Comparative Legal Systems, Legal Evolution, and Private Contracting

Political Economics: Week 9

Giacomo A. M. Ponzetto

CREI – UPF – Barcelona GSE

5th and 8th March 2010

The Three Branches of Government

1. Political-economy models often assume that policy is determined by a unitary decision-maker, representing the government.
   - Reasonable fit to the choices of a strong executive, as in the UK.
2. Models of legislative bargaining recognize that laws, and particularly the budget, are written by parliaments whose members may have individual agendas.
   - Coalitions in proportional-representation systems.
   - National parties v. local constituencies in the U.S. Congress.
3. The judiciary has received much less attention as a co-equal third branch of government
   - Law and economics exists as a separate field; but so do public economics and public finance.

Judge-Made Law

- Courts are tasked with resolving disputes, and they are crucial actors in the implementation of policy, particularly in the field of public regulation of private activity.
- In several domains that affect commercial activity, the law itself is formulated not only by the legislature, but also by courts through their decisions.
  - Judicial precedent is an especially important source of tort law.
- Posner (1972) conjectured that the evolution of case law gradually leads to the adoption of efficient legal rules, because all judges try to achieve economic efficiency, even if they may have different distributional preferences.
- Empirical evidence shows that judges hold heterogeneous biases and political preferences that need not be consistent with efficiency maximization.

A Simple Model of Tort

- An offender O may cause an accident that inflicts unit harm to a victim V.
- O can take precautions that cost C and reduce the probability of an accident from p_N to p_P.
- The conditional probability of an accident depends on two attributes a and u which are independently and uniformly distributed on $[0,1]$ in the population of potential cases:

$$p_N - p_P = \begin{cases} \underline{\Delta} & \text{for} \quad a+u<1 \\ \bar{\Delta} & \text{for} \quad a+u \geq 1 \end{cases} \quad \text{with } \bar{\Delta} > C > \underline{\Delta}.$$

- Damages are so high that O takes precautions whenever he is liable.
  - E.g., a jury could award punitive damages.
- The efficient rule holds O liable if and only if $a+u \geq 1$.

Imperfect Information

- When the first rule is set, the only available information concerns a, while u is unknown: thus the court's ruling can only be based on a.
- Any non-perverse rule takes the form of a threshold $A \in [0,1]$ such that O is liable if and only if $a \geq A$.
- Imperfect information implies that any rule induces statistical errors.

  Type I: false positives occur with probability

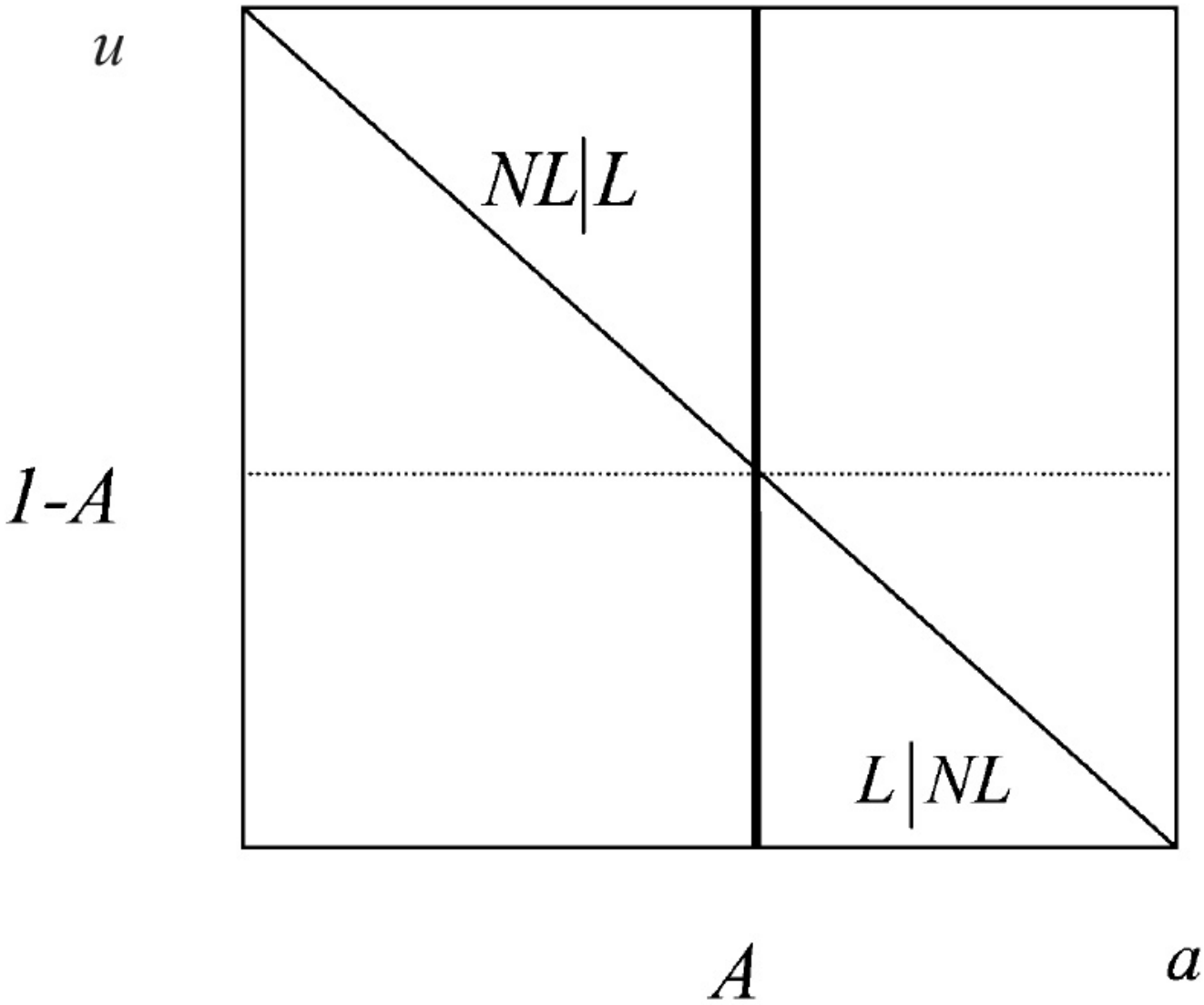
$$o(A) \equiv \Pr(L|NL) = (1-A)^2/2.$$

  Type II: false negatives occur with probability

$$v(A) \equiv \Pr(NL|L) = A^2/2.$$

- The social welfare function attaches a cost $\lambda_O > 0$ to inefficient over-precautions and $\lambda_V > 0$ to inefficient under-precautions. $\lambda \equiv \lambda_O/\lambda_V$ is the relative cost of over-precautions.
- The efficient one-dimensional rule is $A^* = \lambda/(1+\lambda)$.

One-Dimensional Rules



Biased Judges

- Judges dislike all errors, but they need not weigh different types of error in the same way as the social welfare function.
  - Crucially, no judge is so biased that he consciously prefers introducing mistakes in the rule.

$\Rightarrow$ Judge i perceives a cost $\beta_{O,i} > 0$ of false positives and $\beta_{V,i} > 0$ of false negatives.

- Judges derive utility from the quality of the legal rule their decision establishes.
  - They do not care only about the outcome of the case they are deciding.
  - They do not engage in a strategic interaction with future judges.

$\Rightarrow$ Judge i's preferred one-dimensional rule is $\hat{A}_i = \beta_{O,i} / \left(\beta_{O,i} + \beta_{V,i}\right)$.

- All judges have the same preference intensity, $\beta_{O,i} + \beta_{V,i} = 1$.

Judicial Biases

1. Fraction γ of judges are *unbiased* efficiency maximizers with

$$\frac{\beta_{O,i}}{\beta_{V,i}} = \lambda \Leftrightarrow \hat{A}_i = A^* = \frac{\lambda}{1+\lambda}.$$

2. Fraction $(1-\gamma)/2$ of judges have a *pro-O* bias and

$$\frac{\beta_{O,i}}{\beta_{V,i}} = \lambda\pi \Leftrightarrow \hat{A}_i \equiv \hat{A}_O = \frac{\lambda\pi}{1+\lambda\pi} > A^*.$$

3. Fraction $(1-\gamma)/2$ of judges have a *pro-V* bias and

$$\frac{\beta_{O,i}}{\beta_{V,i}} = \frac{\lambda}{\pi} \Leftrightarrow \hat{A}_i \equiv \hat{A}_V = \frac{\lambda}{\pi+\lambda} < A^*.$$

- $\pi > 1$ measures the extent of judicial polarization.

Distinguishing

- Judges are required to abide by the holding of the first court, but they are allowed to refine legal rules by distinguishing, i.e.,

  "not[ing] a significant factual, procedural, or legal difference (in an earlier case), usu[ally] to minimize the case's precedential effect or to show that it is inapplicable."

- A long-standing tradition of legal formalism holds that the *ratio decidendi* of a case is unambiguously identifiable, and that therefore distinguishing is confined within strict, logically defined boundaries (Goodhart 1930).
- Gennaioli and Shleifer's (2007) model captures such a binary view of the evolution of case law:
  1. Some legal changes are impossible for judges to effect.
  2. Any possible change can be effected by incurring a fixed effort cost k.

Permissible Distinguishing

1. Distinguishing must introduce a new empirical dimension b into adjudication.
2. The new judge must respect the original precedent in the sense that he can only choose two threshold $B_0 \in [0,1]$ and $B_1 \in [0,1]$ such that O is liable if and only if

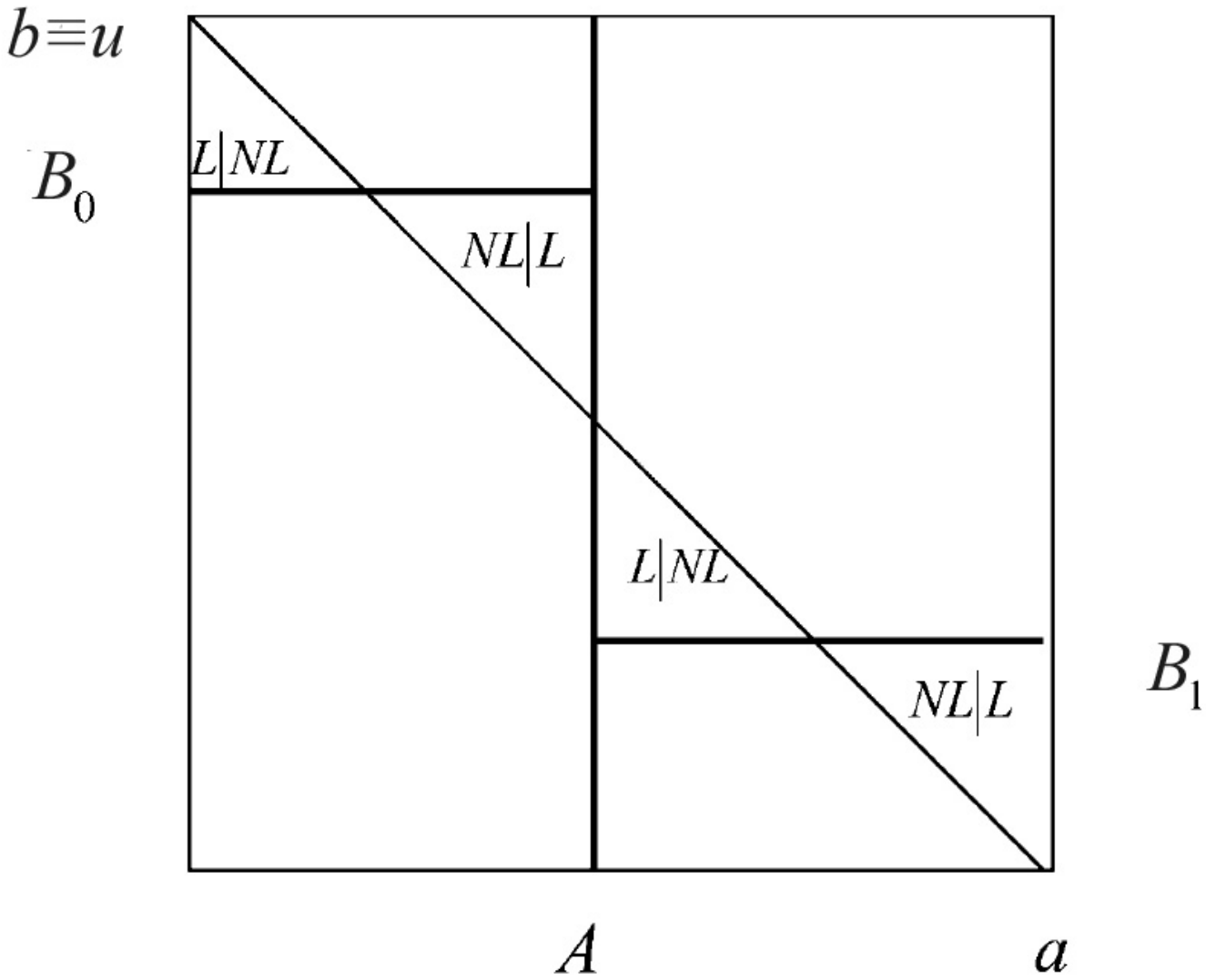
$$a < A \text{ and } b \geq B_0$$

or

$$a \geq A \text{ and } b \geq B_1.$$

3. The new dimension must provide previously unexploited economic information about the optimal allocation of liability: i.e., b must coincide with u, because "the *materiality* of a dimension is a physical characteristic that even the most biased judges cannot subvert."

Piecewise Constant Rules



The Effects of Distinguishing on Errors

- If judge j distinguishes precedent A, he sets

$$B_{0,j} = 1 - A + A\hat{A}_j \text{ and } B_{1,j} = (1 - A)\,\hat{A}_j.$$

1. The ratio of type II to type I errors changes from

$$\frac{v(A)}{o(A)} = \left(\frac{A}{1-A}\right)^2$$

to

$$\frac{v(A,\hat{A}_j)}{o(A,\hat{A}_j)} = \left(\frac{\hat{A}_j}{1-\hat{A}_j}\right)^2 = \frac{v(\hat{A}_j)}{o(\hat{A}_j)}.$$

2. The overall prevalence of errors decreases from

$$v(A) + o(A) = \left[A^2 + (1-A)^2\right]/2$$

to

$$v(A,\hat{A}_j) + o(A,\hat{A}_j) = \left[A^2 + (1-A)^2\right]\left[\hat{A}_j^2 + (1-\hat{A}_j)^2\right]/2.$$

The Effects of Distinguishing on Welfare

1. Distinguishing introduces u into the rule, increasing its precision.
   - This increases welfare, since all judges share the goal of reducing errors.
2. Distinguishing replaces the first judge's relative bias $\hat{A}_1 / \left(1 - \hat{A}_1\right)$ with the second judge's $\hat{A}_2 / \left(1 - \hat{A}_2\right)$.
   - This reduces welfare iff the second judge is more biased than the first.

$\Rightarrow$ The social loss changes from

$$\Lambda(A) = \left[A^2 + \lambda (1 - A)^2\right] / 2$$

to

$$\Lambda\left(A, \hat{A}_j\right) = \left[A^2 + (1 - A)^2\right] \left[\hat{A}_j^2 + \lambda \left(1 - \hat{A}_j\right)^2\right] / 2.$$

- Welfare may fall if $\lambda \neq 1$ and π is large:
  1. if $\lambda > 1$ and the second judge is pro-V;
  2. if $\lambda < 1$ and the second judge is pro-O.

The Benefits of Distinguishing

1. If $\lambda = 1$, distinguishing improves the legal rule with certainty.
2. If $\lambda \neq 1$, distinguishing is beneficial on average.
   - Given the opportunity, judge j distinguishes judge's i precedent if

$$k \leq \frac{1}{2}\left(\hat{A}_i - \hat{A}_j\right)^2 + \hat{A}_i\left(1 - \hat{A}_i\right)\hat{A}_j\left(1 - \hat{A}_j\right).$$

   - Ex ante, there is equal probability of judge j distinguishing judge's i precedent, or judge i distinguishing judge's j precedent.
   - Expected welfare is higher if the probability is not zero, because

$$\Lambda\left(\hat{A}_i, \hat{A}_j\right) - \Lambda\left(\hat{A}_i\right) + \Lambda\left(\hat{A}_j, \hat{A}_i\right) - \Lambda\left(\hat{A}_j\right) < 0$$

   for all $\hat{A}_i, \hat{A}_j \in \left\{\hat{A}_V, A^*, \hat{A}_O\right\}$.

$\Rightarrow$ Judges' biases "wash out" on average, while more information is always good.

Judicial Polarization

1. Greater polarization implies that the legal rules are more severely biased, at least in expectation.
2. Greater polarization of either the first or the second judge reduces the increase in precision that the latter effects by distinguishing.
3. On the other hand, greater polarization increases the likelihood that a judge inherits a precedent he strongly disagrees with and that he is willing to distinguish in spite of the effort cost k.

- A trade-off between lower bias and greater precision can emerge.
- Some polarization can be optimal because it makes judges more willing to incur a private cost in order to provide a public good, i.e., distinguishing.

The Inefficiency of Legal Evolution under Distinguishing

1. Distinguishing cannot achieve the first-best efficient rule.
   - The efficient rule is not piecewise constant.
2. Distinguishing cannot achieve the optimal piecewise constant rule unless $\lambda = 1$ and $k \leq 1/16$.
   - The optimal piecewise constant rule does not involve $A = A^*$ unless $\lambda = 1$, but judges set the first rule myopically.
   - A judge only introduces u into the rule if it is not too costly to do so.
3. If $\lambda = 1$ and $k \leq 1/16$, distinguishing achieves the optimal piecewise constant rule with probability γ^2.
   - Both the first and the second judge must be unbiased, or else some bias is irreparably introduced.

Overruling

- Judges could also make law without respecting precedent, by overruling the decisions of previous courts and establishing a new rule.
- Suppose this is possible at a fixed cost k.
  - Judges might also falsely claim that they are distinguishing, while effectively overruling.
- If k is sufficiently low and π is sufficiently high, overruling enables legal changes even while u remains unobservable.
- As long as u is unobservable:

1. Greater legal polarization unambiguously reduces welfare.
2. Changes in the rule have no average effect on welfare.
   - Opposite biases cancel each other in expectation.
3. Ex ante welfare is identical whether or not judges may overrule.
   - If only distinguishing is possible, no legal change occurs.
   - This would be optimal if society had preferences for the early resolution of uncertainty.

Efficient Overruling

- u becomes observable exogenously.
  - Technological progress.
  - Courts can only rule on the facts that come up during the trial.
- With overruling, judges do not have to respect precedents and set a piecewise constant rule.
  - Balancing tests based on marginal trade-offs: O is liable if and only if $b \geq B(A)$, for any function $B : [0,1] \longrightarrow [0,1]$
  - E.g., Hand's Formula for the assessment of negligence liability.

$\Rightarrow$ As soon as u becomes observable, overruling establishes the first-best efficient rule: O is liable for $u \geq 1 - a$.

  - With complete information, all judges want to eliminate all errors.

- Greater legal polarization increases welfare by spurring judges to incur the private cost of adopting the first-best rule.

Why Stare Decisis?

- With these assumptions, overruling dominates distinguishing:
  1. Identical welfare while u is unobservable.
  2. Strictly greater welfare under overruling once u becomes observable.
- If a judge is willing to pay k to introduce u by distinguishing, he is willing to pay $K > k$ to introduce it by overruling.
  - Every judge strictly prefers the first-best rule he can set by overruling to the constrained-optimal piecewise constant rule he can set by distinguishing.
- Is u more likely to become observable when its value is lower both to society and to individual judges?
- Distinguishing would be disastrous if u remained unobservable but judges could introduce a spurious dimension $b \perp u$.
  - This strategy only could, and with sufficient polarization would, be used by strongly biased judges to destroy the precision of the rule and make O either always or never liable.

Stare Decisis and Judicial Incentives

- Hutchinson (2005): *contemporary judges have insisted that following precedent is not an all-or-nothing choice between blind adherence and total disregard.*
- Stare decisis means that judges must justify their decisions in terms of adherence to binding precedent.
- Calabresi (1982): *the major effective control on courts stems precisely from their duty to explain what they are doing.*
- Fernandez and Ponzetto's (2009) model captures the pragmatist view that rhetorical requirements shape the logical structure of the law, rather than vice-versa (Dewey 1924).
  1. All legal changes can potentially be effected through distinguishing.
  2. The effort cost of reconciling a judgement with the rhetorical demands of stare decisis is an increasing and convex function of the magnitude of the effected change.

Rhetorical Distinguishing

- Let μ_t be the probability of cases (a, u) for which either the old $(t-1)$ rule assigned liability but the new one (t) does not, or vice-versa.
- The effort cost required to change the rule is

$$k(\mu_t) = \frac{1}{2} c \mu_t^2 \text{ for } c > 0.$$

- While u is unobservable, the rule at time t is fully described by the threshold A_t, and

$$k(A_t, A_{t-1}) = \frac{1}{2} c (A_t - A_{t-1})^2 .$$

- Once u becomes observable, all judges want to minimize errors and achieve the first-best efficient rule.
- Legal evolution under perfect information is summarized by the probability of false positives o_t and of false negatives v_t, and

$$k(o_t, v_t, o_{t-1}, v_{t-1}) = \frac{1}{2} c (o_t + v_t - o_{t-1} - v_{t-1})^2 .$$

The Incessant Evolution of Case Law

- While u is unobservable, case law evolves as an $AR(1)$ Markov process:

$$A_t = \frac{c}{1+c} A_{t-1} + \frac{1}{1+c} \hat{A}_j.$$

- Case law is constantly evolving and does not converge in probability to any single rule (e.g., the optimal one-dimensional rule).
- Case law converges to the ergodic distribution

$$A_\infty \sim N\left(\mathbb{E}\left(\hat{A}_j\right), \frac{1}{1+2c}\sigma^2\left(\hat{A}_j\right)\right).$$

- The legal rule always incorporates, with different weighting, the perspectives of all previous judges as well as the current one.
- Judicial polarization unambiguously reduces welfare by increasing the variance of legal rules.
  - For $\lambda \neq 1$, it also increases the average bias, given Gennaioli and Shleifer's definition of $\hat{A}_V$ and $\hat{A}_O$.

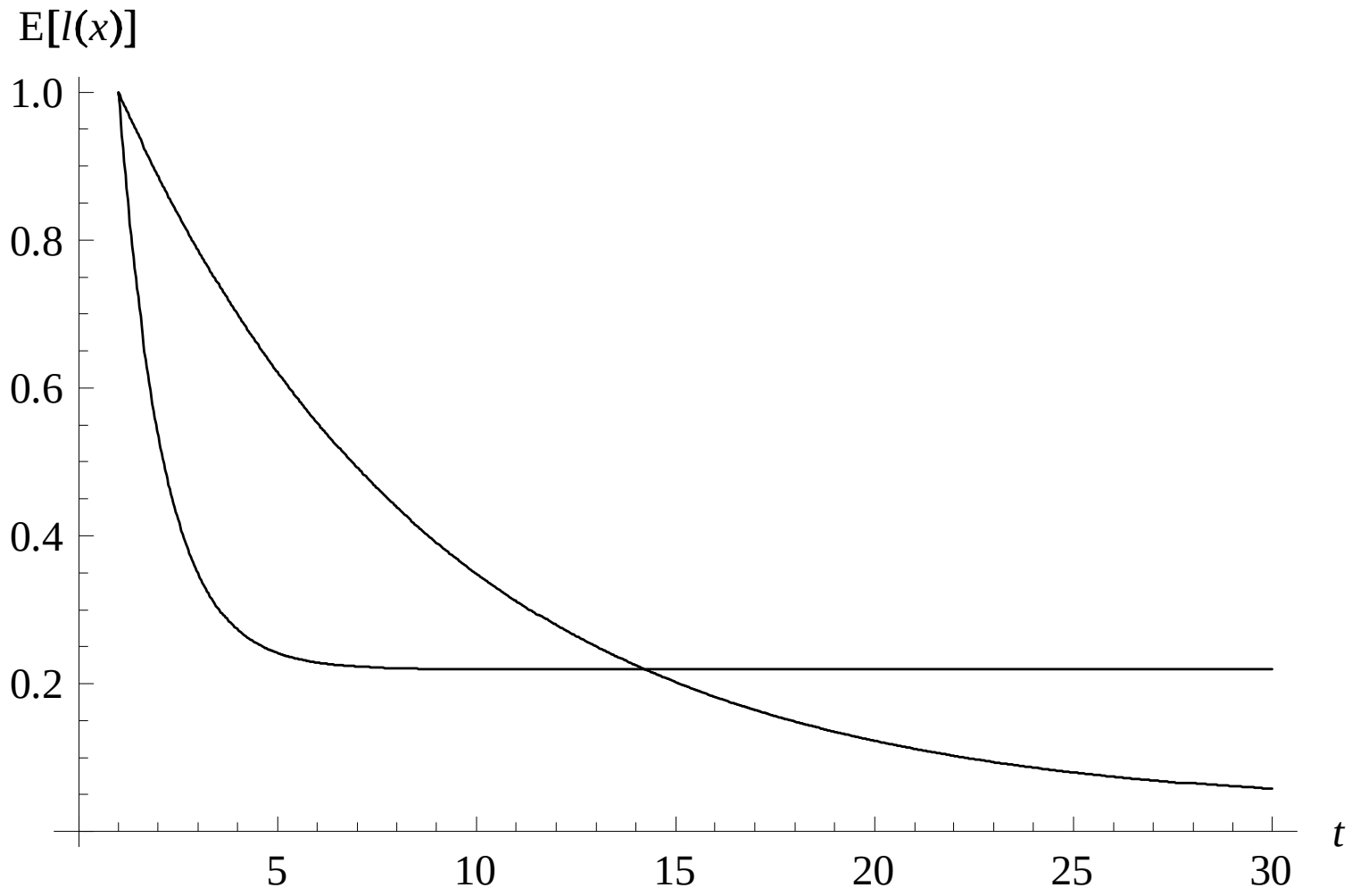
Convergence of Case Law towards Efficiency

- Stare decisis makes case law evolve gradually towards greater efficiency and welfare compared to either the absence of legal change or unchecked overruling.
- The social loss in period t is an affine transformation of $l(A_t) = (A_t - A^*)^2$, whose ex ante expectation is

$$\mathbb{E}l(A_t) = \sigma^2(A_t) + [\mathbb{E}(A_t) - A^*]^2.$$

- The expected value $\mathbb{E}(A_t) = \mathbb{E}(\hat{A}_j)$ is identical in all systems.
- The variance $\sigma^2(A_t)$ coincides with the variance of judicial preferences $\sigma^2(\hat{A}_j)$ both with no change and with unchecked overruling.
- Stare decisis promotes the certainty of the law: $\sigma^2(A_t)$ is monotone decreasing in t, and $\sigma^2(A_t) < \sigma^2(\hat{A}_j)$ for all $t > 1$.
- The cost of legal innovation c captures a trade-off between the speed of convergence and asymptotic efficiency.

The Benefits of Stare Decisis



$\sigma^2(A_t)/\sigma^2(\hat{A}_j)$ for $c = \{1.78, 15.7\}$

Convergence of Case Law to the First Best

- Once u becomes observable, case law under stare decisis converges gradually to the first-best efficient rule.
- Every ruling marginally decreases the prevalence of errors of the type the judge dislikes the most.
- Welfare is higher when judges are more active:
  1. when the cost of legal innovation is lower;
  2. when judicial polarization is higher.
- With complete information, stare decisis ensures that case law converges almost surely to the first-best efficient rule.
  - A strong endorsement of Posner's efficiency hypothesis.
  - Probably not too relevant: how often are courts perfectly informed?
- Overruling would reach efficiency in a single leap if k were low enough, but might never attain it if it were too high.

Democratic Representation

- A probabilistic-voting model in which special-interest groups have disproportionate political power.
  - Campaign contributions from lobbies, superior information, pivotality.
- The power of groups fluctuates randomly over time. At each point S groups are influential, and each has policy preferences $\hat{x}_s$ i.i.d. with mean x^* and variance σ^2.
  - It would not matter for the comparison of case law and statute law if interest groups had an average bias.
- The legislature is controlled by the majority party, and its policy preferences are distributed as

$$\hat{x}^L = (1 - \iota)\, x^* + \iota \sum_{s=1}^{S} \frac{\hat{x}_s}{S}.$$

$\iota \in (0, 1)$ measures the imperfection of democratic politics.

The Legislature and the Judiciary

- The preferences of the judiciary reflect two sources of bias.
  1. Judges are, at best, elected or appointed by elected representatives.
  2. Special-interests are directly involved in trials, and can sway courts.
- The preferences of each judge are distributed as

$$\hat{x}^J = (1-b)\,\hat{x}^L + b\hat{x}_s$$

  $b \in (0,1)$ measures outside influence on judicial decisions.
- The legislature is the more representative branch of government; its preferences have lower variance than the judiciary's:

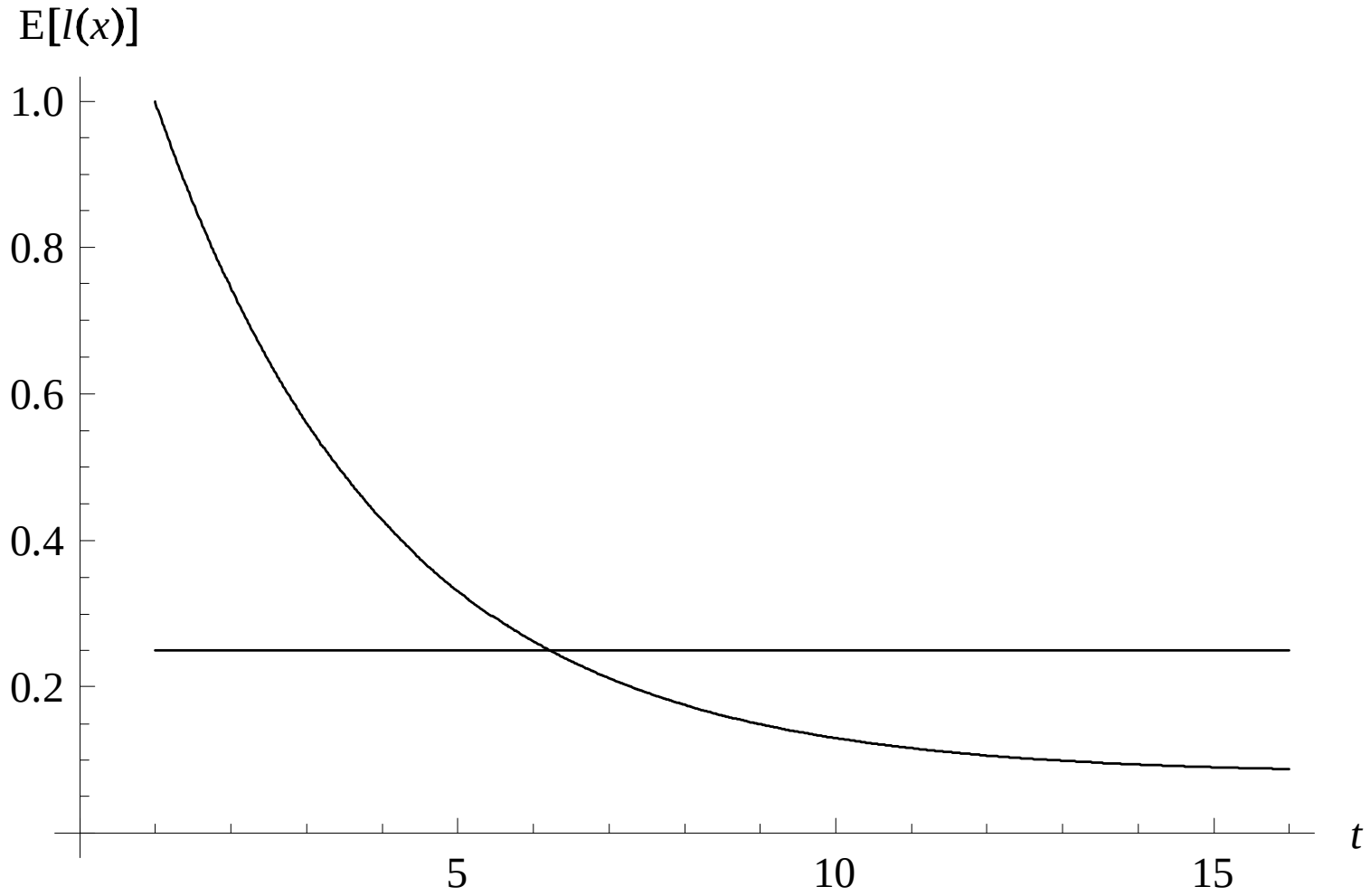
$$0 < \hat{\sigma}_L(\iota) < \hat{\sigma}_J(\iota, b) < \sigma^2.$$

- Political imperfections make both institutions less representative, but affect the legislature relatively more: $\partial\hat{\sigma}_L/\partial\iota > 0$, $\partial\hat{\sigma}_J/\partial\iota > 0$, and $\partial\left(\hat{\sigma}_L/\hat{\sigma}_J\right)/\partial\iota > 0$.
- Outside influence on judicial decisions makes judicial preferences more volatile: $\partial\hat{\sigma}_J/\partial b > 0$.

Case Law v. Statute Law

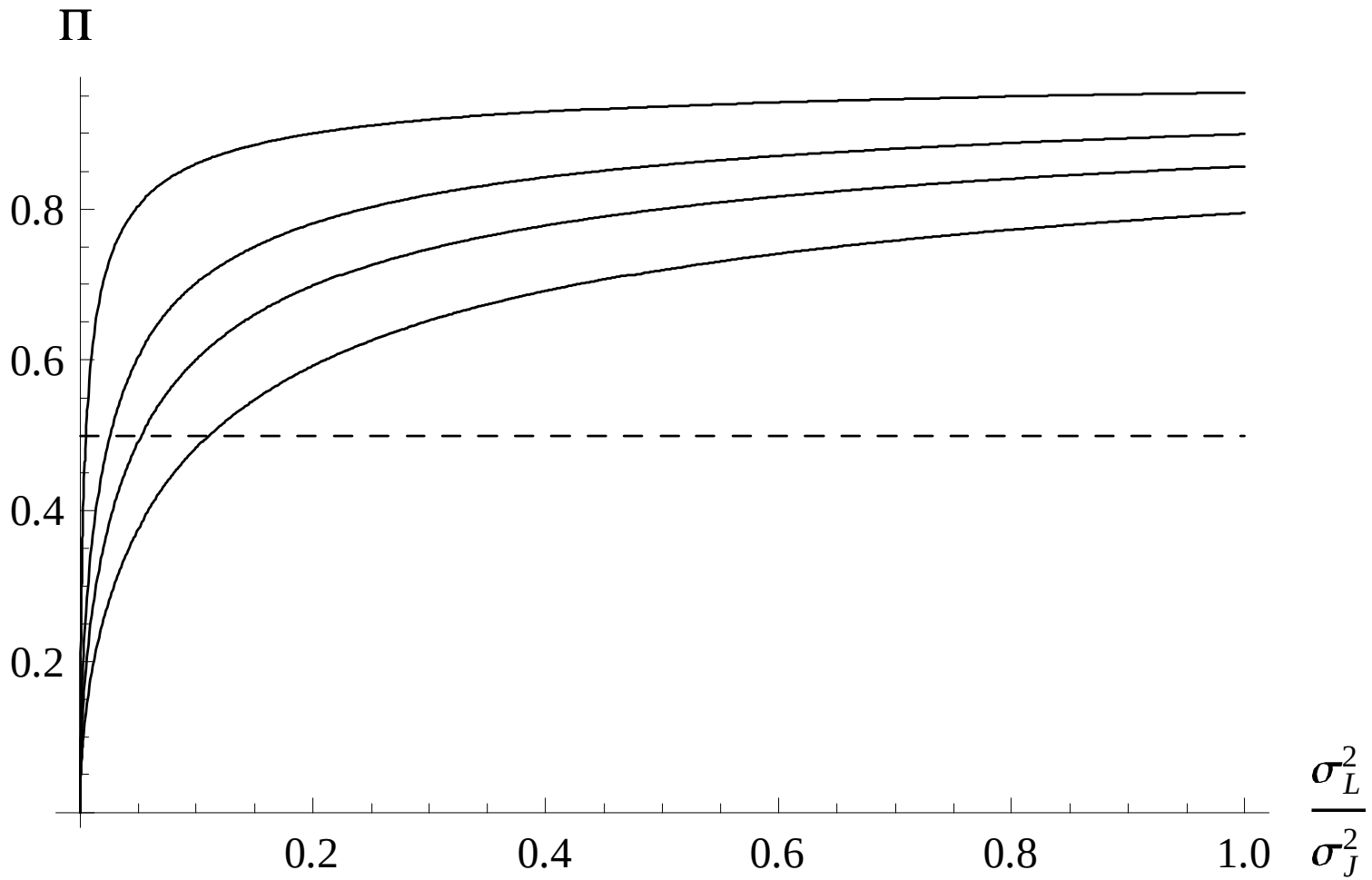
- Legislation is not bound by stare decisis. It is analogous to overruling, in that it has fixed cost of changing the law, regardless of the magnitude of the change.
- Statute law evolves as a strictly stationary Markov process whose unconditional distribution is $x_t^L \sim \hat{x}^L$ for all t.
- Case law evolves as an $AR(1)$ Markov process: $x_{t+1}^J = \psi x_t^J + (1-\psi)\hat{x}^J$, where $\psi = c/(1+c)$ is the effort cost of legal innovation.
- If stare decisis is sufficiently respected, there exists a finite value T such that the expected efficiency of statute law is lower than that of case law if and only if the latter embodies at least T precedents.
- In a stationary environment, stare decisis is not only expected but overwhelmingly likely to produce more efficient rules than legislation in the long run.

The Expected Efficiency of Case Law and Statutes



$\hat{\sigma}_J^2 = 4\hat{\sigma}_L^2 = 1$ and $\psi = .85$

Asymptotic Superiority of Case Law



The probability that case law is more efficient than statutes

Multiple Sources of Law

- The first-best efficient rule x_t^* varies over time as a result of of technological innovation and social change.
- A jump process with constant hazard rate $\lambda \in [0, 1]$ and i.i.d. random increments with mean zero and variance $\sigma_\xi^2 > 0$.
- Jumps in the optimum are met with a legislative response.
  - Statute-writing after 9/11, the Enron scandal, the financial crisis.
  - The new statute tends to, but need not, be an improvement.
- The legislature always desires more opportunities to intervene. It gets a chance to do so in the absence with social change with probability $1 - \alpha$, where $\alpha \in [0, 1]$ is a reduced-form measure of political accountability.
- Statutes are interpreted by courts in accordance with stare decisis.

The Optimal Legal System

- Pure statute law is optimal if and only if legislation is perfect ($\hat{\sigma}_L^2 = 0$); nonetheless, it is preferable to any system of pure case law if social change is sufficiently intense: $\lambda\sigma_\xi^2 \geq \hat{\sigma}_L^4 / \left(\hat{\sigma}_J^2 - \hat{\sigma}_L^2\right)$.
- The optimal system is pure case law when the social optimum is sufficiently stable and a mixed system when it is more volatile: $\sigma_\xi^2 \geq \Xi\left(\hat{\sigma}_J^2, \hat{\sigma}_L^2, \alpha, \lambda\right)$.
- The mixed system is more likely to be optimal when:

1. The extent of judicial biases is greater: $\partial\Xi/\partial\hat{\sigma}_J^2 < 0$.
2. Legislative preferences are more aligned with welfare: $\partial\Xi/\partial\hat{\sigma}_L^2 > 0$.
3. The legislature is more accountable: $\partial\Xi/\partial\alpha < 0$.
4. Social change is more frequent: $\partial\Xi/\partial\lambda < 0$.

Legal Origins Revisited

- For any specification of the process of social and political change, the introduction of precedent-bound judicial interpretation in a statute law system increases social welfare.
- The importance of judge-made law in different legal families is ranked, from common law at the top to French civil law at the bottom.
- Empirically, reliance on case law matters more than judicial independence (Beck, Demirgüç-Kunt, and Levine 2003a, 2005).
- French civil law does badly without *jurisprudence constante*:
  1. The French deviation (Merryman 1996): French civil law works well in France, but French courts have never stopped relying on case law.
  2. The transplant effect (Berkowitz, Pistor, and Richard 2003): French civil law hurts "unreceptive" transplants that are not OECD members.
- Efficiency pressures explain the gradual convergence of developed countries to a mixed systems of statutes and case law.

The Organization of Trials

- Another difference between legal systems, and particularly between common law and civil law, is the organization of trials.
  - Common law countries favour an adversarial system and adjudication of facts by a jury.
  - Civil law countries favour an inquisitorial system and adjudication by professional judges.
- Bringing adjudication under the centralized control of the state protects law enforcement from coercion by powerful litigants.
- Conversely, such control makes law enforcers beholden to the state and politicizes justice.
- Glaeser and Shleifer (2002) explain institutional choice as an efficient response to the relative risk of subversion of justice by a powerful central government or by powerful local interests.

Competing Interests in Adjudication

- Powerful special interests violate the law and are brought to court.
- Violations differ on two dimensions:
  1. $D > 0$ is the social value of punishing the violator.
  2. $R \in \mathbb{R}$ captures the separate impact of punishment on the government.
     - Independent distributions $F(D)$ and $G(R)$ such that $\mathbb{E}R = 0$.
- The government values punishment $D + \theta R$.
  - $\theta > 0$ measures the representativeness of the government.
- Through an appropriate incentive scheme, a government-appointed judge can be induced to punish if and only if punishment is in the interest of the government.
- A local jury values punishment $D - A$.
  - $A > 0$ measures the ability of special interests to influence justice.

Independent Juries v. Government-Appointed Judges

- Social welfare with an independent jury is

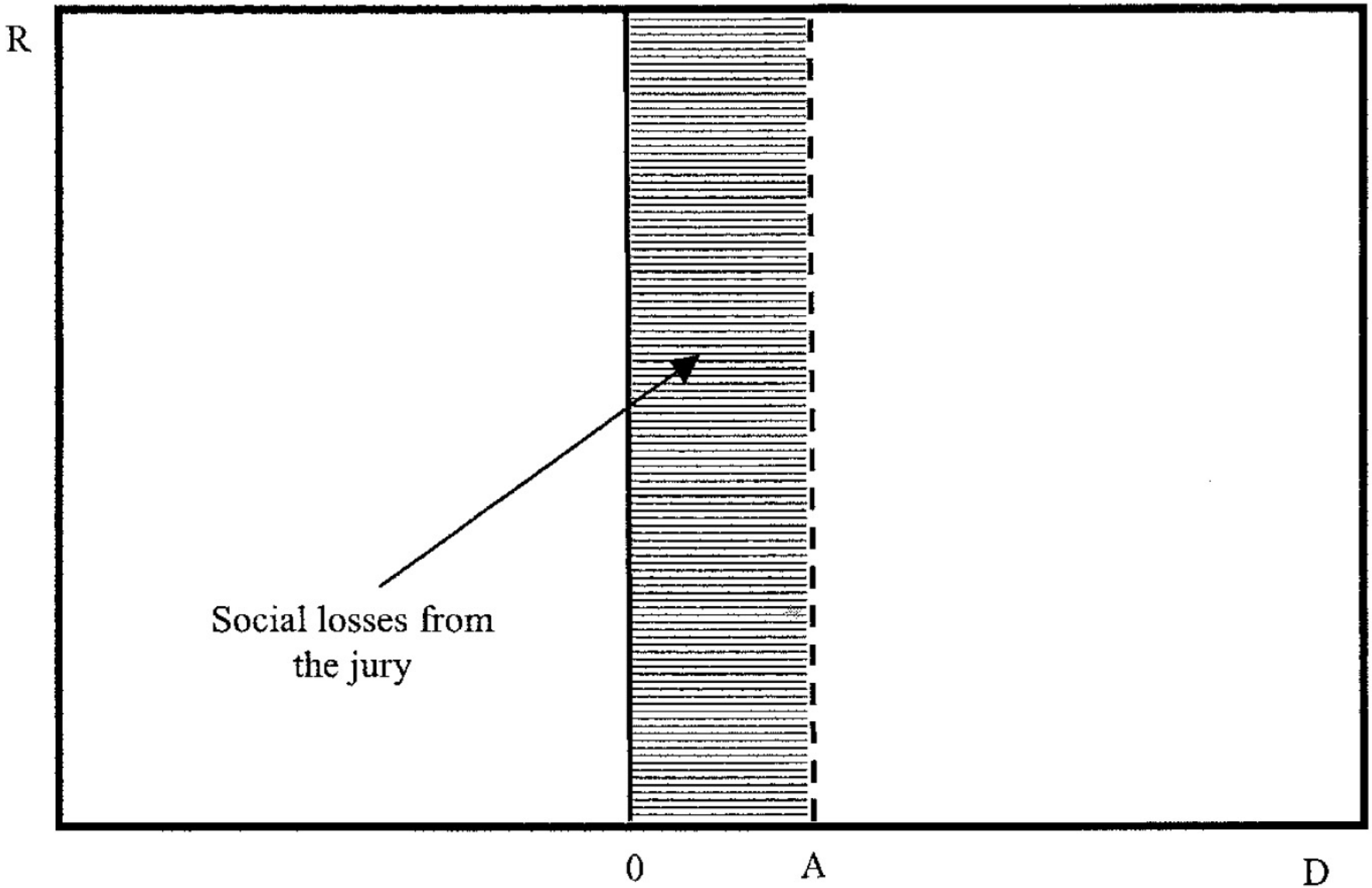
$$\int_A^{\infty} DdF(D).$$

- Social welfare with a government-appointed judge is

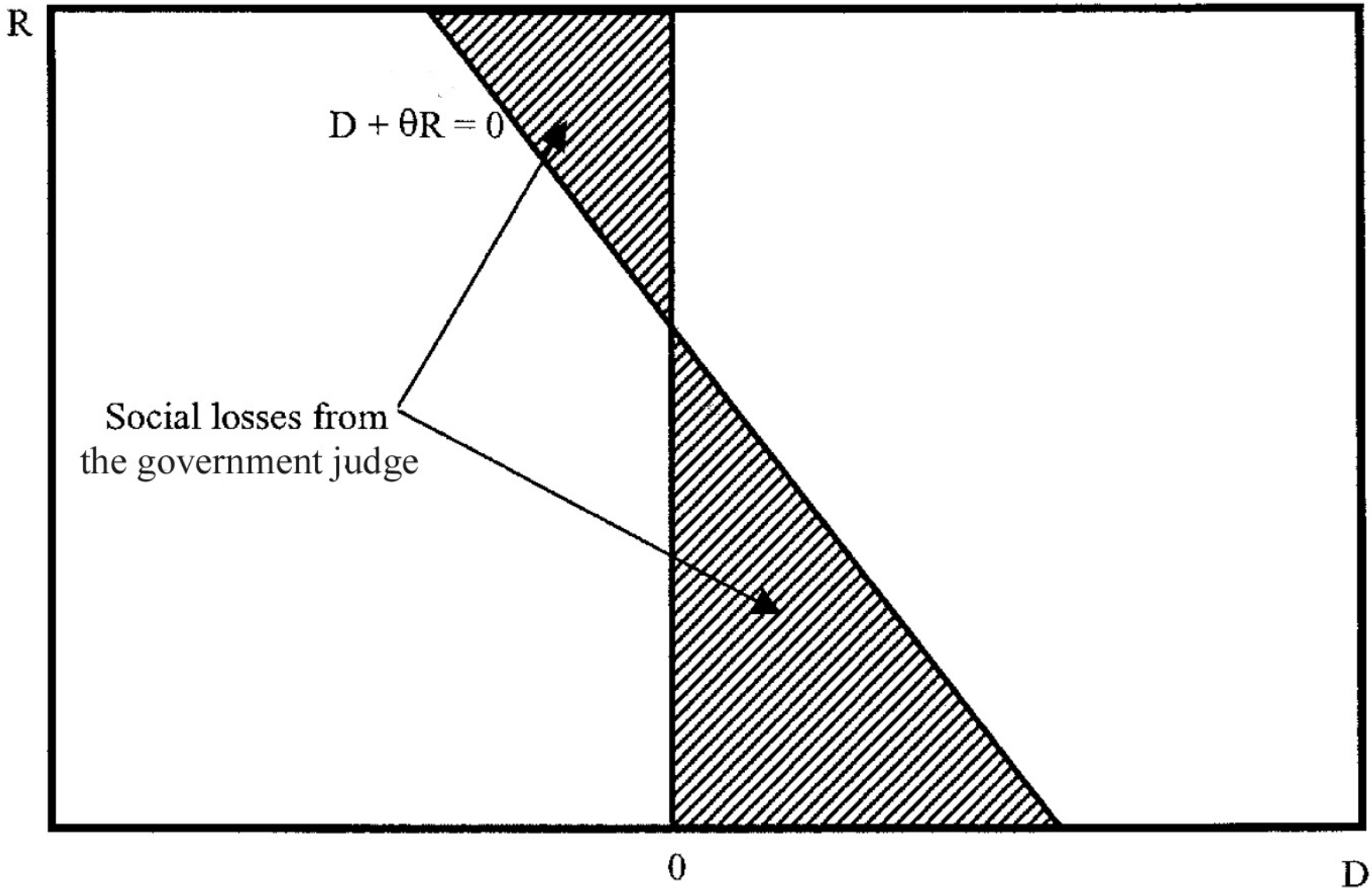
$$\int_0^{\infty}\int_{-D/\theta}^{\infty} DdG(R)\,dF(D) = \int_0^{\infty} D\left[1 - G\left(-\frac{D}{\theta}\right)\right] dF(D).$$

- There exists a threshold $A^*(\theta) > 0$ at which social welfare is the same with independent juries or government-appointed judges.
- For $A > A^*$, government-appointed judges yield higher welfare.
- For $A < A^*$, independent juries yield higher welfare.
- Government-appointed judges are more desirable, the more representative the government: $\partial A^*/\partial\theta > 0$.

Social Losses from Jury Coercion



Social Losses from Government Interference



Bright-Line Rules

- The choice of government-appointed judges creates an incentive problem if the central authorities cannot observe D and R.
- If the judges do not share the government's objective function a priori, it can be optimal to codify a bright-line rule that they are required to base adjudication upon.
- If the government only observes whether $D \geq \bar{D}$, the adoption of bright-line rules means it no longer interferes with justice to pursue its policy goals.
- There remains a trade-off between the advantage of greater information the jury can use, and the cost of subversion of justice by special interests. The results are unchanged for $A^*(\bar{D})$.
- Institutional progress reducing A and $\bar{D}$ generates convergence between the substantive outcomes of the two systems.

The Legal Origins Theory

- The model is a formal example of the Legal Origins Theory: civil law is primarily concerned with protecting individuals from each other, and common law with protecting them from the state.
- The transplant effect follows naturally: government-appointed judges are good in countries with representative government (high θ) but terrible in autocracies (low θ).
- In the origin countries, the respective institutions could result from an efficient Coasian bargain.
- Glaeser and Shleifer stress the medieval origins of legal systems:
  1. Weak barons and a strong monarch in England.
  2. A weak monarch and strong noblemen in France.
- In general, when, where and by whom are Coasian institutional bargains struck?
- Do legal institutions persist after becoming inefficient?

Contracts and Enforcement Risk

- Incentive schemes devised contractually by private parties rely on courts (or arbitration) for enforcement.
- Contract theory is traditionally concerned with verifiability: what outcomes can be observed not only by the parties, but also by the enforcement agency?
- The most common assumption is that some variables are public information and others are technologically impossible to verify (Grossman and Hart 1986).
- Some work on costly state verification (Townsend 1979).
- Gennaioli (2009) studies imperfect verification due to the limits of the court system, which can be both imperfectly informed and biased.
- Gennaioli and Perotti (2009) extend this line of research to the endogenous joint evolution of contracts and their enforcement.

A Financing Relationship

- An investor I finances the project of a pennyless entrepreneur E.
- The cost of the project is $k > 0$.
- Under E-control, the return is r, which equals $\bar{r}$ with probability μ and $\underline{r}$ with probability $1 - \mu$.
- E can pledge at most αr to I: $(1 - \alpha) r$ can be embezzled, or represents and intangible private benefit.
- Under I-control, the return is λ, which can be paid to I in full.
- The project can always be financed, but optimal control is state-contingent:

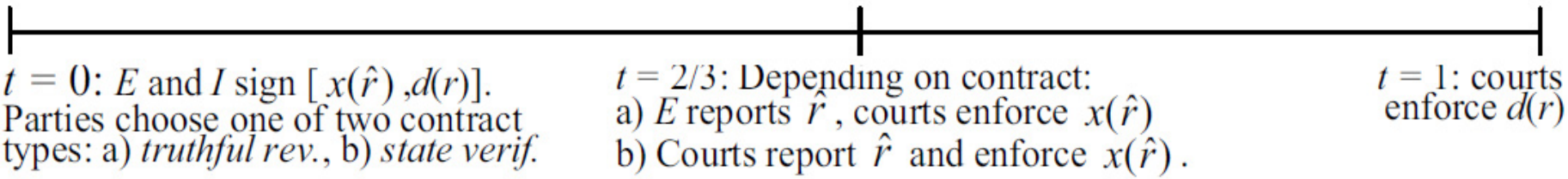
$$\bar{r} > \lambda > \underline{r} > 0 \text{ and } \lambda \geq k.$$

- The first best is attainable with a fully contingent contract:

$$\mu \alpha \bar{r} + (1 - \mu) \lambda \geq k.$$

Imperfect Contracting

- I and E do not need outside enforcement if they use a *truthful revelation* contract in which E reports state $\hat{r} \in \{\bar{r}, \underline{r}\}$ and is given incentives not to lie.
- Under a *state verification* contract, the report is made by a judge instead.
- In either case, the allocation of control rights is given by a probability $x(\hat{r})$ ofr E-control.
- Repayment is $d_E(r) \leq \alpha r$ under E-control or $d_I \leq \lambda$ under I control.
  - Non-embezzled returns are observed ex post.
  - E is judgement proof: no criminal penalties for misreporting.



Judicial State Verification

- Judges observe a noisy signal $s \sim N\left(r, \theta^2\right)$.
  - Higher θ captures lower judicial competence.
- Judges have idiosyncratic biases, and desire E-control if and only if $\beta \mathbb{E}\left(r|s\right) \geq \lambda$, i.e.,

$$s \geq \frac{\bar{r} + \underline{r}}{2} - \frac{\theta^2}{\bar{r} - \underline{r}} \ln\left(\frac{\mu}{1-\mu} \frac{\beta\bar{r} - \lambda}{\lambda - \beta\underline{r}}\right).$$

  - $\beta > 1$ denotes a bias favouring E-control, $\beta < 1$ I-control.
- Moderately biased judges with $\beta \in (\lambda/\bar{r}, \lambda/\underline{r})$ are more affected by their biases when noise θ is larger.
- The unbiased ruling is governed by

$$\eta \equiv \frac{\mu}{1-\mu} \frac{\bar{r} - \lambda}{\lambda - \underline{r}},$$

which captures the expected relative cost of misallocating control.

Judicial Biases

- Analytical tractability is obtained by assuming a distribution for β such that
$$\ln\left(\frac{\mu}{1-\mu}\frac{\beta\bar{r}-\lambda}{\lambda-\beta\underline{r}}\right) \sim N\left(\ln\eta, \sigma^2\right).$$
- $e_I = 1 - p_{\bar{r}}$ denotes the probability of pro-I errors (when $r = \bar{r}$), and $e_E = 1 - p_{\underline{r}}$ the probability of pro-E errors (when $r = \underline{r}$).
- Higher η reduces e_I and increases e_E.
- There are thresholds $\eta_1 \leq 1 \leq \eta_2$ such that higher σ increases e_I for $\eta \leq \eta_1$, e_E for $\eta \geq \eta_2$, and both errors for $\eta \in (\eta_1, \eta_2)$.
- Greater polarization increases the more socially costly error.
- Greater polarization also increases the total number of errors, just like greater noise θ.
- If $\sigma \to \infty$, all judges are so biased that they disregard s: then judgement is a coin toss.

State Verification Contracts

- E asks for financing under the contract

$$\max \mathbb{E}\left\{\omega\left(r\right)\left[r-d_{E}\left(r\right)\right]+\left[1-\omega\left(r\right)\right]\left(\lambda-d_{I}\right)\right\},$$

where the probability that E-control is enforced is

$$\omega\left(\bar{r}\right)=x\left(\bar{r}\right)p_{\bar{r}}+x\left(\underline{r}\right)\left(1-p_{\bar{r}}\right),$$

$$\omega\left(\underline{r}\right)=x\left(\underline{r}\right)p_{\underline{r}}+x\left(\bar{r}\right)\left(1-p_{\underline{r}}\right),$$

and I must break even:

$$\mathbb{E}\left\{\omega\left(r\right)d_{E}\left(r\right)+\left[1-\omega\left(r\right)\right]d_{I}\right\}\geq k,$$

for $x\left(r\right)\in\left[0,1\right]$, $d_{R}\left(r\right)\leq\alpha r$, and $d_{I}\leq\lambda$.

Slack Participation Constraint

- If the break-even constraint is slack, $x(\bar{r})$ and $x(\underline{r})$ have marginal benefits

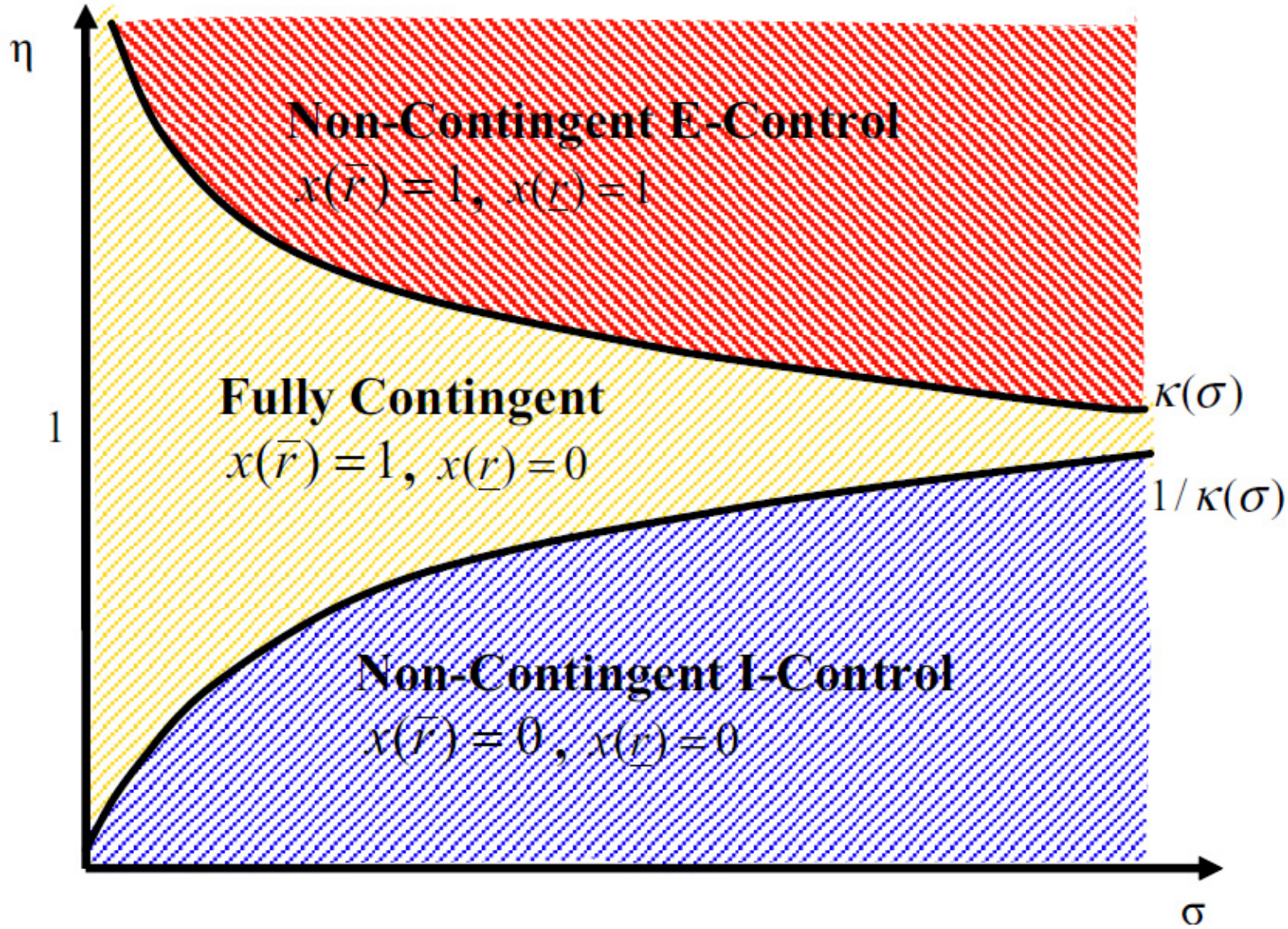
$$MB_{x(\bar{r})} = \eta(1 - e_I) - e_E > MB_{x(\underline{r})} = \eta e_I - (1 - e_E).$$

- The optimal contract is fully state contingent for

$$\frac{e_E}{1 - e_I} \leq \eta \leq \frac{e_I}{1 - e_E},$$

  while it assigns control with certainty to I for lower and to E for higher values of η.
- Trade off between exploiting the signal s and letting judges indulge their biases.
- The cost of judicial partiality is minimal for $\eta \approx 1$, when society is almost indifferent between the two types of error.
- If judges are unbiased, the contract is fully state contingent regardless of θ. More information is always good, no matter how poor its quality, if it is not misused.

State Verification with Slack Participation Constraint



Binding Participation Constraint

- If the participation constraint binds and has a Lagrange multiplier $\Lambda > 1$, the optimal contract has the same form with

$$\eta_b \equiv \frac{\eta + (\Lambda - 1)\frac{\mu}{1-\mu}\frac{\alpha\bar{r}-\lambda}{\lambda-\underline{r}}}{1 + (\Lambda - 1)\frac{\lambda-\alpha\underline{r}}{\lambda-\underline{r}}} < \eta$$

  replacing η.
- Pro-E errors are more costly because they also undermine repayment.
- Courts then also have a time-inconsistency problem:
  - Contracts specify I-control when it is necessary ex ante.
  - Unbiased judges prefer E-control when it is efficient ex post.

$\Rightarrow$ It is no longer guaranteed that unbiased courts are always used.

- Some pro-I bias could help counterbalance dynamic inconsistency.

Truthful Revelation Contracts

- Any $x(\bar{r}) > x(\underline{r})$ can be implemented by truthful revelation by satisfying the incentive-compatibility constraint

$$\bar{r} - d_E(\bar{r}) \geq \lambda - d_I \geq \underline{r} - d_E(\underline{r}).$$

- Incentive compatibility is consistent with a maximum repayment

$$d_E(r) = \alpha r \text{ and } d_I \geq \lambda - (1-\alpha)\underline{r}.$$

- The participation constraint is simultaneously satisfied if

$$\mu x(\bar{r})\alpha\bar{r} + [1 - \mu x(\bar{r})][\lambda - (1-\alpha)\underline{r}] \geq k$$

- The first best is attainable if

$$\mu\alpha\bar{r} + (1-\mu)[\lambda - (1-\alpha)\underline{r}] \geq k,$$

- Truthful revelation succeeds when it is difficult to tunnel resources from a struggling company, i.e., when $(1-\alpha)\underline{r}$ is low.
- We can assume parameter conditions such that no truthful revelation contract is feasible when the first best is not.

Optimal Contracts with Enforcement Risk

- Focus on the participation constraint by assuming $\eta = 1$.

There exists a threshold $\lambda^* \in (\underline{r}, \bar{r})$ such that:

1. If $\lambda \geq \lambda^*$ the parties write a truthful revelation contract.
2. If $\lambda < \lambda^*$ the parties write a state verification contract.
   There are two thresholds $\tilde{\lambda} \leq \lambda^*$ and $\sigma^* \geq 0$ such that:
   1. If $\lambda \geq \tilde{\lambda}$ or $\sigma < \sigma^*$, parties set $x(\bar{r}) = 1$, $x(\underline{r}) = 0$.
   2. If $\lambda < \tilde{\lambda}$ and $\sigma \geq \sigma^*$, parties set $x(\underline{r}) = 0$, but to attain break-even they must set $x(\bar{r}) < 1$, decreasing in σ and such that $x(\bar{r}) = 0$ for $\sigma \to \infty$.

- If E's informational rent is large, judicial enforcement is required to break even.
- A fully state-contingent contract is possible if the cash flow under I-control is sufficiently large and judges are not too biased.
- Excessive judicial polarization jeopardizes break-even and induces the use of less flexible contracts, leading to underinvestment and welfare losses.

A Supply Relationship

- A seller S and a buyer B contract over the supply of a relationship-specific widget whose outside value is nil.
- The transaction is characterized by its size $\bar{v} \in [0,1]$, which varies in the population with density $f(\bar{v})$.
- The timing of the relationship is the following.

1. S must undertake an unobservable investment $k\bar{v}^2$ to acquire relationship-specific human capital, where $k \in (0, 1/6)$.
2. B's valuation v is realized from a uniform distribution on $[0, \bar{v}]$.
3. S exerts unobservable effort e at a cost $e^2/2$. With probability e the widget is successfully produced.

Incomplete Contracts

- The efficient level of effort is

$$e_{FB}(v) = \arg\max_{e} \left\{ ev - \frac{1}{2}e^2 \right\} = v.$$

- Expected surplus in the first best is

$$W_{FB}(\bar{v}) = \frac{1}{2}\mathbb{E}\left(v^2|\bar{v}\right) - k\bar{v}^2 = \left(\frac{1}{6} - k\right)\bar{v}^2.$$

- The first best can be achieved by a contract that specifies ex ante a contingent price $p(v) = v$.
- The standard problem: v need not be verifiable by courts, so fully contingent contract may be unavailable ex ante.
- Gennaioli and Perotti (2009) assume that bargaining is impossible after the realization of v.

Imperfect Enforcement

- In a trial, B tries to convince the court that the value is low and S that it is high.
- S wins if and only if $v/\bar{v} > \beta$, where $\beta \in [0,1]$ denotes B's relative power.
- The coarseness of state verification limits the parties to a contract with a base payment p and a bonus Δ, which induce effort

$$e_{LF}(v) = \begin{cases} p & \text{if} \quad v \leq \beta\bar{v} \\ p+\Delta & \text{if} \quad v > \beta\bar{v} \end{cases}.$$

- The optimal laissez-faire contract is

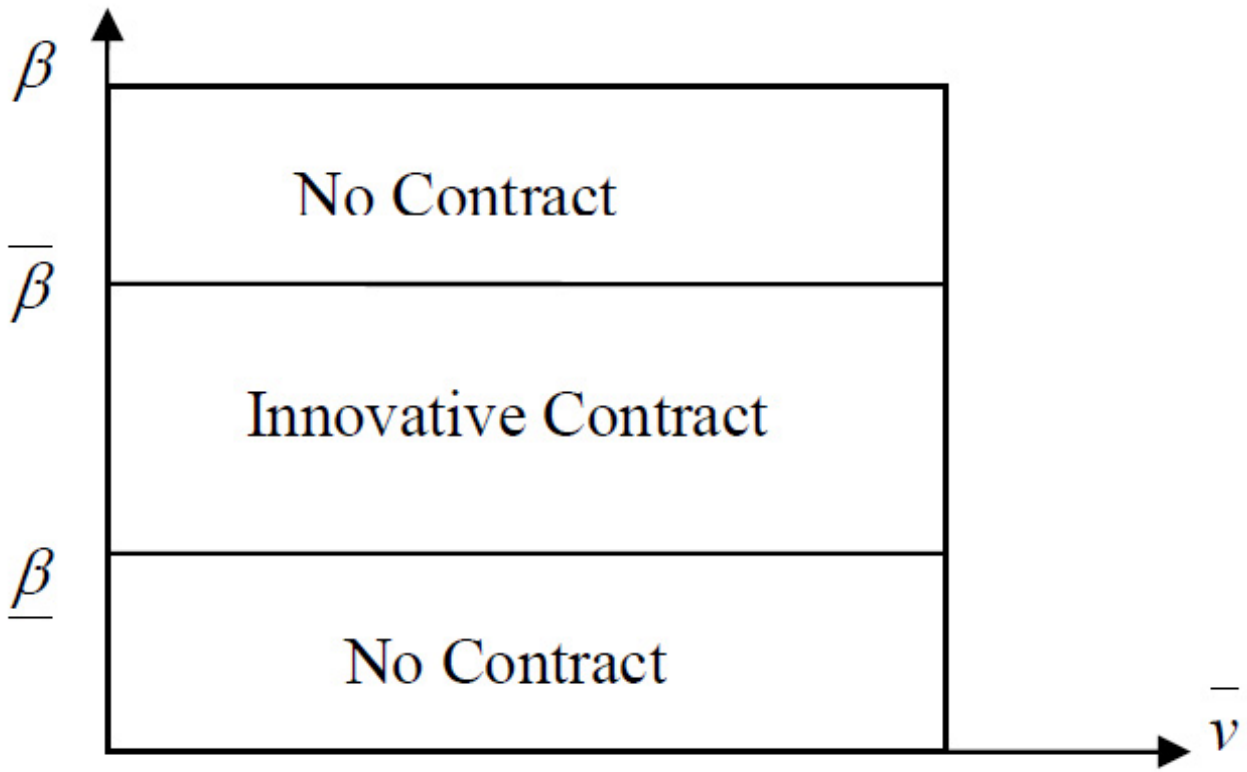
$$p = \frac{1}{2}\beta\bar{v} \text{ and } \Delta = \frac{1}{2}\bar{v}.$$

- Expected surplus with this contract is

$$W_{LF}(\bar{v},\beta) = W_{FB}(\bar{v}) - \frac{1-3\beta+3\beta^2}{24}\bar{v}^2.$$

Static Contracting under Laissez Faire

- As inequality $|\beta - 1/2|$ increases, enforcement problems make the contract less and less state-contingent and reduce its value.
- For $k \in [1/8, 5/32]$ there are two thresholds $\underline{\beta}$ and $\bar{\beta}$ such that B and S contract if and only if $\beta \in \left[\underline{\beta}, \bar{\beta}\right]$.



Contract Standardization

- A standard contract is less exposed to enforcement risk: the court verifies if $v > v_s$, independent of β.
- The drawback of standard contracts is that one size must fit all: v_s is also independent of $\bar{v}$.
  - The standard contract is useless for transactions of size $\bar{v} \leq v_s$.
- Given verifiability v_s, the optimal standard contract is

$$p = \frac{1}{2} v_s \text{ and } \Delta = \frac{1}{2} \bar{v}.$$

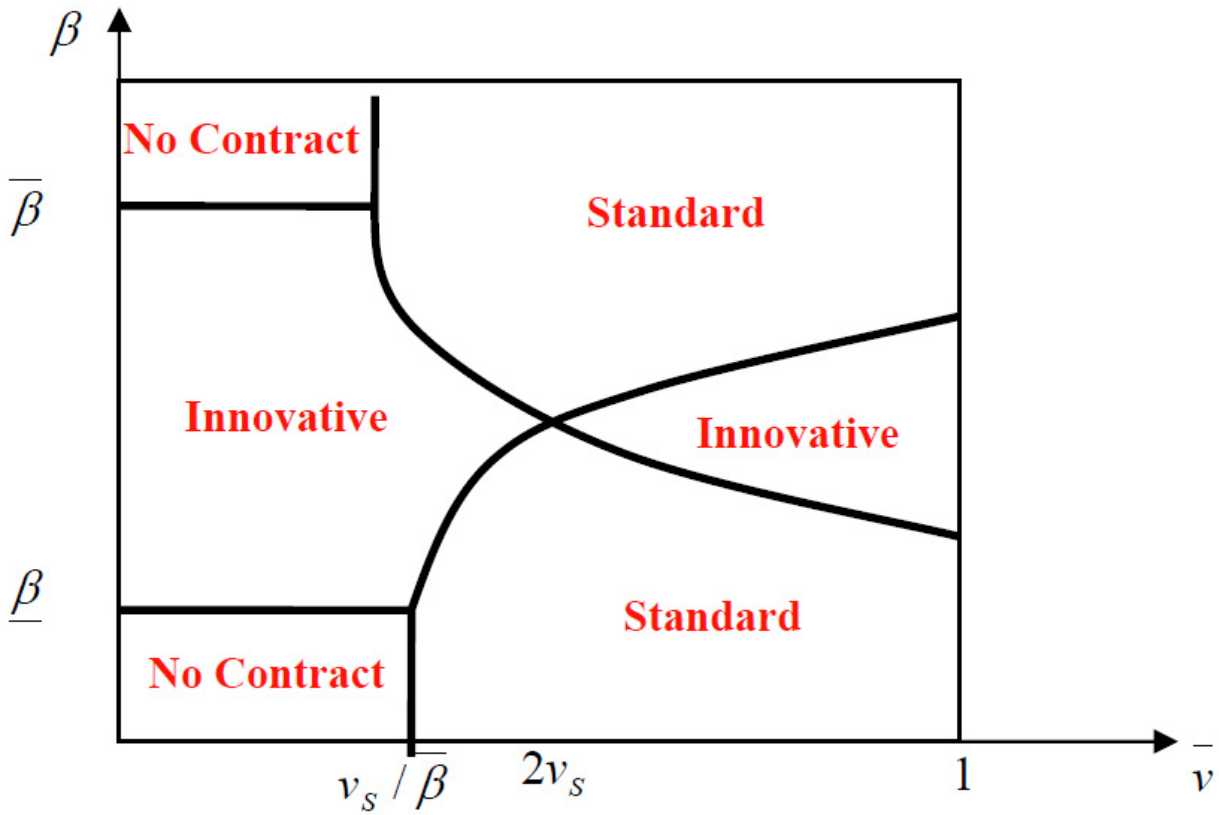
- Expected surplus with this contract is

$$W_{SC}(\bar{v}, v_s) = W_{FB}(\bar{v}) - \frac{\bar{v}^2 - 3v_s\bar{v} + 3v_s^2}{24}.$$

- Standard contracts cannot be customized to create a three-fold incentive scheme that conditions on both v_s and $\beta\bar{v}$.

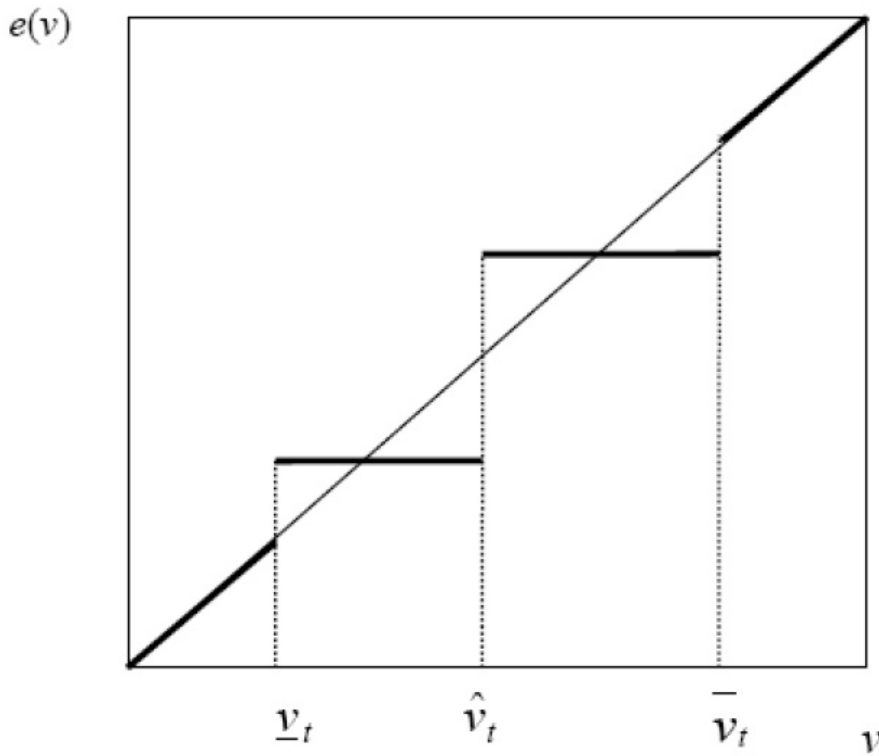
Static Contracting with Standardization

- Standardization increases social welfare in a static environment:
  1. Unequal parties that could not contract under laissez faire can do so.
  2. Parties with expected value $\bar{v}/2$ close to v_s improve their contracts.



Precedents and Contracts

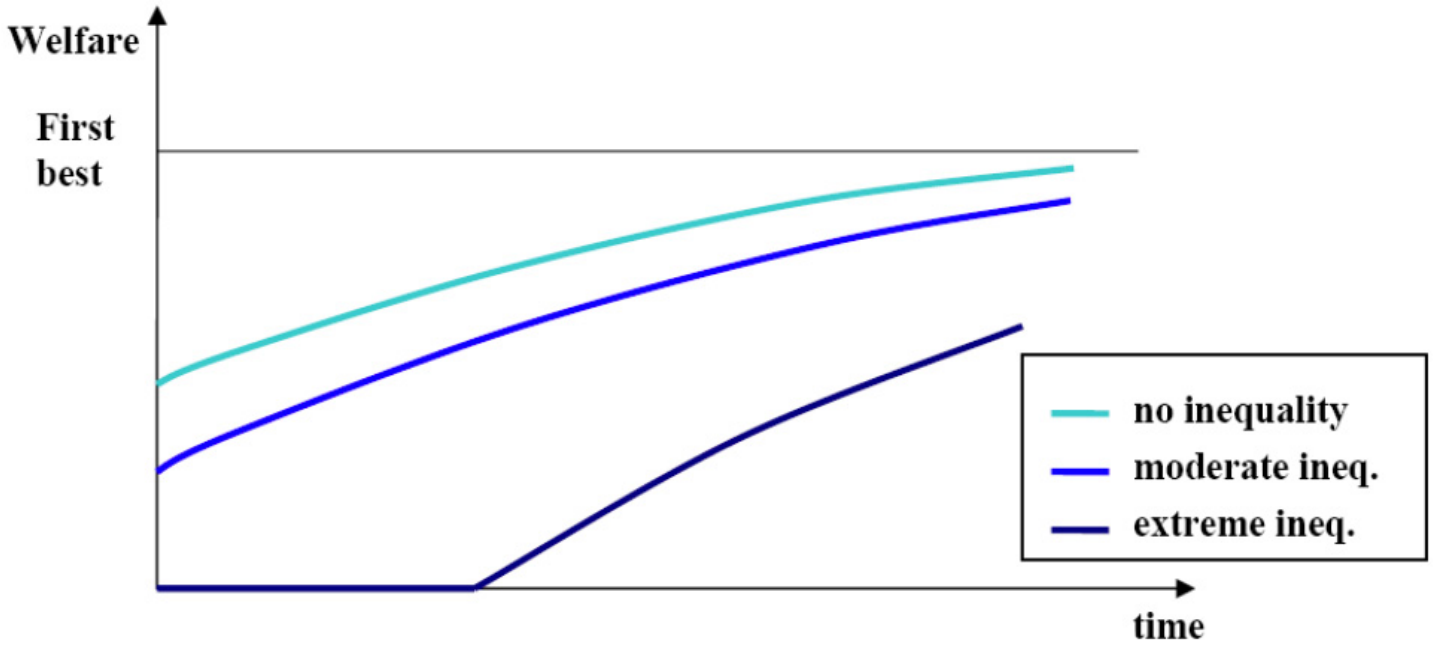
- Under laissez faire, case law teaches courts how to recognize when the value is high or low.
  - Extreme values $v \leq \underline{v}_t$ and $v \geq \bar{v}_t$ are perfectly verifiable at time t.
  - $(\underline{v}_t, \bar{v}_t)$ is the area of uncertain enforcement in which only a two-part price can be used.
  - S wins if and only if $v > \hat{v}_t \equiv (1-\beta)\,\underline{v}_t + \beta \bar{v}_t$



The Evolution of Contracting under Laissez Faire

- Given legal uncertainty $g_t \equiv \bar{v}_t - \underline{v}_t$, parties contract if $\beta \in \left[\underline{\beta}_t, \bar{\beta}_t\right]$, with $\underline{\beta}_t$ decreasing and $\bar{\beta}_t$ increasing in g_t.
- Litigation of innovative contracts creates precedents: $\dot{g}_t = -g_t \left[F\left(\bar{\beta}_t\right) - F\underline{\beta}_t\right]$.

$\Rightarrow$ The joint evolution of contracts and case law converges to the first best $g_\infty = 0$.



Legal Evolution under Standardization

- The introduction of a standard contract immediately increases the volume of trade.
- Innovative contracts are used more rarely than under laissez faire

$\Rightarrow$ Legal evolution is slower, due to the lower rate of litigation of innovative contracts.

- The result is true even if the standard contract is continuously updated. In fact, the long-run cost is the larger, the larger the short-run benefits.
  - A more efficient standard contract ($v_s \approx 1/2$) crowds out innovation more strongly.
- There exists a threshold t^* such that social welfare at time t is higher under standardization than under laissez faire if and only if $t < t^*$.
- The short-run benefits of standardization last longer in more unequal societies: t^* is increasing in the variance of β.

Dynamic Consequences of Contract Standardization