, p. 360). For example, in case of the discovery of a seminal innovation, some further discoveries of minor importance can follow more easily (Crepon and Duguet 1997).

Our investigation is not so much motivated by the huge amount of zeros or overdispersion in the data, but rather uses hurdle models to allow for potentially different processes driving innovation. Employing the Poisson model for the mean equation along with clustered standard errors cannot accommodate differing processes.

In summary, these considerations concerning excess zeros, overdispersion, and potentially different determinants in the innovation process can be addressed by two-part hurdle models. Specifically, these consider the case that there are two different processes driving either the 'first innovation' (does a company own at least one citation-weighted patent) or the 'continuing innovation' decision (if a company has a positive number of citation-weighted patents, how many of them does it possess).[3]

Positive outcomes are observed whether the zero hurdle is crossed and are modeled through a truncated (from the left) negative binomial model, whereas the probability to cross the hurdle is modeled via a censored negative binomial model (see, e.g., Cameron and Trivedi 1998). As we use a negative binomial model for the count as well as for the binary part of the hurdle model, we restrict both dispersion parameters to the same value (as recommended by Winkelmann 2010, p. 183).

Table 2 shows the same models as Table 1, but using negative binomial hurdle models instead of the (single-equation) Poisson models. As in the original analysis, standard errors are clustered at the firm level.

---

[3] Strictly speaking, firms first decide whether or not to patent. However, we use citation-weighted patents per year as the dependent variable in both parts of the hurdle model, as the differences in the results for citations versus patents per year are small. Also, one could consider a different aggregation period for the patents/citations but certainly the technological challenges for new first innovations have been changing over the course of the decade.

Table 2 Basic—hurdle models

Model Dependent variable	Hurdle NegBin (1) Cites		Hurdle NegBin (2) Cites		Hurdle NegBin (3) Cites	
	Count	Zero	Count	Zero	Count	Zero
Institutional owners	0.006**	0.013***	0.005**	0.012***	0.003·	0.009***
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)
ln(capital/labor)	0.549***	0.505***	0.318***	0.289*	0.253***	0.248*
	(0.099)	(0.143)	(0.084)	(0.116)	(0.070)	(0.097)
ln(sales)	0.440***	0.659***	0.199***	0.357***	0.118***	0.191***
	(0.048)	(0.064)	(0.058)	(0.056)	(0.036)	(0.047)
ln(R&D) stock			0.407***	0.339***	0.155***	0.109***
			(0.042)	(0.041)	(0.031)	(0.033)
Fixed effects controls	No		No		Yes	
Observations	6208		6208		6208	
AIC	47,486.20		46,582.55		44,965.53	
BIC	49,472.61		48,595.89		47,005.81	
Hurdle test (p value)	0.000		0.007		0.001	

Standard errors (in parentheses) are clustered by firm. There are 803 different firms. Additionally, all regressions include time dummies for each year (with reference category 1991) as well as four-digit industry dummies. Fixed effects controls are included using the presample mean-scaling estimator developed by Blundell and Powell (2004). The zero part is a censored negative binomial with log link. The count part is a truncated negative binomial with log link. The hurdle tests are based on all coefficients of the respective model given in the table. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '·' 0.1 ' ' 1

Furthermore, a test for the presence of a zero hurdle is conducted. It is a Wald test to check for pairwise equality between all reported coefficients from the two parts of the hurdle model, where the null hypothesis claims that no hurdle is needed (Zeileis et al. 2008). In the event of two identical processes, the estimated parameters in both model parts should be similar as well. On the contrary, we find evidence that our estimated parameters differ in both model parts, which allows us to draw the conclusion that a single equation is not enough.[4]

The results indicate that the coefficient of institutional ownership is for most of the estimated hurdle models positive and significant. The only exception is the count part of the Hurdle NegBin (3) model, where the coefficient of institutional ownership is no longer significant. Quantitatively, a dampening of the coefficient of institutional ownership in the count part of the model can be observed, and the more variables are included into the model. In the binary part of the model, the coefficient of institutional ownership is quite stable. For most of the other explanatory variables, a dampening of the coefficients in both parts of the model can be observed and the more variables are included into the model.

3 Competition models: career concerns versus lazy manager

Two hypotheses are considered by Aghion et al. (2013), the lazy manager hypothesis and the career concerns hypothesis. Both hypotheses predict a positive relationship between institutional ownership and innovation, but they differ with regard to product market competition.

The lazy manager approach is based on 'quiet life' models developed by Bertrand and Mullainathan (2003). It is assumed that managers prefer a quiet life, and want to keep the status quo, as '…he [the average manager] seems to avoid creating new plants as much as he avoids destroying old ones' (Bertrand and Mullainathan 2003, p. 1072). If competition is high, there is no need for intensified monitoring, because 'competiton is a source of discipline.' (Hart 1983, p. 366). Hence, the lazy manager approach implies that institutions and competition are substitutes (Aghion et al. 2013, p. 292).

The career concerns hypothesis pays closer attention to the managerial labor market. In a seminal paper, Fama (1980) investigates how pressure imposed by managerial labor markets can discipline managers. As a manager's talent is at least partially unknown to other market participants '…investment decisions become tests that provide information about talent. Perceptions about talent, in turn, determine the manager's future opportunity wage, and this is what makes investments risky from the manager's perspective even if income is not explicitly tied to profits' (Holmstrom 1999, p. 178). Institutional investors insure the manager against career risks, as they collect independent information (the crucial point here is the assumption of better monitoring abilities by institutional investors) and assess the ability of a manager based on that information, and not solely based on revenue realization (Aghion et al. 2013).

[4] If we not only take into account the reported coefficients, but also include time dummies for each year as well as four-digit industry dummies, the obtained p values are even lower.

Table 3 Competition Poisson models—replication

Model	Poisson	Poisson	Poisson
Competition over time	Varies	Constant	Constant
Sample	Pooled	High comp.	Low comp.
Dependent variable	Cites	Cites	Cites
	(1)	(2)	(3)
Institutional owners	−0.064*	0.009***	0.000
	(0.030)	(0.001)	(0.003)
ln(capital/labor)	0.452***	0.564***	0.206*
	(0.141)	(0.115)	(0.081)
ln(sales)	0.189*	0.267***	0.086
	(0.075)	(0.070)	(0.063)
ln(R&D) stock	−0.001	−0.063	0.047
	(0.084)	(0.107)	(0.088)
Competition	−3.694		
	(3.330)		
Competition ×	0.082*		
Institutional owners	(0.035)		
Fixed effects controls	Yes	Yes	Yes
Observations	6208	3125	3083
AIC	444,893.694	298,559.057	131,075.768
BIC	445,923.934	299,000.502	131,636.898

This table shows the replication of a number of models from Table 2 in Aghion et al. (2013). All models are Poisson regression models. Standard errors (in parentheses) are clustered at three-digit industry level. There are 803 different firms. Additionally, all regressions include time dummies for each year (with reference category 1991) as well as four-digit industry dummies. Fixed effects controls are included using the presample mean-scaling estimator developed by Blundell and Powell (2004). An industry is classified as one with high competition if it is above the median of the $1 −$ Lerner index. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Thus, if competition is high, the positive effect of institutional investors on innovation should be stronger. According to the career concerns hypothesis, institutions and competition are complements (Aghion et al. 2013, p. 292).

Table 3 contains the replication of a selection of Poisson models conducted by Aghion et al. (2013, p. 293), with product market competition included as a further explanatory variable. The replication of all models from Table 2 in Aghion et al. (2013, p. 293) was successful. For the sake of brevity, only a selection of these models is presented here.

Column 1 displays a Poisson model with an interaction between product market competiton and institutional ownership. As the coefficient has a positive sign and is significant, this is indicative for the career concerns hypothesis and against the lazy manager hypothesis.

Columns 2 and 3 of Table 3 show two Poisson models based on a high- and low-competition sample, respectively. The coefficient of institutional ownership is positive and significant for the model based on the high-competition sample, whereas it is zero

and insignificant for the low-competition model. Again, this result delivers evidence for the career concerns hypothesis.

To show that the Poisson model with interaction term provides quantitatively nearly the same conclusions as the two Poisson models with the data split into high versus low product market competition, we use the following illustration. As a rough guideline, assume a value of 0.9 for high competition, and a value of 0.8 for low competition.[5] For high competition, the coefficient of institutional ownership (Table 3, column 1) is $-0.064 + 0.9 \cdot 0.082 = 0.0098$, whereas the coefficient has a value of $-0.064 + 0.8 \cdot 0.082 = 0.0016$ for low competition. These coefficients are very close to those in the models based on the high- and low-competition sample in Table 3, columns 2 and 3, with values of 0.009 and 0.000. Thus, with regard to the interpretation of the institutional ownership coefficient, there is hardly any difference between a modeling strategy based on an 'interaction view' and a 'sub-sample view.'

Table 4 contains the same models as Table 3, but again hurdle models are employed, instead of Poisson models. If both zero and nonzero citations were driven by the same determinants, then both parts of the hurdle model should deliver similar results as the Poisson model. Again, all hurdle tests accomplished indicate that a single equation is not enough (see Table 4). However, it is observed that the clear preference for the career concerns hypothesis and against the lazy manager hypothesis heavily depends on the model framework. When hurdle models are used, the empirical results do not show such a clear picture. Even more, there is some evidence that neither the lazy manager hypothesis nor the career concerns hypothesis can satisfactorily explain the mechanisms at work.

The coefficient of the interaction term of competition times the share of institutional owners [Hurdle NegBin (1)] is negative and significant (at the 5 % significance level) in the zero part of the model, whereas it is positive but not significant in the count part of the model. The fact that the interaction term in the zero part of the model has a negative sign cannot be explained by either the career concerns hypothesis or the lazy manager hypothesis because only the former leads to a significant interaction but requires that the coefficient of the interaction term is positive. However, here the influence of institutional ownership is different in the two parts of the model, which can also be observed in the models with the sample split into high and low competition. In the high-competition sample, the coefficient of institutional ownership is positive in both parts of the model, but only in the count part significant [Hurdle NegBin (2)], whereas the coefficient is positive in both parts of the model for the low-competition sample, too, but only significant in the zero part of the model [Hurdle NegBin (3)]. Furthermore, a similar numerical illustration as shown above also applies to hurdle models.[6]

---

[5] The first and third quartile of product market competition constant over time are 0.83 and 0.89, respectively. For the sake of simplicity, 0.8 and 0.9 are used in the illustration leading to qualitatively the same interpretation.

[6] Using the same numbers as before, the coefficient of institutional ownership [Table 4, count part of Hurdle NegBin (1)] is $-0.014 + 0.9 \cdot 0.019 = 0.0031$ for the high-competition sample and $-0.014 + 0.8 \cdot 0.019 = 0.0012$ for the low-competition sample, which is again approximately equal to the coefficients of the count parts in Table 4, Hurdle NegBin (2) and Hurdle NegBin (3), with values of 0.003 and 0.002. For the binary part of the models, the coefficient is $0.083 - 0.9 \cdot 0.086 = 0.0056$ and $0.083 - 0.8 \cdot 0.086 = 0.0142$

Table 4 Competition—hurdle models

Model	Hurdle NegBin (1)		Hurdle NegBin (2)		Hurdle NegBin (3)	
Competition over time	Varies		Constant		Constant	
Sample	Pooled		High comp.		Low comp.	
Dependent variable	Cites		Cites		Cites	
	Count	Zero	Count	Zero	Count	Zero
Institutional owners	−0.014	0.083*	0.003***	0.005·	0.002	0.012***
	(0.020)	(0.033)	(0.001)	(0.003)	(0.002)	(0.003)
ln(capital/labor)	0.252***	0.254**	0.239***	0.345**	0.229**	0.191·
	(0.051)	(0.081)	(0.049)	(0.129)	(0.086)	(0.102)
ln(sales)	0.117***	0.191***	0.111·	0.306***	0.118***	0.117*
	(0.031)	(0.059)	(0.060)	(0.080)	(0.030)	(0.059)
ln(R&D) stock	0.155**	0.109*	0.225**	0.094	0.081·	0.116*
	(0.051)	(0.053)	(0.085)	(0.107)	(0.046)	(0.047)
Competition	−1.965	1.873				
	(2.878)	(4.138)				
Competition×	0.019	−0.086*				
Institutional owners	(0.022)	(0.038)				
Fixed effects controls	Yes		Yes		Yes	
Observations	6208		3125		3083	
AIC	44,968.75		24,328.19		20,576.24	
BIC	47,035.96		25,217.12		21,704.54	
Hurdle test (*p* value)	0.000		0.000		0.000	

Standard errors (in parentheses) are clustered at three-digit industry level. There are 803 different firms. Additionally, all regressions include time dummies for each year (with reference category 1991) as well as four-digit industry dummies. Fixed effects controls are included using the presample mean-scaling estimator developed by Blundell and Powell (2004). The count part is a censored negative binomial with log link. The zero part is a truncated negative binomial with log link. The hurdle tests are based on all coefficients of the respective model given in the table. An industry is classified as one with high competition if it is above the median of the 1 − Lerner index. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '·' 0.1 ' ' 1

A possible explanation might be the risk behavior of institutional investors. Xu et al. (2015) find that institutional owners are risk averse preferring loss aversion over superior performance (see also Bodnaruk and Simonov 2016; Porter and Trifts 2014), and do not encourage (or even impede) innovation activities which carry greater risk. Our results indicate that this might be the case in highly competitive and thus potentially riskier markets. Summing up, the influence of institutional ownership on innovation is in both parts of the models positive, but not in every part significant. This result can be seen as an image with a higher resolution than is possible to generate within the Poisson approach.

In a nutshell, the discrimination between the career concerns hypothesis on the one hand and the lazy manager hypothesis on the other hand based on Poisson models does not produce the same results as a modeling strategy based on a hurdle (or even a negative binomial) model.

As a consequence, this raises the issue whether the conclusions drawn in the Aghion et al. (2013) paper are robust. Of course, we most certainly cannot stand here and say that the use of hurdle models is the most suitable solution. In the alternative, should it be held that the Poisson model would be the most appropriate model, the hurdle model should deliver similar results. The results we have achieved do not point in that direction, though. Instead, we find that innovation is differently affected by institutional owners in the two parts of a hurdle model. Any conclusion that shows preference for the career concerns hypothesis and against the lazy manager hypothesis cannot be drawn in such a setting.

4 Conclusion

Based on a successful replication in a narrow sense of (parts of) the Aghion et al. (2013) paper, a refinement of the modeling strategy reveals differences with regard to the sign and significance of the coefficients of interest, and hence also to the interpretation. The results obtained from an extended analysis with negative binomial hurdle models differ materially from the outcomes of the single-equation Poisson approach carried out by Aghion et al. (2013). While the Poisson-based model delivers a clear indication for the career concerns hypothesis and against the lazy manager hypothesis, there is evidence for a deviation from the Poisson model. Both economic and statistical considerations suggest that different determinants may drive zero and nonzero patent citations. If a single set of determinants is sufficient for describing the underlying innovation processes, then both parts of the hurdle model should lead to estimates that are similar to the coefficients from the Poisson equation. However, both sign and significance of the coefficients of interest change when employing the hurdle model (using the same type of clustered standard errors as in the Poisson case), providing evidence for differing determinants. In the event that the innovation process can accurately be

---

Footnote6 continued

for the high- and low-competition sample, respectively. The coefficients of institutional ownership in the binary part in Table 4, Hurdle NegBin (2) and Hurdle NegBin (3) are 0.005 and 0.012. Equivalent to the illustration based on Poisson models, hardly any differences should be found between an 'interaction view' and a 'sub-sample view.'

described as one equation, both parts of the hurdle model should deliver similar results as the Poisson model. All the hurdle tests performed indicate that a single equation is not enough.

From an economic perspective, this is a difference in determinants of 'first innovation' and 'continuing innovation,' respectively. If it is the case that the first innovation (the first citation-weighted patent count) is especially hard to obtain in comparison with succeeding innovations, hurdle models offer a useful way that allows for a distinction to be made between these two processes. The rationale behind this is the notion of nonlinearities in the innovation process (König and Licht 1995; Licht and Zoz 1996; Crepon and Duguet 1997; Brouwer and Kleinknecht 1999; Mohnen and Röller 2005; Mäkinen 2007). Then, the processes that drive innovative outcome might themselves be affected by a number of different driving factors.

The empirical findings resulting from the extended analysis concerning the discrimination between the lazy manager hypothesis and the career concerns hypothesis indicate that neither hypothesis is completely satisfactory. The unambiguous, interpretable picture arising out of the analysis of Aghion et al. (2013) cannot be fully confirmed when hurdle models are used. A possible explanation for our results is the risk behavior of institutional investors, who do not encourage more hazardous innovation activities (as also found by Xu et al. 2015). Nevertheless, there is some evidence for a remarkably stable positive correlation of citation-weighted patents and institutional ownership across specifications.

5 Computational details

The results of the empirical analysis are obtained using R 3.1.0 (R Core Team 2013) with the packages AER 1.2-1, foreign 0.8-61, MASS 7.3-33 (Venables and Ripley 2002), countreg 0.1-3 (Zeileis et al. 2008) especially for the function `hurdle()`, and sandwich 2.3-0 (Zeileis 2006).

References

Aghion P, Van Reenen J, Zingales L (2013) Innovation and institutional ownership. Am Econ Rev 103(1):277–304. Data and online appendix, doi:10.1257/aer.103.1.277. Accessed 30 July 2014

Archibugi D, Planta M (1996) Measuring technological change through patents and innovation surveys. Technovation 16(9):451–519

Basberg BL (1987) Patents and the measurement of technological change: a survey of the literature. Res Policy 16(2):131–141

Baysinger BD, Kosnik RD, Turk TA (1991) Effects of board and ownership structure on corporate R&D strategy. Acad Manag J 34(1):205–214

Bertrand M, Mullainathan S (2003) Enjoying the quiet life? Corporate governance and managerial preferences. J Polit Econ 111(5):1043–1075

Blundell RW, Powell JL (2004) Endogeneity in semiparametric binary response models. Rev Econ Stud 71(3):655–679

Bodnaruk A, Simonov A (2016) Loss-averse preferences, performance, and career success of institutional investors. Rev Financ Stud. doi:10.1093/rfs/hhw053

Brouwer E, Kleinknecht A (1999) Innovative output and a firm's propensity to patent: an exploration of CIS micro data. Res Policy 28(6):615–624

Bushee BJ (1998) The influence of institutional investors on myopic R&D investment behavior. Acc Rev 73(3):305–333

Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge

Comanor WS, Scherer FM (1969) Patent statistics as a measure of technical change. J Polit Econ 77(3):392–398

Crepon B, Duguet E (1997) Research and development, competition and innovation pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity. J Econom 79(2):355–378

David P, Hitt MA, Gimeno J (2001) The influence of activism by institutional investors on R&D. Acad Manag J 44(1):144–157

Fama EF (1980) Agency problems and the theory of the firm. J Polit Econ 88(2):288–307

Graves SB (1988) Institutional ownership and corporate R&D in the computer industry. Acad Manag J 31(2):417–428

Griliches Z (1990) Patent statistics as economic indicators: a survey. Tech. rep, National Bureau of Economic Research

Hansen GS, Hill CW (1991) Are institutional investors myopic? A time-series study of four technology-driven industries. Strateg Manag J 12(1):1–16

Hart OD (1983) The market mechanism as an incentive scheme. Bell J Econ 14(2):366–382

Hausman JA, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents–R&D relationship. Econometrica 52(4):909–938

Holmstrom B (1999) Managerial incentive problems: a dynamic perspective. Rev Econ Stud 66(1):169–182

Kleinknecht A, Van Montfort K, Brouwer E (2002) The non-trivial choice between innovation indicators. Econ Innov New Tech 11(2):109–121

Kochhar R, David P (1996) Institutional investors and firm innovation: a test of competing hypotheses. Strateg Manag J 17(1):73–84

König H, Licht G (1995) Patents, R&D, and innovation: evidence from the Mannheim innovation panel. ifo Studien Zeitschrift für empirische Wirtschaftsforschung 4(95):521–543

Licht G, Zoz K (1996) Patents and R&D: an econometric investigation using applications for German, European, and US patents by German companies. ZEW Discussion Paper no. 96-19, Zentrum für Europäische Wirtschaftsforschung, Mannheim

Mäkinen I (2007) To patent or not to patent? An innovation-level investigation of the propensity to patent, VTT

Menkhoff L (2002) Institutional investors: the external costs of a successful innovation. J Econ Issues 36(4):907–933

Mohnen P, Röller LH (2005) Complementarities in innovation policy. Eur Econ Rev 49(6):1431–1450

Porter GE, Trifts JW (2014) The career paths of mutual fund managers: the role of merit. Financ Anal J 70(4):55–71

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

StataCorp (2011) Stata data analysis statistical software: Release 12. StataCorp LP, College Station, TX

Van Der Panne G, Kleinknecht A (2005) Proximity and knowledge spillovers: evidence from new product announcements in the Netherlands. In: World Bank–Cambridge-MIT Workshop on University-Industry Linkages in Europe and North America, Cambridge, UK, September, vol 26. http://homepages.ulb.ac.be/~mcincera/INIR2003/PROGRAM.HTML

Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York

Winkelmann R (2010) Econometric analysis of count data. Springer, New York

Xu X, Wang C, Cheng S (2015) Myopic investor or active monitor? The role of institutional investors in corporate innovation. Int J Financ Res 6(2):23–32

Zeileis A (2006) Object-oriented computation of sandwich estimators. J Stat Softw 16(9):1–16

Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. J Stat Softw 27(8):1–25