Screening Property Rights for Innovation

Supplemental Appendices

William Matcham and Mark Schankerman

A Additional Tables and Figures

Table A.1. Regression Results

Variable	(1) Negotiation	(2) Grant
GS-7	-0.002 (0.004)	0.003 (0.004)
GS-9	-0.016 (0.004)	0.035 (0.004)
GS-11	-0.020 (0.004)	0.066 (0.004)
GS-12	-0.034 (0.004)	0.092 (0.004)
GS-13	-0.045 (0.004)	0.126 (0.004)
GS-14	-0.091 (0.004)	0.178 (0.004)
Chemicals (17)	0.063 (0.001)	0.067 (0.001)
Comp. Software (21)	0.105 (0.002)	0.196 (0.002)
Comp. Networks (24)	0.123 (0.002)	0.192 (0.002)
Communications (26)	0.047 (0.002)	0.198 (0.002)
Electronics (28)	-0.010 (0.001)	0.244 (0.001)
Other (36)	0.065 (0.001)	0.136 (0.001)
Engineering (37)	0.042 (0.001)	0.139 (0.001)
Small Entity	-0.120 (0.001)	-0.169 (0.001)
Year Dummies	Yes	Yes
N	1,641,333	1,759,313

Notes: Column (1) shows estimates from a regression of a binary variable equal to one if the application process lasts more than one round, against dummies for examiner seniority grade, technology center, year of application, and a small entity indicator (applying firm having fewer than 500 employees). Column (2) is the same except the dependent variable is equal to one if the examiner grants a patent. The omitted grade is GS-5, and the omitted technology center is Biotechnology and Organic Fields (16). Technology center "Other" refers to Center 3600, which is "Transportation, Electronic Commerce, Construction, Agriculture, Licensing and Review." Following Frakes and Wasserman (2017), we omit GS-15 examiners. Standard errors are clustered at the examiner level.

Table A.2. Estimated and Assigned Parameters

Estimated Parameters			
Variable	**Notation**	**Distribution**	**Parameters**
Examiner			
Intrinsic motivation	$\theta \sim G_{S,\theta}(\cdot)$	Log-normal	σ_θ, μ_θ^J or μ_θ^S
Error	$\varepsilon \sim G_{e,\varepsilon}(\cdot)$	Normal	σ_ε
Applicant			
Initial claim returns	$v_j^* \sim G_v(\cdot)$	Log-normal	μ_v, σ_v
Initial claim distances	$D_j^* \sim G_D(\cdot)$	Multivariate Beta	α_D, γ_D, ρ
Obsolescence	$\boldsymbol{\omega}$	Bernoulli	P_ω^{pre} or P_ω^{post}
Application legal costs	f_{app}	Log-normal	$\boldsymbol{\mu}_{f,\text{app}}, \boldsymbol{\sigma}_{f,\text{app}}$
Issuance legal costs	f_{iss}	Log-normal	$\boldsymbol{\mu}_{f,\text{iss}}, \boldsymbol{\sigma}_{f,\text{iss}}$
Maintenance legal costs	f_{main}	Log-normal	$\boldsymbol{\mu}_{f,\text{main}}, \boldsymbol{\sigma}_{f,\text{main}}$
Amendment legal costs	f_{amend}	Log-normal	$\boldsymbol{\mu}_{f,\text{amend}}, \boldsymbol{\sigma}_{f,\text{amend}}$
Narrowing	η	-	-
Threshold by technology center	τ_T		Range from 0.47 to 0.50

Assigned Parameters		
Variable	**Notation**	**Values**
Discount rate	β	0.95
Depreciation	δ	$\frac{0.14-P_\omega^{\text{post}}}{1-P_\omega^{\text{post}}}$
Credits	$g_y^r(S,T)$	See Appendix D
Finalizing fee	ϕ	\$1,770
RCE fees	$F_{\text{round}}^3 = F_{\text{round}}^5$	\$930
	F_{renew}^4	\$1,600
Renewal fees	F_{renew}^8	\$3,600
	F_{renew}^{12}	\$7,400

Notes: To generate correlated multivariate Beta draws for unpadded distances, we draw a vector of size M_0 from a standard multivariate normal with correlation coefficient ρ. We apply the quantile function of the normal to the draws to create correlated uniform random variables. Then for the estimation initial values $(\tilde{\alpha}_D, \tilde{\beta}_D)$, we apply the inverse CDF of a Beta distribution with these parameters to the uniform draws to generate correlated beta distributed initial distances. For ρ, we use the empirical correlation of granted distances. Simulations confirm that the correlation of the multivariate copula is very close to the correlation of the distances. See Nelsen (2007) for details.

Table A.3. Estimated Attorney Costs by Technology Area (Application)

Parameter	Symbol	Estimate	S.E.
Chemical application fighting cost log-mean	μ_f^{chem}	9.14	0.003
Chemical application fighting cost log-sigma	σ_f^{chem}	0.44	0.009
Electrical application fighting cost log-mean	μ_f^{elec}	9.20	0.002
Electrical application fighting cost log-sigma	σ_f^{elec}	0.11	0.002
Mechanical application fighting cost log-mean	μ_f^{mech}	8.98	0.003
Mechanical application fighting cost log-sigma	σ_f^{mech}	0.75	0.006

Notes: Standard errors are bootstrapped.

Table A.4. Estimated Attorney Costs by Technology Area (Other)

Parameter	Symbol	Estimate
Simple amendment fighting cost log-mean	$\mu_{f,\text{amend}}^{\text{simple}}$	7.60
Simple amendment fighting cost log-sigma	$\sigma_{f,\text{amend}}^{\text{simple}}$	0.37
Chemical amendment fighting cost log-mean	$\mu_{f,\text{amend}}^{\text{chem}}$	8.13
Chemical amendment fighting cost log-sigma	$\sigma_{f,\text{amend}}^{\text{chem}}$	0.45
Electrical amendment fighting cost log-mean	$\mu_{f,\text{amend}}^{\text{elec}}$	8.07
Electrical amendment fighting cost log-sigma	$\sigma_{f,\text{amend}}^{\text{elec}}$	0.38
Mechanical amendment fighting cost log-mean	$\mu_{f,\text{amend}}^{\text{mech}}$	7.95
Mechanical amendment fighting cost log-sigma	$\sigma_{f,\text{amend}}^{\text{mech}}$	0.43
Issuance cost log-mean	$\mu_{f,\text{iss}}$	6.55
Issuance cost log-sigma	$\sigma_{f,\text{iss}}$	0.62
Maintenance cost log-mean	$\mu_{f,\text{main}}$	5.67
Maintenance cost log-sigma	$\sigma_{f,\text{main}}$	0.46

Notes: Standard errors are not included here, since we only observe the fighting costs moments for the external GMM estimation rather than the underlying data.

Table A.5. Robustness of Model Estimates

Symbol	(1) Baseline	(2) Distance FF	(3) 1% τ	(4) Unpurged	(5) η Round	(6) η Seniority	(7) IM FF (i)	(8) IM FF (ii)	(9) π Round	(10) σ_ε Seniority
μ_v	10.60	9.81	10.70	11.07	10.41	11.48	10.89	11.02	10.14	10.47
σ_v	0.86	1.29	0.90	1.19	0.94	0.53	2.31	0.55	0.58	1.07
P_ω^{pre}	0.13	0.14	0.16	0.14	0.16	0.13	0.14	0.16	0.13	0.12
P_ω^{post}	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
α_D	3.96	7.00	4.00	4.10	4.35	3.35	4.41	4.74	3.74	4.28
γ_D	7.71	6.34	7.64	8.01	8.34	6.62	8.34	8.76	7.46	8.18
μ_f^{simple}	8.55	8.66	8.49	8.73	8.68	8.65	8.56	8.45	8.38	8.64
σ_f^{simple}	0.88	0.77	0.87	0.38	0.77	0.62	0.81	0.97	1.11	0.87
μ_θ^J	3.81	3.77	3.50	3.85	3.94	4.36	3.74	3.71	3.69	3.61
μ_θ^S	2.77	2.98	2.55	2.94	3.05	3.17	2.74	2.98	2.69	2.98
σ_θ	0.98	0.95	0.95	1.12	1.09	1.27	0.98	1.11	0.95	0.98
η	0.27	0.28	0.29	0.28	η^1 =0.29 η^{2+} =0.22	η^J=0.26 η^S=0.32	0.27	0.29	0.27	0.28
π	1.01	2.38	0.48	0.90	1.08	0.37	1.04	0.83	π^{INIT} =0.73 π^{RCE} =1.54	0.50
σ_ε	0.18	0.16	0.20	0.22	0.20	0.10	0.21	0.23	0.18	σ_ε^J =0.23 σ_ε^S=0.32
ϑ	-	1.79	-	-	-	-	-	-	-	-
ς	-	-	-	-	-	-	0.89	-	-	-

Notes: This table provides estimates of the model parameters across various model alternatives. Column (1) reproduces the baseline estimates; (2) adjusts the functional form (FF) for padded distance to $\tilde{D}_j = (D_j^*)^\vartheta p^{-1}$; (3) uses the 1% of each examiner's distance granted for the threshold estimator; (4) uses unpurged distances; (5) allows narrowing to vary by rounds 1 and 2+; (6) allows narrowing to vary by seniority; (7) adjusts the functional form for intrinsic motivation cost to $\mathcal{R}(M_r, \theta) = \theta \left(\frac{M_r}{M_0}\right)^\varsigma$; (8) adjusts it to $\mathcal{R}(M_r, \theta) = \theta M_r$; (9) allows delay cost to differ in RCEs; (10) allows error variance parameter to vary by seniority.

TABLE A.6. NET SOCIAL COSTS OF PATENT PROSECUTION: ROBUSTNESS

	Patent Premium (Ψ) = 0.05							Patent Premium (Ψ) = 0.025						
Counterfactual	$\boldsymbol{T_1}$ **(1.5)**	$\boldsymbol{T_1}$ **(2.0)**	$\boldsymbol{T_2}$ **(1.5)**	$\boldsymbol{T_2}$ **(2.0)**	$\boldsymbol{T_3}$	**Σ (1.5)**	**Σ (2.0)**	$\boldsymbol{T_1}$ **(1.5)**	$\boldsymbol{T_1}$ **(2.0)**	$\boldsymbol{T_2}$ **(1.5)**	$\boldsymbol{T_2}$ **(2.0)**	$\boldsymbol{T_3}$	**Σ (1.5)**	**Σ (2.0)**
Baseline ($Bn)	4.8	4.8	0.7	1.3	18.6	24.1	24.7	5.3	5.3	0.2	0.4	21.0	26.5	26.7
50K Round Fee	4.4	4.3	1.1	2.1	15.9	21.3	22.3	4.9	4.9	0.3	0.7	17.2	22.4	22.8
Three Rounds	4.0	3.8	1.9	4.0	13.7	19.6	21.4	4.4	4.4	1.0	2.1	14.9	20.4	21.4
Two Rounds	3.0	2.9	3.5	7.3	8.8	15.3	19.0	3.2	3.2	0.5	1.1	9.5	13.2	13.8
One Round	0.6	0.6	3.1	6.5	3.6	7.2	10.6	0.7	0.7	0.4	0.8	3.8	4.8	5.3
15% IM	19.2	17.7	0.8	1.7	18.9	38.9	38.3	20.9	19.6	0.4	0.8	21.1	42.4	41.4
Credit↘	4.5	4.4	0.7	1.3	18.6	23.8	24.4	4.6	4.5	0.2	0.5	21.0	25.8	25.9
Credit↘ + 15% IM	11.6	3.4	1.4	3.0	20.6	33.6	27.0	15.9	9.4	0.9	1.9	22.9	39.6	34.2

Notes: This table provides the values of net social costs for alternative values of the patent premium and social multiplier. Columns denoted $\boldsymbol{T_j}$ **(1.5)** and $\boldsymbol{T_j}$ **(2.0)** provide values of type j net social costs when $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}$ is equal to 1.5 and 2.0, respectively. Columns Σ **(1.5)** and Σ **(2.0)** provide the total net social costs when $\frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}$ is equal to 1.5 and 2.0, respectively.

TABLE A.7. COUNTERFACTUAL CONFIDENCE INTERVALS

Counterfactual	**Not Apply** (%)	**Pad** (%)	**R1 Gr** (%)	$\tilde{v}_j$	**T1 Err** (%)	**T1 Egr** (%)	**T2 Err** (%)	**T2 Egr** (%)	$\boldsymbol{T_1}$ **Cost**	$\boldsymbol{T_2}$ **Cost**	$\boldsymbol{T_3}$ **Cost**
Baseline	[4.82, 6.36]	[5.31, 5.95]	[10.82, 11.56]	[42.95, 53.19]	[17.40, 17.89]	[4.20, 4.29]	[37.48, 39.74]	[17.42, 18.59]	[4.28, 4.91]	[0.26, 2.27]	[17.48, 19.00]
50K Round Fee	[14.64, 18.39]	[0.70, 1.39]	[15.76, 17.69]	[45.60, 55.60]	[16.71, 16.96]	[4.22, 4.34]	[42.59, 45.28]	[21.64, 23.30]	[3.74, 4.59]	[1.17, 5.99]	[14.39, 16.25]
Three Rounds	[23.79, 24.24]	[0.33, 0.66]	[13.86, 14.61]	[46.00, 57.10]	[16.64, 17.23]	[4.06, 4.17]	[46.76, 49.16]	[22.82, 24.35]	[3.28, 3.94]	[1.85, 8.98]	[12.98, 13.84]
Two Rounds	[45.49, 45.93]	[-3.29, -3.09]	[23.34, 24.23]	[49.57, 61.69]	[14.27, 14.60]	[4.23, 4.40]	[50.61, 53.12]	[30.44, 32.20]	[2.20, 3.14]	[5.79, 11.41]	[8.41, 8.91]
One Round	[73.87, 74.37]	[-8.31, -8.23]	[88.73, 88.94]	[53.65, 67.09]	[5.23, 5.55]	[1.47, 1.56]	[74.91, 75.94]	[55.79, 56.44]	[0.46, 0.61]	[4.74, 10.09]	[3.44, 3.61]
15% IM	[2.22, 3.04]	[10.34, 11.20]	[31.47, 32.44]	[50.89, 63.25]	[80.01, 80.50]	[52.61, 53.21]	[21.68, 23.22]	[10.33, 11.23]	[13.98, 19.04]	[0.51, 2.68]	[17.95, 19.29]
Credit↘	[4.77, 6.28]	[5.07, 5.79]	[10.81, 11.61]	[42.92, 53.16]	[16.95, 17.42]	[3.87, 3.95]	[37.63, 39.75]	[17.74, 18.96]	[3.81, 4.75]	[0.01, 2.27]	[17.48, 18.99]
Credit↘ + 15% IM	[1.60, 2.25]	[29.94, 32.99]	[32.55, 34.12]	[59.19, 74.30]	[81.23, 81.63]	[52.52, 53.72]	[14.57, 15.86]	[5.82, 6.65]	[1.49, 6.14]	[1.24, 3.86]	[19.51, 21.00]

Notes: This table provides 95% percentile bootstrapped confidence intervals for outcomes and social costs across counterfactual scenarios. See Tables 3 and 4 for description of columns and rows.

Figure A.1. Distribution of Padded Granted Distances

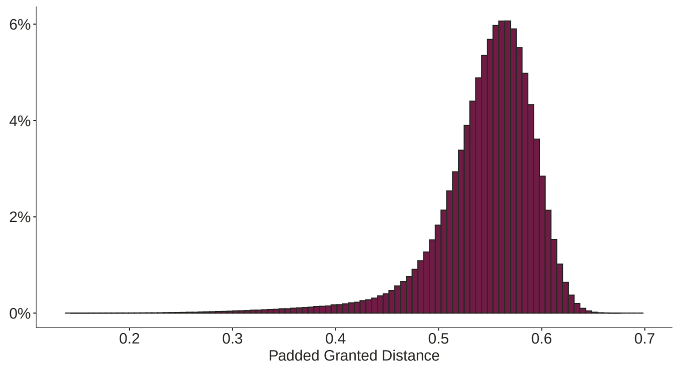
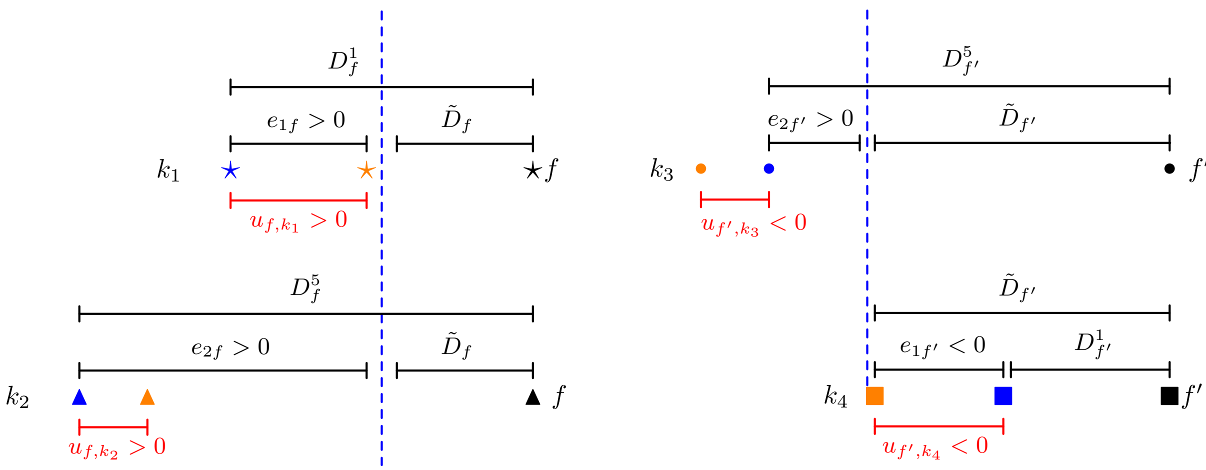


Figure A.2. Algorithmic Errors and Latent Variable Model



(A) Subfigure 1 (B) Subfigure 2

Notes: The figures depict the variables of the latent variable model ($\tilde{D}, D^1, D^5, e_1$, and e_2), and the algorithmic error, denoted u, for two focal claims f and f'. The aim is to illustrate that positive correlation between the *algorithmic* errors need not imply any particular correlation in *latent variable* errors e_1 and e_2. Black color depicts the focal claim location, orange color the true position of a claim, and blue color the algorithm's measurement of the position of a claim. For the sake of illustration, assume that focal claims have no algorithmic error, the latent model has no controls, and $\xi_2 = 1$, so that $D_f^1 = \tilde{D}_f + e_{1f}$ and $D_f^5 = \tilde{D}_f + e_{2f}$.

In Subfigure 1, claim f (shown by the black star and the black triangle) is the focal claim whose distance to prior art we want to measure. Claim k_1 is the closest, and k_2 the fifth closest in reality. The distance between the black and orange stars represents $\tilde{D}_f$. The distance between the orange and blue stars is the positive algorithmic error in measuring the distance between f and k_1. The algorithm measures the distance between f and k_1 as the distance between the black and blue stars. This length is shorter than the measured distance between the black and blue triangles, so, the ordering is correct. In Subfigure 1, the algorithmic errors u_{f,k_1} and f_{f,k_2} are positive, and so are e_{1f} and e_{2f}. In Subfigure 2, while both algorithmic errors f_{f',k_3} and f_{f',k_4} are negative, $e_{1f'}$ is negative whereas $e_{2f'}$ is positive. Hence, in Subfigure 1, the signs of the algorithmic errors match those of e_1 and e_2, but in Subfigure 2 the signs do not match.

Figure A.3. SMM Objective Value Plots

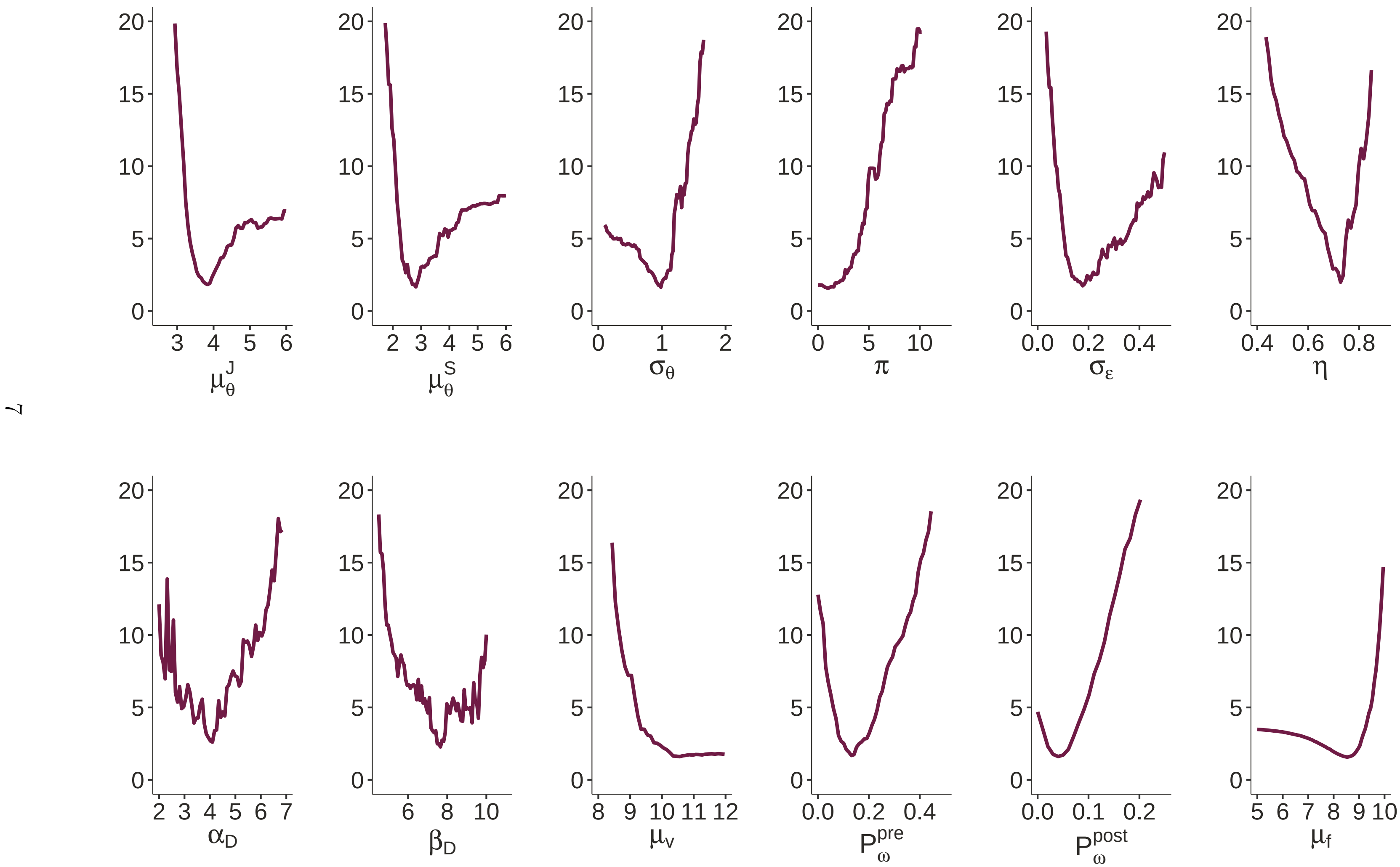


Figure A.4. Match of internal data and model moments

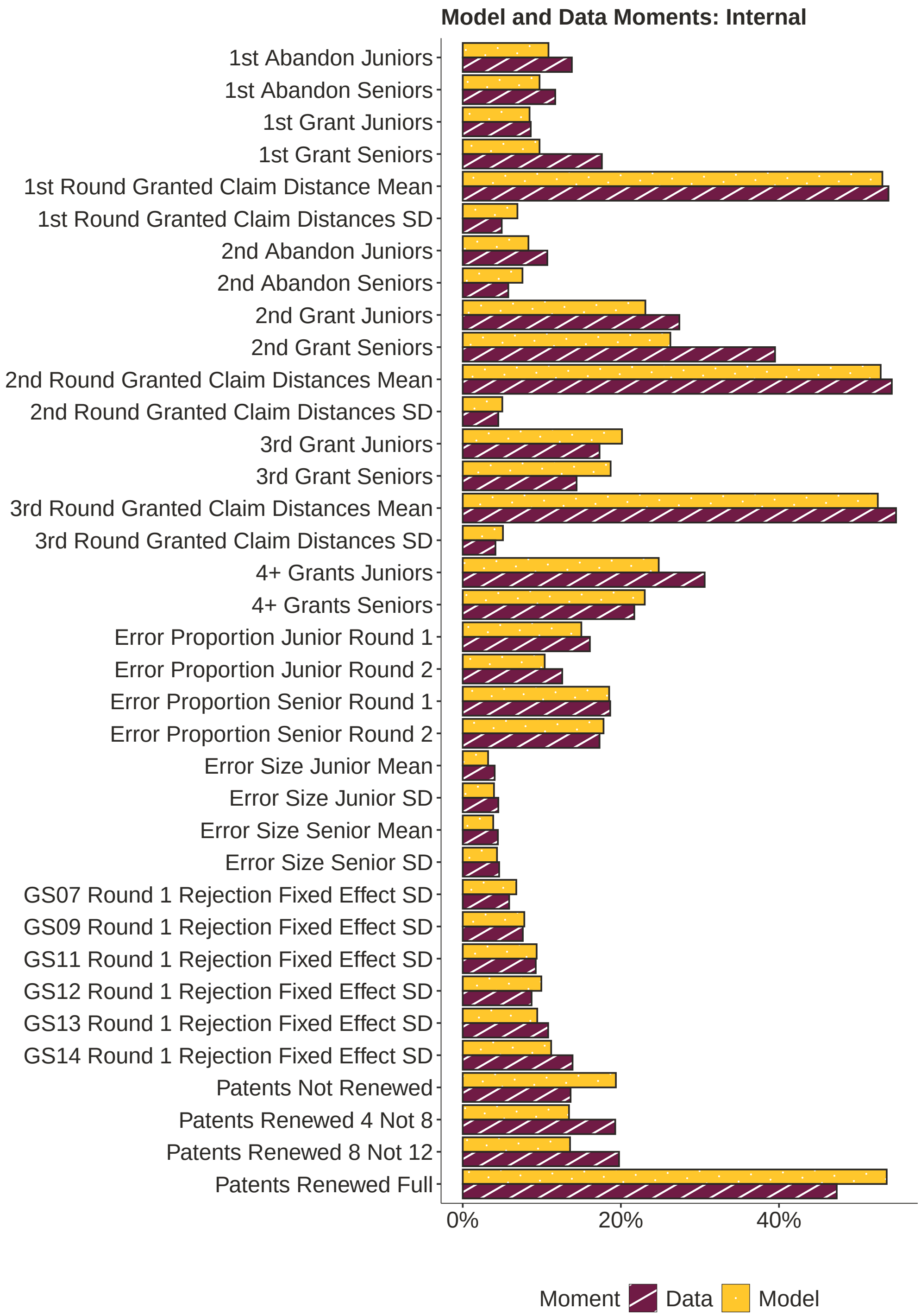
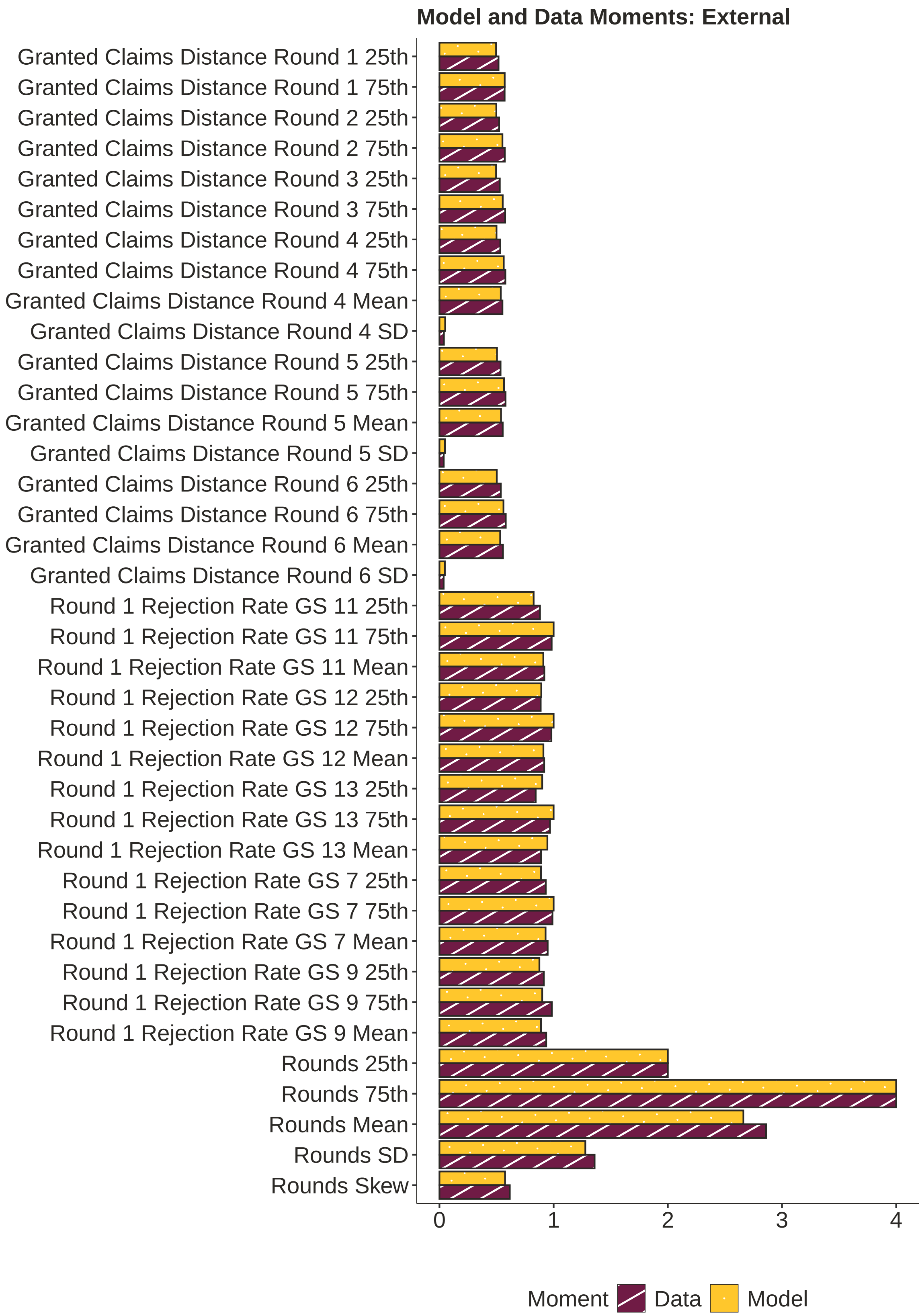


FIGURE A.5. MATCH OF EXTERNAL DATA AND MODEL MOMENTS



B Proofs

B.1 Proposition 1

Proof. $\Longrightarrow$: Suppose that Condition 1 holds. Take $V(p, v_j^*) = \tilde{v}_L(p, v_j^*, \mathbf{0}) = \mathcal{V}(p, v_j^*) = \tilde{v}_j^1$ and $D(p, D_j^*, \varepsilon) = \hat{D}_L(p, D_j^*, \varepsilon, \mathbf{0}) = \mathcal{D}(p, D_j^*, \varepsilon) = \hat{D}_j^1$. Then, by the condition we assume, we have $\tilde{v}_L(p, v_j^*, \boldsymbol{\eta}) = W_v(V(p, v_j^*), \boldsymbol{\eta})$ along with $\hat{D}_L(p, D_j^*, \varepsilon, \boldsymbol{\eta}) = W_D(D(p, D_j^*, \varepsilon), \boldsymbol{\eta})$ implying we can take $\tilde{v}_C = W_v$ and $\hat{D}_C = W_D$.

It remains to show that $\tilde{v}_C$ and $\hat{D}_C$ are increasing in their first argument at $\boldsymbol{\eta} = \mathbf{0}$. Note that for all $\tilde{v}_j^1$, $\tilde{v}_j^1 = \tilde{v}_C(\tilde{v}_j^1, \mathbf{0})$. Therefore, for $\tilde{v}_{j,1}^1 < \tilde{v}_{j,2}^1$,

$$\tilde{v}_C(\tilde{v}_{j,1}^1, \mathbf{0}) = \tilde{v}_{j,1}^1 < \tilde{v}_{j,2}^1 = \tilde{v}_C(\tilde{v}_{j,2}^1, \mathbf{0})$$

therefore $\tilde{v}_C$ is increasing in its first argument at $\boldsymbol{\eta} = \mathbf{0}$. The same argument works for $\hat{D}_C$. This completes the necessity of being able to separate the terms (p, v_j^*) and $\boldsymbol{\eta}$ in the arguments of $\tilde{v}_L$ (similarly for $\hat{D}_L$). Next, we move to the sufficiency.

$\Longleftarrow$: Assume that (2) and (3) hold. We need to show that Condition 1 holds. Note that by definition, $\tilde{v}_j^1 = \tilde{v}_C(V(p, v_j^*), \mathbf{0})$ and $\hat{D}_j^1 = \hat{D}_C(D(p, D_j^*, \varepsilon), \mathbf{0})$. Because both $\tilde{v}_C$ and $\hat{D}_C$ are strictly increasing at $\boldsymbol{\eta} = \mathbf{0}$, by the implicit function theorem, we can write $V(p, v_j^*) = \tilde{v}_C^{-1}(\tilde{v}_j^1, \mathbf{0})$ and $D(p, D_j^*, \varepsilon) = \hat{D}_C^{-1}(\hat{D}_j^1, \mathbf{0})$. Plugging in to (2) and (3) respectively, we obtain

$$\tilde{v}_L(p, v_j^*, \boldsymbol{\eta}) = \tilde{v}_C(\tilde{v}_C^{-1}(\tilde{v}_j^1, \mathbf{0}), \boldsymbol{\eta}) \equiv W_v(\tilde{v}_j^1, \boldsymbol{\eta}) \tag{B.1}$$

and

$$\hat{D}_L(p, D_j^*, \varepsilon, \boldsymbol{\eta}) = \hat{D}_C(\hat{D}_C^{-1}(\hat{D}_j^1, \mathbf{0}), \boldsymbol{\eta}) \equiv W_D(\hat{D}_j^1, \boldsymbol{\eta}) \tag{B.2}$$

as required. The first equalities in (B.1) and (B.2) show directly that the examiner does not need to form or update beliefs on p, v_j^* and D_j^*: they have a way to map initial values (respectively, distance assessments) and narrowing into future padded values (respectively, distance assessments), without any knowledge of p, v_j^*, D_j^*. □

B.2 Proposition 2

Proof. To start, we state and prove a lemma that will be used in the proof of Proposition 2.

Lemma 1. *Suppose condition 2.1 in Proposition 2 holds. Then, for an examiner with sufficiently large intrinsic motivation, $\hat{D}_j \geq \tau$ for all j granted, that is, the examiner will never grant a claim with an assessed distance below the threshold.*

Proof. In round r, an examiner will refuse to grant a patent to an application with a claim below the threshold (i.e., an application with $\mathcal{R}^r > 0$) if

$$g_{GR}^r - \theta\mathcal{R}^r < g_{REJ}^r + \mathbb{E}(W_e^r)$$

where we have dropped the (S,T) terms on credits for ease of notation. We show that if θ is sufficiently large, this inequality holds when replacing $\mathbb{E}(W_e^r)$ with W_e^r, *for all* realizations of W_e^r. This ensures that the inequality will hold with the expected value of W_e^r, as required.

The realizations of W_e^r depend on the terminal round of the application, either through obsolescence, in which case we have abandonment, or from choices to abandon/grant. When the terminal round is $r+s$ for $s \geq 1$ there are two inequalities to consider. In the case of grant in round $r+s$, the inequality is

$$\theta > \frac{-\left(-g_{GR}^r + \beta^s g_{GR}^{r+s} + \sum_{s'=0}^{s-1} \beta^{s'} \left[g_{REJ}^{r+s'} + g_{FIGHT}^{r+s'} - \beta\pi\right]\right)}{\mathcal{R}^r - \beta^s\mathcal{R}^{r+s}}$$

and in the case of abandonment in round $r+s$ the inequality is[1]

$$\theta > \frac{-\left(-g_{GR}^r + \beta^s(g_{REJ}^{r+s} + g_{ABN}^{r+s}) + \sum_{s'=0}^{s-1} \beta^{s'} \left[g_{REJ}^{r+s'} + g_{FIGHT}^{r+s'} - \beta\pi\right]\right)}{\mathcal{R}^r}$$

Both will hold for sufficiently intrinsically motivated examiners. For the denominators, by condition 2.1, $\mathcal{R}^r$ cannot be smaller than $\bar{M}^{-1}$ and for all r, s, we have $\mathcal{R}^r - \beta^s\mathcal{R}^{r+s}$ is positive and bounded, because $\beta < 1$ and by narrowing, and $\mathcal{R}^r \geq \mathcal{R}^{r+s}$ for all $s > 0$. The numerators are either negative, in which case the inequality holds for all θ; otherwise, the numerators are positive but bounded.

Therefore, for a sufficiently motivated examiner, the key inequality holds for all realizations of W_e^r and thus for $\mathbb{E}(W_e^r)$, as required. The intuition is that if the examiner is sufficiently intrinsically motivated, and they are looking at an application with claims they believe invalid ($\mathcal{R} > 0$), it is always better for them to wait for a future round, where $\mathcal{R}$ will fall, potentially to zero. □

Next, to prove the proposition, we reformulate $\hat{\tau}$ in a way that lends itself to the appropriate asymptotic analysis. Note that examiner e's minimum padded distance across all claims they grant can be written as the minimum, across granted patents $a = 1, \ldots, A_e$ that are examined by examiner e, of the minimum padded distance of the granted claims on patent a. The latter quantity just described is given by $\min_{j=1,\ldots,M_a^{GR}} \tilde{D}_j$, where M_a^{GR} is the number of claims granted

[1] The case for abandonment in round r is covered by taking $s = 0$, in which case the latter summation is empty.

on patent a. Hence,

$$\tau_e = \min_{j \in M_e^{GR}} \tilde{D}_j = \min_{a=1,\ldots,A_e} \min_{j=1,\ldots,M_a^{GR}} \tilde{D}_j \tag{B.3}$$

As mentioned in the main text, we focus on the case of $A_e = A$ for all e. To prove consistency, we must show that for every $\vartheta > 0$, $\mathbb{P}(|\max_e \tau_e - \tau| > \vartheta) \underset{A\to\infty}{\longrightarrow} 0$. Since

$$\mathbb{P}(|\max_e \tau_e - \tau| > \vartheta) \leq \underbrace{\mathbb{P}(\max_e \tau_e > \tau + \vartheta)}_{\mathcal{A}} + \underbrace{\mathbb{P}(\max_e \tau_e < \tau - \vartheta)}_{\mathcal{B}},$$

it suffices to show that $\mathcal{A}$ and $\mathcal{B}$ converge to 0. For the first, note that

$$\mathcal{A} = \mathbb{P}\left(\bigcup_{e=1}^{E} (\tau_e > \tau + \vartheta)\right) \leq \sum_{e=1}^{E} \mathbb{P}(\tau_e > \tau + \vartheta). \tag{B.4}$$

Now, using Equation (B.3) and the fact that $\min_{j=1,\ldots,M_a^{GR}} \tilde{D}_j$ is an independent random variable across granted applications, we have that[2]

$$\mathbb{P}\left(\tau_e > \tau + \vartheta\right) \quad = \quad \prod_{a=1}^{A} [G_{1a}(\tau + \vartheta)]$$

where $G_{1a}(\tau + \vartheta) = \mathbb{P}\left(\min_{j=1,\ldots,M_a^{GR}} \tilde{D}_j > \tau + \vartheta\right)$. To complete the argument, we explain why $G_{1a}(\tau + \vartheta) < 1$ for all examiners e and applications a, from which it follows that $\prod_{a=1}^{A} [G_{1a}(\tau + \vartheta)] \underset{A\to\infty}{\longrightarrow} 0$ and thus $\mathcal{A} \underset{A\to\infty}{\longrightarrow} 0$.

For an examiner with infinite intrinsic motivation, their assessment is equal to $\tilde{D}_j$ because they do not make search errors. By the continuity of $\tilde{D}_j$ (ensured by the continuity of D_j^*), there is non-zero probability that they receive and then grant a claim $\tilde{D}_j \in (\tau, \tau + \vartheta)$ and thus $G_{1a}(\tau + \vartheta) < 1$. For an examiner with finite intrinsic motivation, $G_{1a}(\tau + \vartheta)$ is strictly less than 1 because $\tilde{D}_j = \hat{D}_j / \varepsilon_a$, implying that the middle term satisfies

$$\begin{aligned}
\mathbb{P}\left(\min_{j=1,\ldots,M_a^{GR}} \frac{\hat{D}_j}{\varepsilon_a} > \tau + \vartheta\right) &= \mathbb{P}\left(\frac{\hat{D}_j}{\varepsilon_a} > \tau + \vartheta \quad \forall j\right) \leq \mathbb{P}\left(\frac{1}{\varepsilon_a} > \tau + \vartheta \quad \forall j\right) \\
&= \mathbb{P}\left(\varepsilon_a < \frac{1}{\tau + \vartheta}\right) = \Phi\left(\frac{(\tau + \vartheta)^{-1} - 1 - \mu}{\sigma}\right) < 1
\end{aligned}$$

The middle inequality here relies on the fact that $\hat{D}_j \leq 1$ for all j. Now, we show that $\mathcal{B}$ converges to 0. Consider the examiner meeting condition in Equation (7) in the text, and denote them by e^*. Then,

$$\mathcal{B} = \mathbb{P}(\tau_e < \tau - \vartheta, \, \forall e) \leq \mathbb{P}(\tau_{e^*} < \tau - \vartheta).$$

[2] $\min_{j=1,\ldots,M_a^{GR}} \tilde{D}_j$ is independent across examinations because examiner errors, padding, and unpadded distances are independent across examinations.

Note that, the minimum of $\tilde{D}_j$ among $j \in M_{e^*}^{GR}$ is strictly less than $\tau - \vartheta$ if and only if there exists $j \in M_{e^*}^{GR}$ such that $\tilde{D}_j < \tau - \vartheta$. Hence

$$\mathcal{B} \leq \mathbb{P}\left(\min_{j \in M_{e^*}^{GR}} \tilde{D}_j < \tau - \vartheta\right) = \mathbb{P}\left(\bigcup_{a=1}^{A} \bigcup_{j=1}^{M_a^{GR}} \tilde{D}_j < \tau - \vartheta\right) \leq \sum_a \sum_j \mathbb{P}\left(\tilde{D}_j < \tau - \vartheta\right)$$

Note first that the case of $\vartheta \geq \tau$ is not of interest as $\tilde{D}_j$ cannot be negative. Hence, we focus on the case of $0 < \vartheta < \tau$. Now, consider a constant $H \in (0, \frac{\vartheta}{\tau - \vartheta})$, for which it holds that $\tau_l \equiv (\tau - \vartheta)(1 + H) < \tau$. Then for claim j on application a, $\tilde{D}_j = \frac{\hat{D}_j}{\varepsilon_a} < \tau - \vartheta$ implies that either $\hat{D}_j < \tau_l$ or $\varepsilon_a > 1 + H$. This is because otherwise we would have $\hat{D}_j \geq \tau_l$ and $\dfrac{1}{\varepsilon_a} \geq \dfrac{1}{1+H}$ which together would imply

$$\tilde{D}_j = \frac{\hat{D}_j}{\varepsilon_a} \geq \frac{\tau_l}{1+H} = \frac{(\tau - \vartheta)(1+H)}{1+H} = \tau - \vartheta,$$

a contradiction. Hence,

$$\begin{aligned}
\mathcal{B} &\leq \sum_{a=1}^{A} \sum_{j=1}^{M_a^{GR}} \mathbb{P}\left(\hat{D}_j < \tau_l \cup \varepsilon_a > 1 + H\right) \leq \sum_{a=1}^{A} \sum_{j=1}^{M_a^{GR}} \mathbb{P}\left(\hat{D}_j < \tau_l\right) + \mathbb{P}\left(\varepsilon_a > 1 + H\right) \\
&= \underbrace{\sum_{a=1}^{A} \sum_{j=1}^{M_a^{GR}} \mathbb{P}\left(\hat{D}_j < \tau_l\right)}_{\mathcal{C}} + \underbrace{\sum_{a=1}^{A} \sum_{j=1}^{M_a^{GR}} \mathbb{P}\left(\varepsilon_a > 1 + H\right)}_{\mathcal{F}}
\end{aligned}$$

Regarding $\mathcal{C}$, by Lemma 1, since $\tau_l = (\tau - \vartheta)(1 + H) < \tau$, we have that $\mathbb{P}\left(\hat{D}_j < \tau_l\right) = 0$ for all j granted, so $\mathcal{C} = 0$ for this examiner. For $\mathcal{F}$, since $\varepsilon_a \sim \mathcal{N}(1 + \mu, \sigma^2)$ after standardizing, we have

$$\begin{aligned}
\mathcal{F} &= \sum_{a=1}^{A} \sum_{j=1}^{M_a^{GR}} \left[1 - \Phi\left(\frac{H - \mu(\theta)}{\sigma(\theta)}\right)\right] = \sum_{a=1}^{A} M_a^{GR} \left[1 - \Phi\left(\frac{H - \mu(\theta)}{\sigma(\theta)}\right)\right] \\
&\leq \sum_{a=1}^{A} \bar{M} \left[1 - \Phi\left(\frac{H - \mu(\theta)}{\sigma(\theta)}\right)\right] = A\bar{M} \left[1 - \Phi\left(\frac{H - \mu(\theta)}{\sigma(\theta)}\right)\right]
\end{aligned}$$

Where the penultimate inequality follows from the fact that $M_a^{GR} \leq \bar{M}$ for all a. The proof is completed by noting that, for any A, the final term can be made arbitrarily small for an examiner as given by the condition in Equation (7) in the text. □

C Microfoundation of Examiner Search

An examiner with intrinsic motivation θ chooses time spent searching prior art, denoted $\mathcal{T}$. From their search, the examiner makes an error ε, denominated as a proportional error in interpreting claim distance (i.e. $\varepsilon = 1.2$ means a 20% overestimation of claim distance). The distribution of ε is $N(1+\mu(\mathcal{T}), \sigma^2(\mathcal{T}))$, where $\mu(\mathcal{T})$ is positive and decreasing in $\mathcal{T}$. The intuition for the *mean* decreasing in $\mathcal{T}$ towards one is that for any finite time spent reading, the examiner may miss relevant content, and more time reading reveals more relevant content. However, the examiner may misinterpret what they read, which can lead to realizations of errors below one. We also assume that the variance is decreasing in $\mathcal{T}$.

The examiner pays a search cost $c(\mathcal{T})$, increasing and convex in time spent $\mathcal{T}$.[3] The examiner wants to minimize their mean-squared error of search, defined as $\mathcal{E} = E\left[(\varepsilon-1)^2\right]$. Examiners with higher intrinsic motivation experience greater utility costs from search errors, so they minimize $\mathcal{H}(\theta, \mathcal{E})$, which is increasing in both arguments with positive cross derivative. For simplicity, we specify this as $f(\theta) \times \mathcal{E}$. Hence, the examiner solves

$$\min_{\mathcal{T}} f(\theta) \int (\varepsilon-1)^2 \phi\left(\frac{\varepsilon-\mu}{\sigma}\right) d\varepsilon + c(\mathcal{T})$$

The first order condition is

$$c'(\mathcal{T}) - f(\theta) \underbrace{\int (\varepsilon-1)^2 \left(\frac{\varepsilon-\mu}{\sigma}\right) \phi\left(\frac{\varepsilon-\mu}{\sigma}\right) Z(\mathcal{T}, \varepsilon) d\varepsilon}_{X(\mathcal{T})} = 0$$

where $Z(\mathcal{T}, \varepsilon) = \frac{d}{d\mathcal{T}}\left(\frac{\varepsilon-\mu}{\sigma}\right)$. The second order condition requires $c''(\mathcal{T}) - f(\theta)X'(\mathcal{T}) > 0$. We assume this condition is met.[4] Differentiating the first order condition with respect to θ yields

$$\frac{d\mathcal{T}}{d\theta} = \frac{f'(\theta)X(\mathcal{T})}{c'' - f(\theta)X'(\mathcal{T})} \geq 0.$$

The inequality follows from the fact that f is increasing in θ and X is non-negative. These comparative statics indicate that the moments of the error distribution $\mu(\theta)$ and $\sigma(\theta)$ decline with intrinsic motivation.

D Examiner Credit Structure

Here, we provide expressions for $g_y^r(S,T)$, for $y \in \{GR,\ ABN,\ REJ,\ FIGHT\}$. We write $g_y^r(S,T) = \nu_y^r \cdot c(S,T)$ and detail the values of the raw credit terms ν_y^r and the adjustment terms $c(S,T)$ in turn. Granting in the first round gives the examiner a payoff of $\nu_{GR}^1 = 2$ credits. Rejecting in the first round gives $\nu_{REJ}^1 = 1.25$. If the applicant abandons in round one,

[3] We can also let the cost depend on an examiner's productivity without any adjustment to the results.

[4] There will be conditions on $\mu(\mathcal{T})$ and $\sigma(\mathcal{T})$ such that the second order conditions hold. We can derive sufficient conditions in the case in which μ depends on $\mathcal{T}$ but σ does not, and the case in which σ depends on $\mathcal{T}$ but μ does not. Details are available on request.

Table D.1. Seniority Corrections for Examiner Credit Adjustments

Seniority Grade	**Signatory Authority**	$c_{SEN}(S)$
GS-5	None	0.55
GS-7	None	0.7
GS-9	None	0.8
GS-11	None	0.9
GS-12	None	1.0
GS-13	None	1.15
GS-13	Partial	1.25
GS-14	Partial	1.25
GS-14	Full (primary examiner)	1.35

Notes: This table provides the seniority factors for credit adjustment. In the empirical work, we use 1.15 for GS-13 and 1.25 for GS-14.

the examiner obtains $\nu^1_{ABN} = 0.75$. Since g^r_{FIGHT} is only received upon submission of an RCE, $v^r_{FIGHT} = 0$ for all odd r. Granting in the second round gives $\nu^2_{GR} = 0.75$ credits. Rejecting in the second round gives $\nu^2_{REJ} = 0.25$ credits, with an extra $\nu^2_{ABN} = \nu^2_{FIGHT} = 0.5$ credits whether the applicant abandons or continues to an RCE. Ultimately, the examiner obtains two credits irrespective of what happens in the first two rounds. The only difference is whether they get the credits immediately (say, from an immediate grant) or spread out over two rounds.

The structure of the payoffs in the first RCE is the same, with $\nu^3_{ABN} = 0.75$, except $\nu^3_{GR} = 1.75$ and $\nu^3_{REJ} = 1$. Similar to before, $\nu^4_{GR} = 0.75$. In the first RCE, irrespective of what occurs, the examiner will obtain 1.75 credits. The distinction is whether examiners earn the full 1.75 credits immediately by granting, or one credit from their non-final rejection and $\nu^4_{REJ} = 0.25$ plus $\nu^4_{ABN} = \nu^4_{FIGHT} = 0.5$ credits from the applicant's response. In the second and any subsequent RCEs, the structure of the payoffs is same, except $\nu^{2r+1}_{ABN} = \nu^{2r+1}_{REJ} = 0.75$ and $\nu^{2r+1}_{GR} = 1.5$ $(r > 1)$. There is no difference for $\nu^{2r+2}_{GR} = 0.75$, $\nu^{2r+2}_{REJ} = 0.25$, and $\nu^{2r+2}_{ABN} = \nu^{2r+2}_{FIGHT} = 0.5$ $(r > 1)$.

Seniority and Technology Complexity Adjustments

The seniority and technology complexity adjustment term is $c(S,T) = \frac{c_{TECH}(T)}{c_{SEN}(S)}$. Table D.1 gives the values of $c_{SEN}(S)$ across the GS categories. Higher seniority factors imply larger values of c_{SEN} and, thus, lower values of credits. Table D.2 gives the values of $c_{TECH}(T)$ we created for the different technology centers and used in estimating the model. The Patent Office does not have adjustments at the technology center level but at the more detailed U.S. Patent Class (USPC)

Table D.2. Technology Center Corrections for Examiner Credit Adjustments

Technology Center T	**USPTO Number**	**Correction** ($c_{TECH}(T)$)
Chemical and Materials Engineering	17	22.2
Computer Architecture Software and Information Security	21	31
Computer Networks, Multiplex, Cable and Cryptography/Security	24	29
Communications	26	26.5
Semiconductors, Electrical and Optical Systems and Components	28	21.4
Transportation, Electronic Commerce, Construction, Agriculture...	36	22.4
Mechanical Engineering, Manufacturing and Products	37	19.9

level. We obtained adjustments at the USPC level from the Patent Office and constructed a patent-application weighted average for each technology center.

E Distance Measure

This section provides extra details on how we construct our distance measure. We also briefly describe our algorithm choice, the paragraph vector approach.[5]

Our approach to constructing the distance metric consists of four steps. The first step is text standardization. We perform basic changes to the content of the text (such as removing stopwords, punctuation, changing all to lowercase, etc.) and remove words that do not carry informational content. After standardization, we drop claims with fewer than two words or illegible text.

Second, we use the paragraph vector approach to represent the text of a patent claim as a numerical vector. The paragraph vector involves training a neural network and offers an improvement on the word vector approach in our context. We implement the paragraph vector approach using Gensim's `Doc2Vec` Python model (Řehůřek and Sojka, 2010). The model is trained to create a dense vector representation of each independent claim. We use the distributed memory method, which learns to predict a target word given the words in its context and the document vector. However, we experimented with the Distributed Bag of Words approach, too. We pass over the data 18 times in training (i.e., 18 epochs), and our dense vectors have 300 elements.

The third step involves taking every focal patent claim vector and calculating its distance to every *previously* granted claim, at the point of grant. We use cosine similarity and angular distance,

[5]At the time of writing this paper, we used the state-of-the-art approach, but there is a fast-moving frontier. Novel approaches use GPT-4 or BERT word embeddings integrated directly into neural networks. See Ash and Hansen (2023) for details on text algorithms.

which are standard in the text matching and NLP literature. We compute the cosine similarity (CS) between claim text vectors x and y as

$$cs(x,y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_j x_j^2 \sum_j y_j^2}}.$$

Then, we calculate the angular distance metric, $AD(x,y) = \arccos(cs(x,y))/\pi$, and finally, double AD to obtain a normalized distance in the interval $[0,1]$. With all distances computed, the final step is to find the closest claim. As robustness, we experimented with using an average of five closest distances. The resulting distance distribution was similar.

F Moment Selection

First, we provide a broad set of moments we could use to estimate our model. Then, we give information the methods we use to prune moments from the full set.

Available Moments

We have seven sets of moments available, which we describe in turn. Our first group of moments corresponds to examiners' issuance and applicants' abandonment decisions. For each round in the model and each seniority level, we calculate the proportion of applications examiners grant and the proportion that applicants abandon. Across nine seniority grades and six rounds, this implies **108** moments.

Second, we observe the distribution of the proportion of claims rejected, both by round and seniority grade. These observations generate another **54** moments. Third, we obtain **4** moments from the proportion of granted patents that renew at four, eight, and twelve years after issuance.

Fourth, we calculate the distribution of claim distances by round. We calculate the mean and standard deviation of the distance distribution by round for six rounds, implying **12** moments on distance. Fifth, we calculate each examiner's *leniency*, which is their average rejection rate across all the applications they examine. Hence, for each seniority grade, we obtain a distribution of examiner rejection rates, for which we can calculate the mean and standard deviation of the distribution of examiner's leniency. From this, we obtain another **18** moments.

Next, given that we can identify the distance threshold externally, we calculate the proportion of granted patents containing at least one invalid claim. Hence, for each round and each seniority level, we calculate the proportion of patents granted containing an invalid claim, implying another **54** moments. Another **108** moments come from calculating the mean and standard deviation of the size of errors (threshold less granted distance) for each seniority and in each round.[6]

Finally, we observe the distribution of application fighting costs. We have six moments on the distribution of legal application fees for four technology categories (simple, chemical, electrical, and mechanical), which we match to the technology centers on which we estimate the model. This implies another **24** moments.

Choosing Moments

We have nearly 400 data moments but only 21 parameters to estimate with simulated method of moments. However, not all moments will aid the estimation procedure in identifying the parameters, so we prune the set of moments for estimation.

We follow a rigorous, data-driven methodology to create a subset of the moments that best estimate the parameters. To do this, we first calculate the sensitivity matrix, described in Section 4.3. If a moment had a negligible value in the sensitivity matrix for all parameters, we considered it as not useful in estimating our model. Further, as described in Jalali, Rahmandad, and Ghoddusi (2015), we plot each moment against each parameter, fixing the other parameters at their estimates. If this curve is flat, this parameter does not influence the value of the moment. For a given moment, if the curve is flat across *all* parameters, it suggests that the moment offers no variation to identify the parameters. We also plot the value of the SMM objective against each parameter, fixing other parameters at their estimates (see Figure A.3 for examples in the estimated model). We experiment with moments until the SMM is approximately convex in each parameter to ensure a well-defined global minimum exists.

By combining the sensitivity matrix with moment and SMM plots, we pruned the moments down to those that assist in estimating the parameters. Since we split many parameters into two seniority groups (junior and senior), we split some moments into the same seniority categories.

Full Set of Moments

1. The proportion of applications granted in each round for juniors and seniors, for rounds one, two, three, and all rounds after four combined [eight moments]
2. The standard deviation of the distribution of examiner rejection rates for the six seniority categories used by the Patent Office (GS levels 7, 9, 11, 12, 13, and 14) [six moments]
3. The proportion of patents granted containing an invalid claim (for juniors and seniors) for rounds one and two [four moments]

[6] To calculate these moments, we take the subset of claims for which the granted distance $\tilde{D}_j$ is below the distance threshold τ, and then work out the mean and variance of $\tau - \tilde{D}_j$, which represents the size of the error.

4. The mean and standard deviation of the error size for juniors and seniors [four moments]
5. The proportion of abandonments in each round, when the assigned examiner is junior and senior, for rounds one and two [four moments]
6. The proportion of granted patents not renewed, renewed at year four but not eight, renewed at year eight but not twelve, and renewed at year twelve [four moments]
7. The mean and standard deviation of the distribution of granted claim distances for rounds one, two, and three [six moments]
8. Mean and median of legal application fees for simple applications and complex applications in electrical, mechanical, and chemical technologies [eight moments]

G Quantification of Social Costs

G.1 Type 1 Social Cost

Type 1 social cost includes two components: litigation costs for litigated patents and deadweight loss for non-litigated patents. Based on the probability of litigation, as discussed in the main text, the expected social cost of granting an invalid patent s of value $\tilde{V}_s$ is

$$S_{1s} = I_s DWL_s + (1 - I_s)\left[0.837 \cdot DWL_s + 0.163 \cdot 2\mathcal{C}(\tilde{V}_s)\right], \tag{G.1}$$

where DWL is the deadweight loss given in the text and $I_s = 1(\tilde{V}_s \leq \check{V})$ represents a dummy equal to one if the patent's value exceeds the exposure threshold. Then, the total type 1 cost is

$$T_1 = \sum_{s \in S_G} E_{1s} S_{1s}$$

where E_{1s} is equal to one if a granted application $s \in S_G$ is invalid and zero otherwise.

Details on Deadweight Loss Calibration

From the main text, we have that

$$DWL = \frac{1}{2}\Delta\wp\Delta q = \frac{1}{2}\frac{\Delta q}{q}q\Delta\wp = \frac{\lambda}{2}\frac{\Delta\wp}{\wp}\tilde{V},$$

by the definitions of $\tilde{V}$ and λ. We calibrate the term $\Delta\wp/\wp$ using the following derivation:

$$\frac{\Delta\wp}{\wp} = \frac{q\Delta\wp}{q\wp} = \frac{\text{lic. rev}}{\text{sales}} = \frac{\text{lic. rev}}{\text{R\&D}} \cdot \frac{\text{R\&D}}{\text{sales}}$$

As described in the text, we refer to Schankerman and Schuett (2022) for the ratio of licensing revenue to R&D and data from the Bureau of Economic Analysis for the ratio of R&D to sales.

Deadweight Loss Extension to Cournot Competition

In the main text, we compute deadweight loss from a patented invention assuming symmetric licensees operate in a perfectly competitive industry. Suppose instead that the licensees compete in a Cournot setting. By standard calculations, the equilibrium price-cost margin is $\dfrac{\wp - c}{\wp} = \dfrac{m^*}{\lambda}$ where $m^* = \frac{1}{N}$ is the average market share, and λ is the demand elasticity. We write this as $\frac{\wp-c}{\wp} = \frac{H^e}{\lambda}$ where H^e is the symmetric-equivalent Herfindahl index (HHI) of concentration. Thus, for $H^e < 1$, we have $\wp = \frac{c}{1-\frac{H^e}{\lambda}}$.

With imperfect competition, the change in equilibrium price is larger than the Arrow royalty due to double marginalization: $\Delta\wp = \frac{\Delta c}{1-\frac{H^e}{\lambda}} > \Delta c$. The associated deadweight loss with Cournot competition is

$$DWL_{\text{cournot}} = \frac{1}{2}\Delta\wp\Delta q = \frac{1}{2}\frac{\Delta c}{1-\frac{H^e}{\lambda}}\Delta q = DWL_{\text{pc}} \cdot \frac{1}{1-\frac{H^e}{\lambda}},$$

where in this case, $\tilde{V} = q\Delta c$ denotes total *royalty payments*. Since $H^e \in (0,1)$ and $|\lambda| > 1$, deadweight loss in this imperfect competition setting is larger than under perfect competition.

Using U.S. Census data for 2007, the value-added weighted-average HHI for manufacturing industries based on the 50 largest firms, H, for manufacturing sectors is 0.05. As is well-known, the HHI can be decomposed as $H = \frac{1}{N} + N \cdot \text{Var}(m) = H^e + N \cdot \text{Var}(m)$, where m is the market share of each firm. Thus, the observed H overstates the unobserved H^e, so the computed deadweight loss will be an upper bound to the true value of DWL. Despite this, the upper bound for the Cournot setting is not materially different from the competitive case in the text.

The value of H varies widely across industries. We do not compute deadweight loss using industry-specific values because it is challenging to assign patents in different patent classes to industries, and the existing Patent Office concordance is problematic (e.g., the mapping is not unique).

Calibrating Litigation Costs

To calibrate litigation costs, $\mathcal{C}(\tilde{V})$, we use data from the American Intellectual Property Law Association (AIPLA) surveys on litigation costs as a function of the value at stake, which we assume is the same for the patentee and challenger. We use the linear specification $\mathcal{C}(\tilde{V}) = \ell_0 + \ell_1\tilde{V}$. Using this same specification, Schankerman and Schuett (2022) estimate $\ell_0 = \$624,000$ and $\ell_1 = 0.162$ (2018 USD). This calibration of legal costs is at the patent, not claim, level.

Implementation of Type 1 Social Cost

A key challenge in implementing our calculation of type 1 social costs is that the estimates of the value of patent rights for invalid patents include potential litigation costs. To impute the "value at stake" in litigation for these patents, we adjust our methodology to exclude these costs.

To make this adjustment, we make two assumptions:

A1: Valid patents are not litigated. This assumption holds in a model with perfect courts, where a competitor either knows or pays a fee to discover whether a patent is valid, and then chooses whether to litigate based on the result. This assumption allows us to calculate the value of patent rights for valid patents, $\tilde{V}$, as equal to the observed value, since there are no litigation costs to net out.

A2: The *distribution* of the value at stake, $G_{\tilde{V}}(\cdot)$, is the same for valid patents as invalid patents. The basis for this assumption is that initial distances and values are uncorrelated in the model. This assumption allows us to draw values from the observed distribution of $\tilde{V} = V$ for valid patents and use them as draws from the distribution of $\tilde{V}$ for invalid patents.

Given A1 and A2, the procedure for calculating type 1 social costs is as follows:

1. Estimate the parameters of a log-normal distribution for the value at stake for *valid* patents.[7] Let the estimated distribution be denoted as $\hat{G}_{\tilde{V}}(\cdot)$.
2. Let $\bar{P}$ be the total number of *invalid* patent grants for the given period we simulate. Then, for each $p = 1, \ldots, \bar{P}$:
   (a) Take a draw from the estimated distribution of *valid* patents' value at stake (ex post value), $\hat{G}_{\tilde{V}}(\cdot)$, to represent the value at stake for the invalid patent p
   (b) Using the draw, calculate S_{1p} from Equation (G.1).
3. Calculate the total social cost of type 1 error as $\sum_{p=1}^{\bar{P}} S_{1p}$.

Finally, note that we calculate the threshold for exposure to litigation from the empirical distribution of the value at stake for valid patents, $\hat{G}_{\tilde{V}}(\cdot)$.

G.2 Type 2 Social Costs

Implementing Type 2 Social Cost Calculation

The primary challenge in calculating type 2 social costs comes from calibrating the value of the invention without patent rights (Π). This task is particularly difficult for inventions with a negative expected value of applying for a patent (Γ^*), where we cannot use the patent premium. In a similar vein to our approach to type 1 social costs, we assume that the distributions of Π for those with positive and negative Γ^* are the same and then draw values of Π from this distribution

[7] The sum of log-normal terms is approximately log-normal, which our simulation displays (Dufresne, 2004).

for those inventions for which Γ^* is negative. To be precise, our specific implementation is:

1. Draw a pilot set of potential inventions, to calculate a distribution of Π. Run this set of potential inventions through the model and calculate Γ^*. For those with positive Γ^*, create a distribution of Π using the relationship $\Gamma^* = \Psi\Pi$, where Ψ is the patent premium.
2. Start the simulation for type 2 social costs by drawing a new set of potential inventions (returns, distances, number of claims, fighting costs, examiner, etc.). For each potential invention i, calculate Γ_i^*. If $\Gamma_i^* > 0$, calculate $\Pi_i = \frac{\Gamma_i^*}{\Psi}$. If $\Gamma_i^* \leq 0$, draw a value of Π_i from the distribution calculated in step 1. Also, draw a development cost κ_i.
3. For each of the potential inventions i, work out the subset $\ell = 1, \ldots, \mathcal{I}_{\text{no dev}}$ that do not develop as those with $\max\{\Gamma_\ell^*, 0\} + \Pi_\ell < \kappa_\ell$
4. For $\ell = 1 \ldots, \mathcal{I}_{\text{no dev}}$, run the potential invention through a model where, at the point of abandonment, the inventor obtains all valid claims they have, and so obtains the patent value of their valid claims, instead of a payoff of zero. By definition, this scenario has the property that all abandoned claims are invalid so that there is no type 2 error. Let Γ_ℓ' denote the expected value of patent rights in this new scenario.
5. From $\ell = 1 \ldots, \mathcal{I}_{\text{no dev}}$, calculate the subset $m = 1, \ldots, \mathcal{I}_{\text{now dev}}$ who have $\max\{0, \Gamma_m'\} + \Pi_m \geq \kappa_m$. This is the set of potential inventions that do not develop with type 2 error but would develop in the absence of type 2 error.
6. For $m = 1, \ldots, \mathcal{I}_{\text{now dev}}$, calculate $SNB_m = \frac{\rho_{\text{soc}}}{\rho_{\text{priv}}}\Big(\max\{0, \Gamma_m^*\} + \Pi_m\Big) - \kappa_m$ and calculate the total type 2 social cost as
$$T_2 = \sum_{m=1}^{\mathcal{I}_{\text{now dev}}} SNB_m.$$

Calibrating Development Costs

We apply the estimates of development costs from Schankerman and Schuett (2022) to our context. They assume that development costs κ are exponential, with mean equal to $k_0 + k_1 z$, where z is the size reduction of the invention and k_0 and k_1 are estimated as 254.6×10^3 and 2.33×10^{10}, respectively. Regarding the size reduction, they assume that z is log-logistic distributed with parameters $\beta_0 = 1.02$ and $\beta_1 = 1.14 \times 10^{-6}$. We use the mean value of z in our calibration.

In the baseline quantification, we draw values of κ from the distribution described above, which assumes that development costs are independent of Γ^* and Ψ. In this model, inventors know their development costs before they decide to develop their idea. We also experiment with another model, which makes the opposite assumption that inventors do not know their development costs and thus use the mean value, $\bar{\kappa} = k_0 + k_1 \bar{z}$, to make their development decision. Both models produce similar conclusions; results are available upon request.

Calibrating the Number of Ideas

To compute the number of ideas, we start with the average annual number of utility patent applications in the period 2011–2013. We convert this into the number of ideas in two steps. First, we use the estimates from Schankerman and Schuett (2022) that about two-thirds of applications are "low type" inventions (defined by them as those that would have been developed even without patent protection), and second, that one-third of ideas become a low type application. Together, this implies about one million ideas for potential inventions for the cohort of applications.

G.3 Patent Prosecution Costs

The amendment cost for application s is the per-negotiation cost $F_{\text{amend},s}$ drawn from the estimated distribution, multiplied by the equilibrium number of negotiations for application s (equal to the number of rounds r_s minus 1). We also include the fixed application attorney cost $F_{\text{app},s}$ implied by the equilibrium padding choice. For administrative costs, we calculate the average Patent Office cost per round and claim, denoted RCC, and multiply it by the number of rounds r_s and claims $M_{0,s}$. Then, the total social cost of patent prosecution is

$$T_3 = \underbrace{\sum_s F_{app,s} + (r_s - 1)F_{\text{amend},s}}_{\text{Attorney Costs}} + \underbrace{\sum_s M_{0,s} r_s RCC}_{\text{Administrative Costs}}.$$

References

Ash, E. and S. Hansen (2023): "Text Algorithms in Economics," *Annual Review of Economics*, 15, 659–688.

Dufresne, D. (2004): "The Log-Normal Approximation in Financial and Other Computations," *Advances in Applied Probability*, 36, 747–773.

Jalali, M., H. Rahmandad, and H. Ghoddusi (2015): *Using the method of simulated moments for system identification*, MIT Press.

Nelsen, R. B. (2007): *An introduction to copulas*, Springer Science & Business Media.

Řehůřek, R. and P. Sojka (2010): "Software Framework for Topic Modelling with Large Corpora," *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.

Schankerman, M. and F. Schuett (2022): "Patent Screening, Innovation, and Welfare," *The Review of Economic Studies*, 89, 2101–2148.