

scottish institute for research in economics



SIRE DISCUSSION PAPER

SIRE-DP-2012-60

Dr. Nicola Searle
University of Abertay

Professor Gavin C Reid
University of St Andrews

Firm size and trade secret intensity: evidence from the Economic Espionage Act
By Nicola C Searle and Gavin C Reid

Dr. Nicola Searle
Senior Knowledge Exchange Associate
Institute for Arts, Media & Computer Games
University of Abertay
Kydd Building
50 Bell St
Dundee
DD1 1HG
Phone: 00-44- (0) 1382 308538
E-mail: n.searle@abertay.ac.uk

Professor Gavin C Reid MA MSc PhD FRSA FFCS
Founder/Director
Centre for Research into Industry, Enterprise, Finance and the Firm (CRIEFF)
School of Economics and Finance
University of St Andrews
St Salvator's College
St Andrews
Fife
KY16 9AL
Phone (Direct Line): 00-44- (0) 1334 462431
Fax (CRIEFF): 00-44- (0) 1334 462444
E-mail: gcr@st-andrews.ac.uk

Abstract:

This paper considers trade secrecy as an appropriation mechanism in the context of the US Economic Espionage Act (EEA) 1996. We examine the relation between trade secret intensity and firm size, using a cross section of 95 court cases. The paper builds on extant work in three respects. First, we create a unique body of evidence, using EEA prosecutions from 1996 to 2008. Second, we use an econometric approach to measurement, estimation and hypothesis testing. This allows us comprehensively to test the robustness of findings. Third, we focus on objectively measured valuations, instead of the subjective, self-reported values used elsewhere. We find a stable, robust value for the elasticity of trade secret intensity with respect to firm size, which indicates that a 10% reduction in firm size leads to a 7% increase in trade secret intensity. We find that this result is not sensitive to industrial sector, sample trimming, or functional form.

Key words: Firm size, economic espionage, intellectual property, trade secrets

JEL codes: D2, K2, L5, O3

1. Introduction

This paper considers trade secrecy as an appropriation mechanism, in the context of the US Economic Espionage Act (EEA) 1996, with the FBI as an enforcement agency, should an infringement of trade secrecy be detected. We report on our findings on the relation between trade secret intensity and the size of firms, using a cross section of ninety five cases. The paper builds on the work of Mansfield (1986), Arundel (2001) and Lerner (2006) in three respects. First, it creates a unique database using EEA prosecutions, arising from FBI investigations, from 1996 to 2008. This extends Lerner's (2006) use of litigation evidence to investigate the relative efficacy of trade secrets, as compared to patents, lawsuits. Second, it uses an explicitly econometric approach to measurement, estimation and hypothesis testing, as opposed to the descriptive and statistical approach of Mansfield (1986) in his exploration of the relation between non-patent IP protection (largely trade secrecy) and firm size. This allows us to engage in explicit, and comprehensive, testing of the robustness of our findings. Third, it focusses on actual IP valuations (by several techniques, and by courtroom cross reference values), as opposed to the subjective, self-reported values used by Arundel (2001), who found that trade secrecy was valued relatively more highly, the smaller was the size of the firm. We determine a value for the elasticity of trade secret intensity with respect to firm size, which indicates that a 10% reduction in firm size leads to a 7% increase in trade secret intensity. This result is found to be stable and robust, in that it is not sensitive to the industrial sector, to sample trimming, or to the functional form.

In the last decade, evidence and analysis have emerged giving a new emphasis to trade secrecy (TS) as a form of intellectual property (IP), especially in its strategic role within firms, see Jensen and Webster (2006), Lerner (2006), Cugno and Ottoz, E. (2006), Cohen, Nelson and Walsh (2000), Zabochnik (2002), and Arundel (2001). Whilst patents and copyrights remain at the forefront of debates about innovation, appropriation and enforcement, trade secrecy, which was once excluded from consideration as a key category of IP is now being taken much more seriously, Graves (2008). Not the least reason for this is the strong enforcement powers the FBI now has (in the USA and beyond) to bring

miscreants to court, Merriam (2009), and the high profiles of some recent cases, including the sentencing of research scientist Kexue Huang to seven years in prison for foreign economic espionage, (see press release, Department of Justice, Office of Public Affairs, December 21, 2011). However, while professional practitioner writings, from both the legal, Halligan (2005), Lemley (2009), and intellectual property camps, Krotoski (2009), have moved rapidly to develop new modes of analysis of trade secrecy, economics writings have been much slower to respond. Those papers which have started to analyse trade secrets usually only do so from the perspective of patents and copyright research, Zwillinger and Genetski (2001), Slowinski, Hummel, and Kumpf (2006). This paper aims to remedy this deficiency, by taking the 'new learning', of trade secrets as IP, as given, and developing thereon an econometric explanation of trade secret intensity (TSI).

Our paper contributes three things, in increasing order of importance: (a) it develops a new database, using court records, on the use of trade secrecy; (b) it uses actual (rather than imputed) valuations of trade secrets; and (c) it develops a well-grounded econometric model of trade secret intensity (TSI). In particular, we find a stable, robust, inverse relationship between firm size and trade secret intensity (TSI), suggesting that small firms have greater TSI than large firms.

The structure of the paper is as follows. Section 2 (Framework) develops a framework for analysing the relationships between firm size and trade secret intensity, by reference to key literature in the intellectual property and industrial organization fields. Section 3 (Data) explains how our unique database, using court records from prosecutions under the USA Economic Espionage Act (1996), was constructed; and defines the ten key variables for use in our econometric modelling. Section 4 (Model) sets out the model, and presents some exploratory data analysis, supported by graphs. Section 5 (Results) reports on seven regressions, including those suggested by functional form analysis, using the Box-Cox transformation. Section 6 (Robustness) reports on further testing of the model (for endogeneity, outliers, trimming etc.). Section 7 (Conclusions) reports

on the overall significance of our findings, in terms of data, empirical evidence and analysis, and future prospects for research in this area.

2. Framework

In this section we establish a framework for our analysis of firm size and trade secret intensity (TSI). Interest in trade secrets as an alternative appropriation measure (e.g. to patenting) has grown in recent years. However, the necessarily undisclosed nature of trade secrets has to some extent thwarted efforts directed at their further study. As a result, there are rather few empirical examinations of their use and purpose. Examples of empirical studies of trade secrets fall largely into two categories: evidence from litigation (for example, Lerner, 2006, Almeling et al, 2010) or survey results (for example, Jensen and Webster, 2006, Cohen et al, 2001, Arundel, 2001.) This study falls into the former category as an empirical study into the criminal prosecution of the theft of trade secrets and their relationship to the firm.

Empirical studies highlight the importance of trade secrets but there is relatively modest analysis of the relationship between trade secrets and firm size.

However, the theoretical literature has put forward arguments to describe this relationship. To capture the scope of these theoretical arguments, we proceed with the assumption that theoretical arguments *for* the use of patents represent a theoretical argument *against* the use of trade secrets. This assumption has its limitations (Willoughby, 2010) as, while patents and trade secrets are typically substitutes, they can also serve as complements. We find literature arguing both for and against an inverse relationship between trade secrets and firm size, as will be discussed later in this section.

To move from theoretical arguments to empirical implementation, we will construct a measure of *Trade Secret Intensity (TSI)*, the purpose of which is to calibrate the firm's use of trade secrets objectively. As befits the term 'intensity,' we seek a relative measure of the extent of use of trade secrets (cf. Bosworth and Rogers, 2001, who use similar measures in their analysis of R&D intensity), of

the form (y/x) . For our purposes, the numerator is a measure of the use of trade secrets (y); and the denominator (x), is a measure of the scale, or size, of the business. The numerator could be measured in a variety of ways; for example, as a count of use of trade secrets, or as the value of some or all of a firm's trade secrets. In the denominator, measures of scale may be used, including employment, sales, and assets. Put generally, we have:

$$\text{Trade Secret Intensity (TSI)} = \frac{\text{(Extent of Use of Trade Secrets)}}{\text{(Scale of Business)}} \quad (1)$$

Here, it will be evident that we will already have to address issues of how we measure both the numerator and the denominator of equation (1). This done, we seek to provide a good explanation of the trade secret intensity (TSI) using econometric estimation on our cross-section of data. This work will explore: the choice of functional form; the choice of appropriate explanatory variables, including behavioural variables and control variables; and the robustness of findings. At its simplest, the theoretical relationship we have in mind is: $TSI = f(\text{firm size})$ which may be written $y = f(x)$ with the restrictions $f' < 0$ and $f'' > 0$: that is, a monotonically decreasing, and convex relationship between trade secret intensity and firm size. This may be extended to the empirical relationship:

$$TSI \equiv y = f(x, \mathbf{Z}, \mathbf{V}; \varepsilon) \quad (2)$$

where, in (1), x is our chosen size measure, \mathbf{Z} is a set of behavioural variables (e.g. measures of value or of scope of innovation for IP protection), \mathbf{V} is a set of control variables (e.g. of sector, or of defendant status), and ε is a random variable.

This is the main point of departure for developing a more complete specification as below, Sections 5 and 6. As will unfold, our general view is that one would expect a negative relation between trade secrecy and firm size, but there are counter-arguments. We shall consider these counter-arguments first, namely that the use of trade secrets is positively related to firm size, that is, $f' > 0$.

To illustrate, Arundel (2001) argues that larger firms prefer trade secrets over patents as larger firms have the power of their market strength to create a lead-

time advantage which is denied to smaller firms. As disclosure via patenting aids the development of competing goods, large firms prefer trade secrecy, which limits disclosure. Scherer (1965) argues that larger firms receive less marginal benefit from patents as larger firms are more sensitive to disclosure via patenting and do not require patents in order to secure financing and enable partnerships. This suggests that smaller firms are more reliant on patents in forming partnerships and obtaining financing. Levin et al (1988) and Cohen et al (2001) concur with this as they argue that smaller firms need a patent portfolio in order to compete in the market. Scherer (1965) and Arundel (2001) further argue that patents can create a protective buffer for smaller firms. They argue that small firms may be unable to exploit innovations quickly due to their limited manufacturing and marketing capacity. Patents create a legal buffer against larger firms, who are able to exploit market power and benefit from economies of scale. Following these arguments, we conclude that there is a tendency for larger firms to prefer trade secrets and for smaller firms to prefer patents as a means of appropriation.

Empirical support for a positive relationship between firm size and the use of trade secrets can also be found in the works of Jensen and Webster (2006) and Cohen et al (2001.) Jensen and Webster combine survey and patent data to argue that larger firms have lower patent intensities than SMEs. Cohen et al (2001) demonstrate, by their survey, that larger firms cite the motive of patenting to enhance their firm's reputation less than frequently than do smaller firms.

A converse line of argument is that use of trade secrets is negatively related to firm size and thus $f' < 0$; and further, that trade secrecy increases more the smaller is the firm, $f'' > 0$. To illustrate, a possibly suitable functional form for $f(.)$ is the equation we have fitted to our raw (untransformed) sample of data on trade secrecy (ordinate) and size (abscissa), in Figure 1 below. It satisfies the restrictions $f' < 0$, and $f'' > 0$. If y is trade secrecy and x is firm size, then this equation can be written $\ln y = a - b \ln x$, with constants $a, b > 0$, or $y = f(x) = Ax^{-b}$, where $\ln A = a$. Then taking derivatives, $dy/dx = f' = -b (y/x) < 0$ for $x, y > 0$, and

$d^2y/dx^2 = f'' = b(b + 1)A x^{-b-2} > 0$, which satisfies the two qualitative restrictions we have considered to imposed on $f(\cdot)$, namely negative monotonicity and convexity. This suggests a log-linear functional form is worthy of consideration for estimation purposes, but of course that is a matter of empirical testing, a task to which we turn in Sections 4, 5 and 6 below.

In support of our main line of argument, the analyses of Lerner (1996), Cordes et al (1999) and Arundel (2001), would hold that the restrictions we have discussed may well occur in practice, as patenting may be too costly for smaller firms, and that this cost constraint binds particularly firmly, the smaller is the firm. Thus, patenting and patent protection are relatively expensive for smaller firms, and trade secrets can prove a more cost-effective protection for innovation. Arundel (2001) also sees favour in the Schumpeterian argument, that smaller firms may produce smaller, more incremental innovations than larger firms, and may therefore produce relatively less patentable innovations. Reid and Ujjual (2008) represent the Schumpeterian hypothesis (which they express in terms of scale economies in R&D) as stemming from Schumpeter's (1942) seminal work, arguing that larger firms are more innovative than smaller firms, because size allows a greater degree of specialisation, and indeed fosters the emergence of a highly creative technological elite within firms, who may be very effective in radical innovation, leading to 'creative destruction'. Such innovations do lend themselves to patent protection. However, they also find a lower-level stable equilibrium for firms, of SME magnitude, who are particularly adept at niche exploitation in their innovation strategy, with more emphasis on incremental innovation. Such innovations do lend themselves to protection by trade secrecy. Arundel (2001) argues larger firms will prefer patents to trade secrecy, as economies of scale reduce the marginal cost of patenting. Scherer (1965) and Jensen and Webster (2006) also use this argument. Jensen and Webster (2006) further argue that relative costs of litigation of patents for larger firms are lower than for smaller firms. Larger firms stand to benefit more from a reputation for aggressive litigation, which will dissuade would-be infringers. Thus, overall, these papers develop a persuasive case for arguing that smaller firms prefer trade secrets and larger firms prefer patents.

Empirical evidence in support of the negative relationship between the use of trade secrets and firm size is found in various survey and patent data studies. Leiponen and Byma (2010), using a survey of small Finnish firms, determine that firms who focus on process innovations, with modest R&D investments or cooperative activities, prefer trade secrecy over speed to market and patents. Arundel (2001) uses survey data to show that small firms value secrecy more than large firms. In another survey, Arundel & Kabla (1998) find that the tendency to patent increases with firm's size. Using a combination of survey and patent data, Mansfield (1986) shows a positive correlation between firm size and the percentage of patentable innovations that are patented. Finally, Scherer (1983) uses R&D and patent data to demonstrate that expenditures on R&D are positively correlated with patenting activity.

Additional insights, which enhance our understanding of the dynamics of innovation, can be found in papers addressing the general relationship between firm size and innovation. Baldwin et al (2000) support the Schumpeterian hypothesis that firm size is positively related to innovation, although empirically they find that the relationship is non-monotonic. They also suggest a positive relationship between the use of trade secrets and innovation. Lunn (1986) finds that the firm's market power and the market concentration of the industry are positively correlated with process innovations. Finally, Cohen et al (1987) find weak support for the Schumpeterian hypothesis as their analysis indicates the size of the firm has only a small positive effect on R&D intensity. However, their research finds that fixed industry effects have a stronger effect on R&D intensity.

The literature detailing the type of innovation and innovators also aids our analysis. As Friedman et al (1991) argue, firms prefer to use trade secrets rather than patents for protecting process innovations. Png (2011) finds that increased protection for trade secrets results in reduced patenting activity in industries where patents for process innovations are effective. In a meta-analysis of firm size and process - product innovation research, Damanpour (2010) finds that the studies support a positive relationship between size and innovation; however, he

fails to find evidence of a relationship between size and a preference for process or product innovations. Indeed, Leiponem and Byman (2010) find small firms engaged in cooperative R&D prefer speed to market over patents and trade secrets.

However, with the partial exception of Baldwin et al (2000), empirical studies do not focus on the relationship between trade secrets and firm size specifically. The lack of consensus regarding the relationship between trade secrets and firm size indicates that this area requires further work. Thus, the empirical work of this paper aims to provide a resolution of the equivocal position presented by analysts of innovation.

3. Data

In this sections we explain how our database was created, and set out the key variables from it that we will be using in our econometric estimation in Section 4 (The Model) of this paper. In 1996, the U.S. enacted the Economic Espionage Act. The act increased the protection of trade secrets by elevating the theft of trade secrets from a civil malfeasance to a criminal felony. In 2011, efforts to increase the prescribed penalties under the act began with The Economic Espionage Act Penalty Enhancement Act. Arguing that the, 'as much as 80 percent of the assets of today's companies are intangible trade secrets,' (Kohl, 2011), supporters of the bill seek to increase the maximum sentence afforded by the EEA.

Our data stem from a sample of 95 cases considered under the EEA from 1996 to 2008. Under the EEA, victim firms report the theft of a trade secret to the FBI who then investigates and prosecutes cases. Records associated with these prosecutions are made public which is a huge boon to researchers for two reasons. First, data regarding these stolen trade secrets would have otherwise remained secret within the victim firm. Second, the EEA represents the harmonisation of trade secret law at the federal level whereas prior to the EEA, trade secrets were dealt with heterogeneous statues at the state level and difficult to research. Thus, these cases provide a unique insight into firms' use of trade secrecy.

Cases were first identified via the Public Access to Electronic Records (PACER). Data was collected using a variety of sources, including court documents, media reports, company databases, FBI press releases, and academic papers.

Variables

In most cases, qualitative information on the nature of the trade secret was available via the Indictment documents or media reports. These descriptions of the trade secrets, such as project plans, prototypes, bid information et cetera, were then categorized according to the type of IP protection available (namely *patentable* or *copyright*.)

Quantitative information on the trade secret was relatively less available. The estimation of the value of the trade secret, which is more of an art than a science (Henry and Turner, 2007), is fraught with lack of availability or conflicting sources. Valuations were found in court documents, media report, and academic papers (Zwillinger and Grenetski (2000) and Carr and Gorman (2001). The method used in the construction of this database is consistent with Carr and Gorman's (2001) 'objective' estimate and relies on the low end (*Low*) of the value range for a particular trade secret.¹

Information on the victim firm was collected from the victim firm's website and databases² to establish the firm's annual revenue (*vsales*), number of employees and primary activities. Firms were then classified using the Small Business Administration's (SBA) definitions of small business by sector (*sbdummy*) and by primary industrial sector of the victim firm (*SIC*). Firms were then grouped into dummy variables to account for the most common sectors (*mandum* or *servdum*.)

¹ Models of value in this context are numerous, a half dozen or so, and are dealt with in detail by Searle (2011). Use of other models for values does not unduly change our results.

² In the first instance, via Edgar Online (the Electronic Data Gathering, Analysis and Retrieval System of the U.S. Securities and Exchange Commission.) When Edgar was incomplete, further information was found via commercial databases.

Finally, information on the defendant was gathered. This allowed for details regarding the relationship of the defendant to the victim firm (*outsider*) and nationality of the defendant (*foreign*). For convictions,³ the number of incarceration months was used to determine the level of loss determined by the court. The resulting variable, *Xref*, provides another estimate of the value of the trade secret. As detailed in Zwillinger and Grenetski (2000), this method uses sentencing to estimate the offence points determined by the court and then infers the loss associated with the theft of the trade secrets.

Given the nature of the data collection process and the EEA cases themselves, data is sometimes not available for all observations. To account for these missing observations, the mean is used to replace missing values. Additional regression work (for example, our extensive testing, reported in Section 5 using trimmed regression) indicates that this model of missing values does not unduly distort the analysis.

³ As some cases are ongoing, not all cases would have concluded in the time frame of the sample.

Table 1: Definitions of Regression Variables

<i>Variable</i>	<i>Definition of Variables⁴</i>				<i>n=95</i>
	type	description	mean	standard deviation	
<i>TSI</i>	continuous	Trade Secrets Intensity (Value of secret/ Victim Firm's No. of Employees)	32,097	114,380	
<i>vsales</i>	continuous	Victim Firm's Annual Sales Revenue (dollars)	10.9 e+9	17.8 e+9	
<i>Xref</i>	continuous	Cross Referenced Value of Trade Secret (2008 dollars)	711,786	2,048,750	
<i>manudum</i>	dummy	Manufacturing sector	0.56	0.50	
<i>servdum</i>	dummy	Service sector	0.18	0.39	
<i>patentable</i>	dummy	Potentially patentable secret	0.39	0.49	
<i>copyright</i>	dummy	Potentially copyrightable secret	0.21	0.41	
<i>outsider</i>	dummy	Defendant is outsider	0.16	0.37	
<i>foreign</i>	dummy	Defendant is foreign	0.22	0.42	
<i>sbdummy</i>	dummy	Victim firm is small business	0.26	0.44	

Note: the mean and standard deviation is calculated on the data after missing value analysis; thus, the reported standard deviation is lower than the pre-missing value input.

4. Model

⁴ The count of non-missing values for the relevant variables are *Low* = 31, *vworkers* = 66, *vsales* = 76, *Xref* = 42 and *outsider* = 85.

The reference point for the model is Equation) with $f' > 0$. To begin with, we construct a suitable proxy of the dependent variable, *TSI*, which is given in expression (3) below. A value measure of trade secrets (*Low*) is preferred to a count measure, as being more accurate, and offering superior economic interpretation. To normalize the value measure, we use the headcount of firm employees as the size measure in the denominator. This allows, in terms of economic interpretation, *TSI* to be treated as a type of productivity measure, gauging the value of trade secrets generated per employee. Furthermore, in terms of our proposed econometric procedures, this ratio measurement has the merit of mitigating the potential statistical problem of heteroskedasticity, which may otherwise cause considerable efficiency loss in the estimation of our key elasticity parameter.

$$TSI = \frac{\text{Extent of use of Trade Secrets}}{\text{Scale of Business}} = \frac{\text{Value of Trade Secrets (low)}}{\text{Number of Firm Employees}} \quad (3)$$

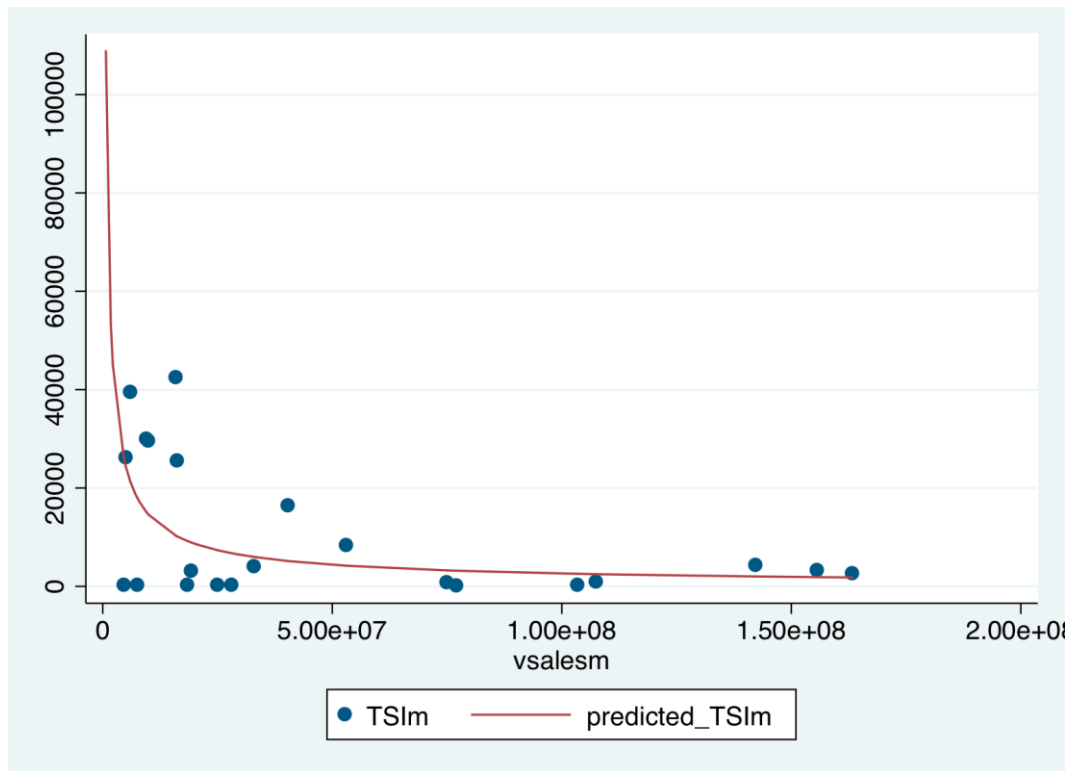
Our focus is on those variables that determine trade secret intensity, and the functional form for representing this relationship. Using the above (3) expression for *TSI*, equation (2) above may be written in a variety of empirical forms for estimation of which one of the most interesting (see Table 2) is:

$$\ln(TSI)_i = \beta_0 + \beta_1 \ln(vsales)_i + \beta_2 \ln(xref)_i + \beta_3 [\ln(vsales)_i \times \ln(xref)_i] + \beta^T \mathbf{X}_i + u_i \quad (i = 1, 2, \dots, N) \quad (4)$$

In equation (4) there is an interaction term between two of the behavioural variables, *vsales* and *xref*, and the remaining behavioural and control variables are subsumed into the vector \mathbf{X} , with corresponding coefficient vector β .

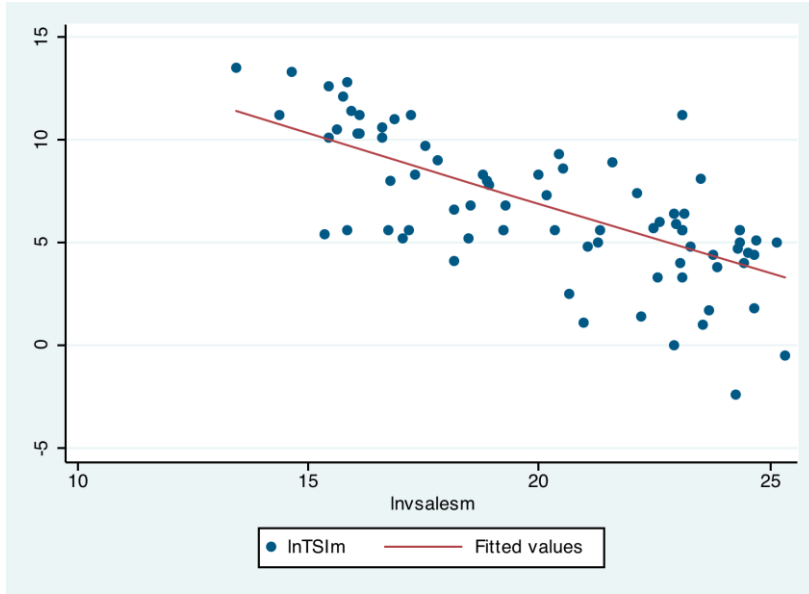
In our investigation of firm's use of trade secrets, we measure the relationship between the use of trade secrets (*TSI*) and the size of the victim firm (*vsales*). A scatter plot of the relationship between *TSI* and *vsales* reveals the following:

Fig. 1: Scatter Plot of TSI and vsales



The scatter plot in Fig. 1 suggests a nonlinear relationship between trade secret intensity (*TSI*) and size (here, as measured by value of sales, or sales revenue). Specifically a negative, convex relationship is suggested by the scatter of data points. We have superimposed the equation $TSI = \text{Exp}(20.47) \times \text{vsales}^{-0.68}$ for illustrative purposes. This is actually an estimate presented below in Table 2, column two, and discussed further later. Applying a log-log transformation to *TSI* and *vsales*, we have the following graph:

Fig. 2: Log-log Scatter Plot of TSI and vsales with Regression Line



Superimposed on Fig. 2 is a log-linear regression line fitted by least squares: $\ln(TSI) = 20.3 - 0.67 \ln(vsales)$, which is a slight adaptation of the estimate reported in Table 2, column 3, below. Thus, a visual interpretation of the data might suggest that a log-log transformation is appropriate. Fortunately, the Box-Cox test, Box and Cox (1964), allows us to test for an appropriate functional form, including the double log form, and this inferential work is undertaken in the next section.

5. Results

We begin with the simple (and clearly unacceptable) bivariate linear regression model using untransformed variables:

$$TSI = \beta_0 + \beta_1 vsalesm + \varepsilon \quad (5)$$

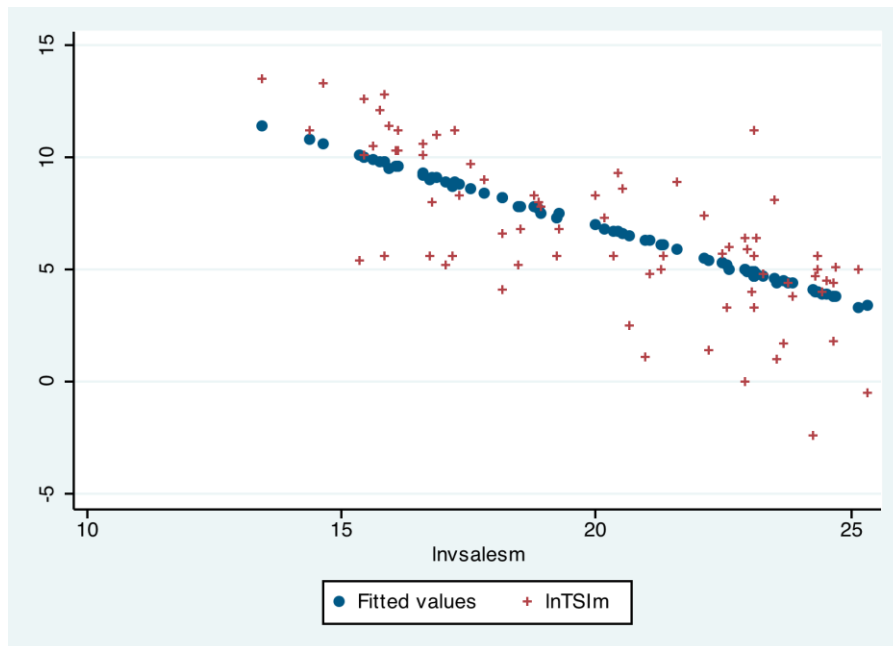
We can test for a variety of alternative functional forms to equation (5) using the Box-Cox transformation. For a generic variable y_i this transformation is: $y_i^{(\lambda)} = (y_i^\lambda - 1)/\lambda$ if $\lambda \neq 0$ and $= \log(y_i)$ if $\lambda = 0$. Starting with equation (5), and exploring its transformation possibilities, the Box-Cox test statistics for λ rejects a reciprocal transformation ($\lambda = -1$) of the dependent and independent variable; and also rejects no transformation ($\lambda = 1$). The test cannot reject $\lambda = 0$ (p-value = 0.59) which therefore suggests using a logarithmic transformation of the dependent and independent variables. Further Box-Cox testing on

generalizations of the simple linear model, including additional variables, also confirms the wisdom of using a log-linear transformation.

To test the relationship between *TSI* and *vsales*, and other variables, in different forms, we generally used a regression strictly linear in the parameters, with *TSI* as the dependent variable. As noted in Hall et al (2011), a number of authors suggest IP strategies vary by industry (e.g. Mansfield, 1986, Arundel & Kabla, 1998, Cohen et al., 2000, Jain and Kiran, 2012). Yoon and Lee (2008) also argue the manufacturing sectors prefer patents whereas service sectors prefer trade secrecy. Here we have therefore used dummy variables for the manufacturing and service sectors (*manudum* and *servdum*)⁵ to account for sectoral effects. The results can be found in Table 2. Overall, this model is a good fit to the data, having an R squared of 0.50 and is statistically significant (p-value of 0.00), but the dummy variables are not statistically significant. Fig. 3 below shows a scatter plot of the predicted (fitted) values of this regression versus the observed values. We note first that the size coefficient is negative (-0.673), highly statistically significant and less than unity in absolute value. Considered as an elasticity, this coefficient suggests that a proportional increase in size, ceteris paribus, leads to a less than proportional decrease in trade secret intensity (TSI).

⁵ The *manudum* and *servdum* dummy variables collectively account for 74% of our observations. So, these manufacturing and service dummies, whilst mutually exclusive, are not exhaustive. Not included, for example, are agriculture, forestry, fishing, mining and construction, for the lower US SIC codes (100-1700) and public administration, for the higher US SIC codes (9100-9900).

Fig. 3: Log Linear Model with Sector Dummies: Scatter Plot of Fitted versus Observed $\ln(TSI)$



However, as observed above, both the *manudum* (US SIC 2000- 3900) and *servdum* (US SIC 4000 -8900) dummy variables are not statistically significant. This suggests that sectoral category of economic activity does not affect the *TSI* of a firm. Further regressions using a variety of sector dummy variables for narrower SIC codes (for example, construction and transportation) also failed to uncover any sectoral effects on $\ln(TSI)$.

As noted in Hall et al (2011), a number of firm and industry characteristics might influence the firm's choice of IP. We consider independent variables including the court's valuation of the trade secret (*Xref*), an interaction term, the thief (*outsider* or *foreigner*), and other available appropriation mechanisms (*patentable* or *copyright*), to control for their effects as reported in Table 2. However, the last four are not significant.

Table 2: Log-linear Regressions with Combinations of Variables

Log-linear Regression Results							
Dependent variable = $\ln(TSI)$							
Variables							
$\ln(vsales)$	-0.68	-0.67	-0.68	-1.48	-0.65	-0.67	-0.67
	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>
manudum		0.18					
		<i>(0.74)</i>					
servdum		0.20					
		<i>(0.78)</i>					
$\ln(xref)$			0.00	-1.37			
			<i>(0.99)</i>	<i>(0.07)</i>			
$\ln(vsales)$ $\times \ln(xref)$				0.07			
				<i>(0.07)</i>			
outsider					0.21		
					<i>(0.76)</i>		
foreign						-0.49	
						<i>(0.38)</i>	
patentable							-0.29
							<i>(0.58)</i>
copyright							-0.60
							<i>(0.34)</i>
constant	20.47	20.26	20.49	37.03	19.82	20.47	20.35
	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>
Overall p-value	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>	<i>(0.00)</i>
R-squared	0.50	0.50	0.50	0.52	0.46	0.50	0.50
N	95.00	95.00	95.00	95.00	85.00	95.00	95.00

Values in parentheses are the p-value of the coefficient; italics indicate a p-value of less than 5%. Bold indicates a coefficient that is significant at the 5% level

Of interest in Table 2 is the the model in column five, which is the estimated version of (4) above. Both $\ln(xref)$ and the interaction term $\ln(xref) \times \ln(vsales)$ are significant at the 10% level. This will prove to be a useful comparison model later, when we consider the robustness of the key elasticity (of trade secret intensity with respect to size).

The dummy variables generally do not add to the explanation of TSI. Thus, the characteristics of the thief, *outsider* and *foreign*, are not significant. As Zwillinger and Grenetski (2000) note, the *Sentencing Guidelines* of the EEA include harsher punishments for those who steal from their employers (*outsider* = 0); however, this control variable does not appear to affect the use of trade secrets.

Additionally, the EEA was designed with the intent to criminalize Economic Espionage which, by its definition, requires the involvement of foreigners. However, *foreign*, as another control variable, again does not appear to influence the use of trade secrets.

Finally, the nature of the trade secrets, as indicated by *patentable* and *copyright*, also fails to influence *TSI*. This result is surprising as the nature of the trade secrets and the availability of alternate methods of protecting the innovation could influence the intensity of the use of trade secrets. Thus, the modelling of these dummy variables in the log-linear regression has resulted in the elimination of additional potential variables that might have seemed to influence *TSI*, but do not.

While the Box-Cox testing for functional form has suggested that the log-linear approach is best, we feel it is nevertheless appropriate to proceed to examine other potential functional forms. The relatively poor performance of these alternative regressions suggests that that the log-linear form is a superior form for our model. For example, estimates of a strictly linear (in variables and in parameters) regression model and quadratic models all have R-squared of less than 0.10. Further, the quadratic term is not significant. Overall, these alternatives did a poor job of predicting the value of *TSI*.

To strengthen the regression results further, we tested for heteroskedasticity. For the equation with sectoral dummies, the Breusch-Pagan test has a p-value of 0.69, so homoskedasticity is not rejected. However, given the small sample size, White's test may be a more appropriate measure. Again, White's test indicates a p-value of 0.71 and again the null hypothesis of homoskedasticity is not rejected. This indicates that neither the sectoral dummy variables, nor the level of sales have significant impact on the error variances of the model.

Concern that the use of trade secrets is correlated with innovation; that is, that innovation can be explained by the use of trade secrets, leads us to test for endogeneity. Baldwin et al (2000, p. 17) make the argument that trade secrets can explain innovation and 'in industries where trade secrets are seen to be effective, the probability that innovation occurs is higher.' Furthermore, a central argument in the Schumpeterian argument is the relationship between size and innovation. Lunn (1986)⁶ argues the analysis of the Schumpeterian (1934, 1942) hypotheses with regards to market structure requires that this relationship be treated endogenously. This gives cause for concern about endogeneity between the size of the firm and the use of trade secrets.

We investigate potential endogeneity in our explanatory variable $\ln(vsales)$ using the Durbin-Wu-Hausman and the Wu-Hausman tests (the DWH tests.) We use our log-linear regression with sectoral dummy variables. The instrumented variable, $\ln(vsales)$, is instrumented by $\ln(xref)$, *outsider* and the sectoral dummies. The tests do not reject the null hypothesis that $\ln(vsales)$ is exogenous (both the DWH tests show a p-value = 0.80.) We can, therefore, proceed without estimators that embrace endogeneity.

6. Robustness

Further analysis via outliers, Cook's (1977) distance, and trimming, tests the robustness of our results. However, one issue that remains unaddressed is error in variables, as it is possible that the data contain measurement errors and that the use of proxies has resulted in errors. The use of our robustness testing

⁶ Lunn (1986), p. 321.

should help to mitigate the influence of such problems such as errors in variables.

Due to the highly heterogeneous estimation of the value of the trade secrets, one outlier has been excluded from the dataset. This observation has been excluded as the estimated value includes significant inputs other than the trade secret itself. This observation is the estimate of the value of source code stolen in the Lucent case⁷ which was valued at \$100M as total sales of the related software. This case is 8.8 standard deviations away from the mean of \$5.6M and 6.1 standard deviations away from its nearest neighbour.⁸ Hence, this case is treated as an outlier and is excluded from the analysis.

Further outlier and trimming analysis confirms the robustness of the regressions performed on the remaining observations. A robust regression investigation of the basic form of the model was used by determining the Cook's (1977) distance of each of the observations. This method detects influential observations in linear regression. Using Cook's distance criteria, five observations were dropped in this robust regression procedure. Nevertheless, our results remained stable, in the face of this. Values of the coefficients remained similar to those found in the regression which used the log-linear model with sectoral dummy variables.

A trimming analysis using a Kernel Density method to trim $\ln(vsales)$ based on the distribution also indicates that the regression results are fairly robust. The trimming analysis, resulting in a cull of up to 20% of observations, reduces the value of R-squared from 0.50 in the complete sample, to 0.36 in the 20% trimmed sample. Additionally, the key coefficient of $\ln(vsales)$, β_1 , is reasonably stable. Thus, it is stable in the sense of being included by the set of values $[-0.62, -0.59]$ under two levels of trimming (10% and 20%); and is also stable in comparison to the values of this parameter under untrimmed analysis (e.g. by comparison to the first row of value in Table 2, excluding that for the model with interactions, in column 5, for which elasticity is calculated in a different way).

⁷ US v. ComTriad et al, 2:01-cr-00365-WHW-3 filed on May 31, 2001 in New Jersey.

⁸ Mean of sample including outlier = \$5.6M with a S.D. of \$12.9M.

An alternate method of trimming is Trimmed Least Squares (LTS), which allows for a specified level of trimming. Performing trims of 10%, 20% and 30% on the equation of column 3, Table 2, again shows β_1 is reasonably stable and includes the set [-0.78, -0.67]. Both methods of trimming analysis confirm the robustness of the results of the regression analysis.

An advantage of the log linear form is the implied constant elasticity (η) of the dependent variable with respect to the independent variables. This measures of the proportional response of the *TSI* with respect to a proportional change in *vsales*. Alternate functional forms of the model, and trimming, as considered above, suggest that the elasticity is fairly stable. Using the usual definition of elasticity $\eta = (\partial y / \partial x)(x/y)$, in the log-linear model η is equal to the value of the coefficient of the relevant variable. Our general conclusion, based on all the estimates reported earlier, is that the value of the elasticity of *TSI* with respect to *vsales* has remained stable, under both alternative specifications and under sample trimming. In the most basic log linear model (column 3, Table 2) the elasticity of *TSI* with respect to *vsales* is $\eta = \beta_1 = -0.67$. If we turn to a more complex functional form, which involves the log-linear model with an interaction term, we find that this elasticity is robust. To illustrate, consider the model of column 5, Table 2:

$$\ln(TSI) = 37.0 - 1.48\ln(vsales) - 1.37\ln(xref) + 0.066\ln(vsales)\ln(xref) + \varepsilon \quad (6)$$

Here the overall fit is highly significant, as are coefficients β_0 and β_1 ; and β_2 and β_3 are also significant at the 10% level. The elasticity of *TSI*, with respect to *vsales* evaluated at the mean of $\ln(xref)$ is $\eta = -1.48 + (0.066 \times -12.18) = -0.68$, which value is very much in line with estimates obtained by different methods, under different sample trimming. This establishes a fairly solid basis for this elasticity as having a (negative) value of about two-thirds. The values of this elasticity obtained under the various forms of regression in Table 2 show considerable stability of η . The robustness checks also find a stable η as the range of the estimations for η is (-0.59, -0.79) under various levels of trimming

noted. Overall, the estimated values of the elasticity of *TSI* with respect to *vsales* remain stable at around $\eta = -0.7$.

The results of this study confirm that the use of trade secrets is negatively related to firm size. Smaller firms are relatively more reliant on trade secrets, and larger firms are relatively less reliant on trade secrets.

7. Conclusion

Empirical investigation into the use of trade secrets remains an under-examined area due to the challenges of data gathering. This study represents a start in beginning to understand the relationship between the use of trade secrets and firm size. Based on a regression analysis of EEA data, we conclude that there is an inverse, convex relationship between firm size and the intensity of trade secrecy. Additionally, it is an inelastic relationship, with an elasticity of -0.7 of *TSI* with respect to *vsales*. Comprehensive testing of the robustness of this relationship, for example, under different functional forms, and different degrees of sample trimming, confirms that this relationship remains stable.

Our results suggest that larger firms prefer patents to trade secrets and that the opposite is true for smaller firms. Given the high costs of both obtaining and maintaining patents, smaller firms may find trade secrets a more efficient method of protecting innovations. Trends in aggressive patent enforcement (as in Lerner and Jaffe, 2004) suggest that the costs of patenting will increase. Future changes in the cost of patenting will create the possibility of empirically testing the assumption that this cost drives smaller firms to use trade secrets. Certainly, there is more work to be done in this area.

The implications that the relationship between the use of trade secrets and firm size has for arguments for and against the Schumpeterian (1934, 1942) hypotheses also merit further scrutiny. Our evidence shows that smaller firms use trade secrets more intensively, extending the similar finding of Arundel (2001), but going beyond the subjective date used there to our use of objective data. As Arundel notes, the Schumpeterian argument supports the concept that smaller firms may produce less patentable innovations and, therefore, are more

reliant on trade secrets. The findings of our study, taken along with those of Arundel (2001) and similar studies, also suggest that empirical studies which use patents as a proxy for innovations will have results that underestimate the innovative activity of smaller firms. As economists shift their focus away from patents and copyright towards alternate methods of protecting innovation, the use of trade secrets will be better understood. Important questions remain unanswered regarding the value of trade secrets, the firm's decision between trade secrets and patents, and the strategic use of trade secrets.

References

- Almeling, D. S., Snyder, D. W., Saponznikow, M., McCollum, W. and Weader, J. (2010) A Statistical Analysis of Trade Secret Litigation in Federal Courts, *Gonzaga Law Review*, available at http://works.bepress.com/david_almeling
- Arundel, A. (2001) The relative effectiveness of patents and secrecy for appropriation, *Research Policy*, 30, 611-24.
- Arundel, A. and Kabla, A. (1998) What percentage of innovations are patented? Empirical estimates for European firms, *Research Policy*, 27, 127-41.
- Baldwin, J., Hanel, P., and Sabourin, D. (2000) Determinants of Innovative Activity in Canadian Manufacturing Firms: The Role of Intellectual Property Rights, *Statistics Canada, Analytical Studies Branch – Research Paper Series*, 11F0019MPE, No. 122.
- Bosworth, D. and Rogers, M. (2001) Market Value, R&D and Intellectual Property: An Empirical Analysis of Large Australian Firm, *The Economic Record*, 77:239, 323-37.
- Box, G.E.P. and Cox, D.R. (1964) An Analysis of Transformations, *Journal of the Royal Statistical Society*, B26 (2), 211-252.
- Carr, C. and Gorman, L. (2001) The Revictimization of Companies by the Stock Market Who Report Trade Secret Theft Under the Economic Espionage Act, *The Business Lawyer*, 57:1, 25-53.
- Cohen, W. M., Nelson, R. R. and Walsh, J. P. (2000) Protecting their Intellectual Assets: Appropriability Condition and Why U.S. Manufacturing Firms Patent (or Not), Paper No. 7552, National Bureau of Economic Research, Washington.
- Cohen, W. M., Levin, R. C., and Mowery, D. C. (1987) Firm Size and R&D Intensity: A Re-examination, *The Journal of Industrial Economics*, 35:4, 543-65.
- Cook, R.D. (1977) Detection of Influential Observations in Linear Regression, *Technometrics*, 19: 1, 15-18.
- Cordes, J. J., Hertzfield, H. R., and Vonortas, N. S. (1999) A Survey of High Technology Firms, *Office of Chief Counsel for Advocacy, United States Small Business Administration*, Contract No. SBA-8141-OA94, 1-71.
- Damanpour, F. (2010) An Integration of Research Findings of Effects of Firm Size and Market Competition on Product and Process Innovations, *British Journal of Management*, 21, 996-1010.
- Friedman, D. D., Landes, W. M., and Posner, R. A. (1991), Some Economics of Trade Secret Law, *The Journal of Economic Perspectives*, 5:1, 62-63.

Jain, V. and Kiran, R. (2012) Technology Management Strategies and Small and Medium Enterprises of Punjab Manufacturing: A Use-based sector analysis, *Journal of Intellectual Property Rights*, 17, 64-72.

Jensen, P. H. and Webster, E. (2006), Firm Size and the Use of Intellectual Property Rights, *The Economic Record*, 82:256, 44-55.

Kohl, H. on Congressional Record: March 30, 2011 (Senate), Page S1978-S1993 available from http://www.fas.org/irp/congress/2011_cr/s678.html

Hall, B., Helmers, C., Rogers, M., and Senia, V. (2011) The choice between formal and informal intellectual property: a review, available at http://www.chelmers.com/projects/IPalternatives_Aug11.pdf

Leiponem, A. and Byman J. (2010) If you cannot block, you better run: Small firms, cooperative innovation, and appropriation strategies, *Research Policy*, 38, 1478-88.

Lerner, J. (1995) Patenting in the Shadow of Competitors, *Journal of Law and Economics*, 38:2, 463-95.

Lerner, J. (2006) Using Litigation to Understand Trade Secrets: A Preliminary Exploration, available at <http://ssrn.com/abstract=922520>

Levin, R. C., Klevorick, A. K., Nelson, R. R., and Winter, S. G. (1988) Appropriating the Returns from Industrial R&D, Discussion Paper No. 862, Cowles Foundation, Yale University, New Haven.

Lunn, J. (1986) An Empirical Analysis of Process and Product Patenting: A Simultaneous Equation Framework, *The Journal of Industrial Economics*, 34:3, 319-30.

Mansfield, E. (1986) Patents and Innovation: An Empirical Study, *Management Science*, 32:2, 173-81.

Mansfield, E., Schwartz, M., and Wagner, S. (1981) Imitation Costs and Patents: An Empirical Study, *The Economic Journal*, 91:364, 907-18.

Png, I.P.L. (2011) Law and Innovation: Evidence from the Uniform Trade Secrets, available at <http://ssrn.com/abstract=1755284>

Reid, G. C. and Ujjual, V. (2008) Firms in Scottish High Technology Clusters, Discussion Paper No. 0804, Centre for Research into Industry, Enterprise, Finance and the Firm, University of St Andrews, St Andrews.

Searle, N.C. (2011) *The Economics of Trade Secrets: Evidence from the Economic Espionage Act*, PhD thesis, University of St Andrews.

Scherer, F.M. (1965) Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions, *The American Economic Review*, 55:5:1, 1097-125.

Scherer, F.M. (1983) The Propensity to Patent, *International Journal of Industrial Organization*, 1, 107-28.

StataCorp (2009), *Stata Statistical Software: Release 11*. College Station, Texas: StataCorp LP.

Willoughby, K. (2010) Intellectual Property Management and Technological Entrepreneurship: One Size Does Not Fit All, available at <http://ssrn.com/abstract=1630666>

Yoon, B. and Lee, S. (2008) Patent Analysis for Technology Forecasting: Sector-specific Applications, Engineering Management Conference, 2008. IEMC Europe 2008. IEEE International

Zwillinger, M. and Genetski C. (2000) 'Calculating Loss Under the Economic Espionage Act of 1996,' *George Mason Law Review*, 9:2, 323-56.