Item # 3
SEMINAR IN LAW, ECONOMICS, AND ORGANIZATION RESEARCH
Professors Louis Kaplow & Oliver Hart

Monday, March 13
**Pound 108**, 12:30 p.m.

“Case Law vs. Statute Law: an Evolutionary Comparison”

Patricio Fernandez and Giacomo Ponzetto
Case Law vs. Statute Law: 
An Evolutionary Comparison*

Preliminary draft: please do not quote without permission

Giacomo A. M. Ponzetto Patricio A. Fernandez
Harvard University Harvard University
March 13th, 2006

Abstract

Recent evidence has shown the correlation of common law with a number of positive economic outcomes. We present a model of the evolution of case law that bears out Cardozo’s and Posner’s hypothesis of its convergence towards efficiency. Since statutes do not share this evolutionary property, common law is the more efficient system in a stationary environment: the difference between the two is particularly stark in less democratic societies. Historically, the emergence of cases law in England and statute law on the Continent can be explained by the different balance of power between the crown and feudal lords. For dynamic contemporary societies, under very general conditions the optimal system is mixed, with legislation dealing with sudden changes but jurisprudence constantly refining the law. This prediction fits with the observation of convergence of the two main legal families.

*We are grateful to Alfred Galichon, Nicola Gennaioli, Edward Glaeser, Anthony Niblett and particularly Louis Kaplow and Andrei Shleifer for helpful comments. Ponzetto acknowledges financial assistance from the Marco Fanno Foundation and Mario Gasbarri Foundation. e-mail: ponzetto@fas.harvard.edu, and pfernand@fas.harvard.edu.
1 Introduction

Recent empirical evidence has shown that countries within the common law tradition that originated in England substantially outperform civil law countries whose legal system sprang up from the Continental European tradition—especially French—in a number of financial, political and economic indicators (La Porta et al. 1997, 1998, 1999, 2002; Mahoney 2001). The reasons for these systematic differences are not yet entirely clear, nor, in general, are the causal mechanisms that link the legal system with economic performance.

Discussion about the characteristics of alternative legal systems and their economic implications has a long history, but so far seems to have received little attention from economists. Different intellectual traditions that go back to the very origins of Western legal thought and political theory have, on the one hand, praised and defended the almost spontaneous character of common law as a gradual process of accumulation of social and political wisdom, which some have seen as akin to the action of nature (cf. Coke, Hale, Hayek); and, on the other, emphasized the need for direct and explicit rational control over the legal process so as to ensure that its evolution conforms to the designs of the legitimate bearer of political power in society, taking into account the good of society (cf. Hobbes, Rousseau, Austin).

The most prominent theory supporting case law in the eyes of economists is surely the hypothesis of its efficiency advanced by Posner (1973), which in the legal literature is usually associated with Cardozo (1921) and his description of an evolutionary process of successive judges who mutually offset each other’s biases: a process by virtue of which “the bad will be rejected and cast off in the laboratory of the years,” leading to legal outcomes that are more uniform, constant, and of “greater value” than the individual judicial decisions considered in isolation.

In spite of this, there is ample evidence of an increasing use of statutes, mainly for the purposes of economic regulation, even in countries with deep-rooted common law legal traditions, in which cases, as opposed to written law, were for centuries the main source of law (cf. Pound 1908, Zweigert and Kötz 1998, Glaeser and Shleifer 2003). Legal scholarship has in fact been emphasizing that legal systems that had been traditionally understood as dissimilar, not only as to their historical and cultural origins, but also with regard to their
procedural and substantive conceptions of the legal process, have nonetheless significantly converged both in terms of final legal outcomes and of significant dimensions of their modus operandi (Coffee 2000, Zweigert and Kötz 1998, Glendon 1984, Glendon et al. 1999).

In this paper we present a simple mathematical model of the evolutionary properties of case and statute law that allows us to evaluate these conflicting claims and more generally to study the optimality of different legal arrangements depending on the different characteristics of a society.

Our premise is the existence of two structurally different technologies of law-making: statute law and case law, that have traditionally characterized respectively the civil and common law legal families.

The fundamental characteristic of judge-made law is that evolves over time in a precisely defined manner, owing to the incessant process of minor distinguishing of settled precedent by different judges. As a consequence, case law does exhibit a convergence property: although it is constantly evolving, and thus never settles on the precisely optimal rule, it provides long-run certainty. We also show that this process is hindered by judicial activism, construed as an explicit intention on the part of the judge to write the law, rather than to decide correctly the case before his court. On the other hand, it is helped by greater efficiency of the court system in going through its case load, and it could be further enhanced by endogenous emergence of cases with greater intensity when the law is less efficient.

On the contrary, civil law has no intrinsic dynamic properties, and its improvement over time can only be due to exogenous enhancements of the electoral process and democratic representativeness. Therefore, in a static setting, the comparison between the two systems can be best pictured as a trade off between the short-run certainty of written statutes and the long-run certainty of common law, a point made in particular by Bruno Leoni (1961).

Following Glaeser and Shleifer (2002) we explain the historical emergence of different legal systems in England and on the European continent based on the relative power of noblemen and monarchs: by extending their analysis, we provide an explanation of the simultaneous evolution of jury trial and stare decisis in England, and the inquisitive trial and written statutes in France.

Beyond the formative stages of a legal system, our model presents strong support for
the view that common law, due to its evolutionary properties, is superior to legislation in a stationary setting. Nonetheless, once social change enters the picture, *stare decisis* can hinder the adaptation of legal institutions to changed circumstances, and therefore become a burden for society. This can explain the particularly keen use of statutes in times of great upheaval, such as the end of Feudalism coinciding with codification in Europe, and the increase in legislative activity in the United States at the end of the nineteenth century.

We conclude, however, that the optimal system is not pure statute law even in the presence of continuous change in social conditions. A mixed system that retains a lawmaking role for judicial decisions, but integrates it with legislation to deal with changed underlying conditions, is proved to be superior to either pure legal arrangement under very general conditions.

### 1.1 A simple model of the law

We will represent the law, or more specifically the rule regulating a given legal issue, by a point on the real line $x \in \mathbb{R}$. Every member $i$ of the population has individual preferences that are additively separable and depend on this legal rule according to the loss function

$$l_i(x) = (x - x_i)^2$$

The bliss points $x_i$ are *i.i.d.* with continuous distribution $F$ on $\mathbb{R}$ with mean $x^*$ and variance $\sigma^2$, which we will initially assume to be constant over time. Social welfare resulting from the law is then described by the loss function

$$l(x) = \int_{-\infty}^{+\infty} (x - x_i)^2 dF(x_i) = \sigma^2 + (x - x^*)^2$$

While the first term captures an intuitive effect—heterogeneity is costly since a single rule has to be adopted for all society—we take the variance of preferences in the population as exogenous and therefore normalize

$$l(x) = (x - x^*)^2$$

where $x^*$ characterizes the socially optimal law, e.g. the allocation of liabilities according to Hand’s Rule.
In a stochastic and dynamic setting, the loss function is assumed to take the standard form of intertemporally separable expected utility:

$$L_t = \mathbb{E}_t \left[ \sum_{s=0}^{\infty} \beta^s l (x_{t+s}) \right] = \sum_{s=0}^{\infty} \beta^s \left[ (E_t (x_{t+s}) - x^*)^2 + Var_t (x_{t+s}) \right]$$

where $x_t$ is the effective law implemented in period $t$. This captures the two fundamental concerns of society: the tendency of legal rules to conform to the optimum every period, and their ex ante predictability. In statistical terms, these are respectively the accuracy and the precision of the process; from the point of view of legal intuition, they capture the equity and the certainty of the law.

2 Case law converges towards efficiency

In a common-law system, the law evolves through the rulings of judges, who have a limited opportunity to deviate from established precedent. Legal principles arise and are modified chiefly through the adjudication of particular cases by independent judges. Common law is essentially case law and has an inductive nature: general legal principles are derived primarily from the solution of a variety of individual cases. As early as the fourth century B.C., Aristotle defended the superiority of legislative over judicial law-making on the grounds that “the decision of the lawgiver is not particular but prospective and general, whereas members of the assembly and the jury find it their duty to decide on definite cases brought before them.” (Aristotle Rhetoric). Even if we disagree on the normative judgment, the diagnosis seems essentially correct.

Moreover, the principal source of law is precedent, by which previous decisions have normative value for subsequent adjudication: “the vast number of volumes of “unwritten” law is the foremost distinguishing feature of the common law tradition” (Glendon et al. 1999, p. 262). In this sense, the judge is not expected merely to apply a preexistent law to a particular case, but rather he must create —or “discover”, in the language preferred by some of the greatest advocates of the common law tradition—on his own the appropriate law, tailored to the case at hand. The legal decision-maker is, foremost, the judge, but his “legislative” power is limited to the adjudication of concrete disputes and does not extend to the promulgation
of abstract and general rules of broader scope. The U.S. Constitution explicitly defines in Article III that the scope of the judicial power extends only to “cases” and “controversies.” (Schauer 2005) In the American legal system, the courts will decide the constitutionality of a law only when that are not only in force, but have been brought to bear in a specific instance. Furthermore, only a party who has suffered from the law’s application is entitled to challenge its constitutionality—i.e. only a party who has the necessary “standing.”

Judges are a heterogeneous group whose individual preferences \( x \) have distribution \( F \) on \( \mathbb{R} \), symmetric around its mean \( x^* \) and having variance \( \sigma^2 \): this heterogeneity can be seen as capturing a combination of scholarly disagreement about different theories of justice, partisan ideology, and even (at least in some countries) corruption and the influence of special-interest groups.

Each judge’s utility function has two components: one is the loss from the difference between the actual law and his bliss point; the other is a cost of deviating from precedent, defined as the law inherited from the previous period. Specifically, we assume that the cost is also quadratic: this could be seen as an analytically tractable second-order approximation to any cost function that captures the core idea that a little amount of distinguishing is almost costless, but fundamental over-ruling is much more costly and therefore rarer.\(^1\)

We follow Posner (2003) in stressing the cost of legal innovation in terms of personal effort. A judge who abides by precedent can use information that has already been generated by previous cases and economize on new information gathering. A decision that drifts away from precedent requires harder work from the judge investigating the novel aspects of the case and the legal arguments connecting them to his ruling.

This effort cost also partially reflects social norms concerning decisions in accordance to precedent. Judges who deviate from previous rulings are subject to more stringent professional criticism by their peers and to scrutiny by extra-judicial institutions and the general public. As a consequence, they are required to make a strong case for their distinguishing, which amounts to an exercise in legal scholarship. Stare decisis instead requires no more than the mention of relevant previous decisions, which has historically been less costly thanks

to the availability of reliable law reports.²

We initially assume that the judge is entirely focused on the case he is deciding, so that his loss function is

\[ l_j = (x_t - \hat{x}_t)^2 +ug(x_t - x_{t-1})^2 \]

with \( c \geq 0 \), which can be conveniently normalized as

\[ l_j = (1 - \psi)(x_t - \hat{x}_t)^2 + \psi(x_t - x_{t-1})^2 \]

where \( \psi = \frac{c}{1+c} \in [0,1] \) represents the institutional importance of precedent, increasing in the cost of deviating from it.

This yields directly the following key result:

**Proposition 1** Case law evolves according to the Markov chain

\[ x_{t+1} = \psi x_t + (1 - \psi) \hat{x}_{t+1} \]

Hence, given any starting point \( x_t \), the conditional distribution of case law after \( s \) decisions has expectation

\[ E(x_{t+s}|x_t) = \psi^s x_t + (1 - \psi^s) x^* \]

and variance

\[ Var(x_{t+s}|x_t) = \frac{1 - \psi}{1 + \psi} \left( 1 - \psi^{2s} \right) \sigma^2 \]

Thus case law converges at rate \(-\ln \psi\) to the asymptotic distribution

\[ x \sim N \left( x^*, \frac{1 - \psi}{1 + \psi} \sigma^2 \right) \]

The proposition captures Cardozo’s intuition on the long-run efficiency of the common law: different judicial biases counterbalance one another over time, making convergence towards an optimal rule possible. However, case law does not deterministically reach the optimal rule, and indeed it will never settle on an immutable rule: this mirrors the conclusion of Gennaioli and Shleifer (2005), who, in a different framework, obtain that pointwise convergence to efficiency is possible only when all judges are efficiency-maximizers.

²Cf. Dawson (1968).
Instead, the evolution of precedent converges to an asymptotic distribution whose variance around the efficient solution is always smaller than the variance of judicial opinions, and can be arbitrarily small if precedents are very strongly binding. The intrinsic trade-off is captured by the speed of convergence, which is intuitively decreasing in the importance of 

stare decisis: thus a legal system with a high importance of precedent will eventually have excellent case law, but the transition may take a long time.

The increasing efficiency of common law can also be seen graphically by plotting the expected loss function as seen from an ex ante perspective: this can be interpreted as the perspective of a hypothetical benevolent planner considering social welfare before the very creation of a jurisprudential system, that is to say behind a veil of ignorance as invoked by Rawls (1999). Since the first judge cannot constrained by stare decisis, given that no precedents exist, he will rule according to his notion of equity, and therefore the expected loss from his ruling will be \( E [l(x_1)] = E [(\hat{x} - x^*)^2] = \hat{\sigma}^2 \); thereafter, the evolution of precedent insures convergence towards efficiency in the sense that

\[
E [l(x_t)] = E \left[ E \left[ (x_t - x^*)^2 | x_1 \right] \right] = \frac{1 - \psi + 2\psi^{2t-1}}{1 + \psi} \hat{\sigma}^2
\]

is monotone decreasing in \( t \).

By plotting two different values of \( \psi \), the graph also conveys visually the aforementioned trade-off between speed of convergence and asymptotic efficiency.
2.1 Judicial activism

While we have suggested that the judge should be focused on deciding the case facing him, it is natural to consider that high-court judges whose rulings constitute binding (though not immutable) precedents could take into account when deciding a case the impact of their decision on the future evolution of the law. In other words, judges could consciously set out to write the law: we take it that this what is popularly meant by judicial activism.

In fact, our analysis confirms and simultaneously qualifies the conventional wisdom that has given the term judicial activism a negative connotation. In a common-law system, judges are in fact continually refining the law: their rulings should be based on settled precedents, but also incorporate, albeit in small part, each judge’s own view of the law. Thus, judges cannot be the mere automata envisaged by Napoleon; however, their activity should ideally be humbly focused on deciding correctly the case before them, and not presume purposefully to write the law of the land.

An explicit concern for one’s legacy on the law does not remove the fundamental convergence properties of case law, but it introduces needless uncertainty in the evolution of precedent. Formally we can prove the following:

**Proposition 2** If judges are concerned about the influence of their decision on the future evolution of the law, with a subjective discount factor \( \beta_j \in [0, 1] \), then case law evolves according to the Markov chain

\[
x_{t+1} = ax_t + (1 - a) \hat{x}_{t+1} + \frac{\beta_j a (1 - a)^2}{1 - \beta_j a} (\hat{x}_{t+1} - x^*)
\]

Hence, given any starting point \( x_t \), the conditional distribution of case law after \( s \) decisions has expectation

\[
E_t(x_{t+s}) = a^s x_t + (1 - a^s) x^*
\]

and variance

\[
Var_t(x_{t+s}) = (1 - a^{2s}) \frac{1 - a}{1 + a} \left( \frac{1 - \beta_j a^2}{1 - \beta_j a} \right)^2 \hat{\sigma}^2
\]

so that case law converges at rate \(-\ln a\) to the asymptotic distribution

\[
x \sim N \left( x^*, \frac{1 - a}{1 + a} \left( \frac{1 - \beta_j a^2}{1 - \beta_j a} \right)^2 \hat{\sigma}^2 \right)
\]
The factor of convergence is a function of judges’ incentives \( a(\psi, \beta_j) \in (0, 1) \) such that

\[
\frac{\partial a}{\partial \beta_j} < 0 \land \frac{\partial a}{\partial \psi} > 0
\]

Moreover, the certainty of the law over any span of time is monotonically decreasing in the amount of judicial activism

\[
\frac{\partial \text{Var}_t(x_{t+s})}{\partial \beta_j} > 0 \forall s \geq 1
\]

The standard model with judges focused on the case at hand is naturally obtained in the limit as

\[
\beta_j = 0 \Rightarrow a = \psi
\]

More generally, the intuition behind the varying persistency of precedent (as measured by \( a \)) is straightforward: the more forward-looking a judge is (the higher \( \beta_j \)), the more he is willing to make the effort to deviate from precedent, since he anticipates that his ruling will not only decide the case at hand, but marginally affect the future evolution of case law. On the other hand, the higher the cost of deviation from *stare decisis*, the smaller such deviations will be. At one extreme \( \psi = 0 \Rightarrow a = 0 \), which is intuitive since then every judge independently and arbitrarily decides the case at hand based on his own preferences; while at the other \( \psi \to 1 \Rightarrow a \to 1 \forall \beta_j \in [0, 1] \), because then deviations from precedent are infinitely costly. It should be noted that this confirms our previous result that the asymptotic variance of case law can be reduced arbitrarily, albeit at the expense of the speed of convergence: \( \lim_{\psi \to 1} V(x) = \lim_{a \to 1} V(x) = 0 \).

While intuitively clear, the change in the factor of convergence \( a \) is not the main effect of judicial activism. In fact, quantitatively, \( a \) always remain remarkably close to \( \psi \): the graph below shows the exact area where \( a(\psi) \) can range as \( \beta_j \) varies in \([0, 1]\): it is largest for \( \psi \approx 0.86 \) where \( a \in (0.74, 0.86] \).\(^3\)

\(^3\) *En passant*, while \( a(\beta_j, \psi) \) does not have a closed-form expression, we can solve both \( \beta_j(\psi, a) = \frac{\psi - a}{a^2(1 - a)} \) for \( 0 \leq a \leq \psi < 1 \) and \( \psi(\beta_j, a) = \frac{a}{1 - \beta_j a^2(1 - a)} \) for \( \beta_j, a \in [0, 1] \).
The most important implication of judges’ concerns for their imprint on future law ($\beta_j > 0$) is the introduction of a strategic element in the judge’s decision process: since the judge expects future rulings on average to tend to revert towards the mean, he is led to be more extreme in order to have a longer-lasting impact on the law.

The model sheds some light on the empirical observation that popular commentaries typically relate activism to a judge’s partisan legal philosophy or ideology. If the causation ran from ideology to activism, we would be left with the puzzle that conservative commentators accuse liberal judges of having an activist stance, while liberal commentators make the same criticism about conservative judges. Needless to say, we cannot rule out in principle that activism is the only common point between two radically different judicial philosophies; but our model suggests a simpler explanation for the observed correlation between partisanship and activism: it is in fact greater activism (higher $\beta_j$) that increases judge’s observed partisanship, namely their tendency to make more extreme rulings, as captured by the variance of judicial decisions.

### 2.2 The timing of judicial decisions

We have described the evolution of the common law as a function of the sequence of judicial rulings by the high court whose precedents are binding for all judges in its jurisdiction. In order to make proper considerations about social welfare and the evolution of common law
as a function of objective calendar time, it remains to explain when such rulings occur.

A simplified but far from unrealistic model sees the primary constraint as technological: the court system operates at, or indeed above, maximum capacity, and the number of cases decided is a function of the number of judges and their efficiency in rapidly reaching a verdict. As a consequence, we can define the fundamental atom of time in our model as the period it takes the high court to process a case. This also creates a natural parallel between the two notions of efficiency: if the legal system can process a greater number of cases (without of course sacrificing the quality of judicial reasoning), the evolution of case law will be faster and the trade-off between speed and eventual optimality will be relaxed. Djankov et al. (2003) show that the quality of courts, measured as the estimated duration of dispute resolution, is highly negatively correlated with procedural formalism (which is systematically greater in civil law countries), but does not respond to the incentives facing the judges.

Of course, it could be objected that, for instance, the United States Federal Courts of Appeals process several thousands of cases a year, and if the process of distinguishing we have outlined above applied to every single one of them, the evolution of case law should be astonishingly rapid.

From a purely mathematical point of view, we feel bound to remark that the model is entirely consistent with this interpretation; in fact, a system that literally had tens of thousands of precedent-setting decisions every year would have extremely appealing dynamic properties: precedent would be almost perfectly binding ($\psi \approx 1$), and yet through a series of extremely frequent but individually imperceptible deviations, the law would converge quite rapidly to a distribution with infinitesimal variance around the optimum.

However, such a rosy view of case law is surely not a faithful depiction of reality. Conceptually, precedents should be seen as following each other, as it is indeed the case for single-bench supreme courts such as the United States Supreme Court or the House of Lords, which naturally issue a fairly limited numbers of opinions every year. Although the Ninth Circuit does decide about five thousand cases a year, it is obvious that one cannot think of an orderly series of precedents following each other at intervals of less than half an hour.

The point is that a majority of lawsuits brought before the courts, even on appeal, will be
primarily factual and break no new legal ground: that is to say they will offer no rationale for distinguishing. This can be simply captured by assuming that every period there is probability \( p \) that a non-trivial case is presented to the court, enabling it to contribute, albeit at some effort cost, to jurisprudential evolution. Such an assumption introduces some analytical complication, and of course an added source of uncertainty in the evolution of the law, namely uncertainty in the timing of relevant precedents. However, it does not alter our conclusions in any substantial way: Proposition 1 can be restated as:

**Corollary 1** If every period the high court has probability \( p \) of being able to refine the existing law, case law evolves according to the Markov chain

\[
x_{t+1} = \begin{cases} x_t & w/ \Pr 1 - p \\ \psi x_t + (1 - \psi) \hat{x}_{t+1} & w/ \Pr p \end{cases}
\]

Hence, given any starting point \( x_t \) the conditional distribution of case law after \( s \) periods has expectation

\[
E(x_{t+s}|x_t) = (1 - p + p\psi)^s x_t + [1 - (1 - p + p\psi)^s] x^*
\]

and variance

\[
Var(x_{t+s}|x_t) = \frac{1 - \psi}{1 + \psi} \left[ 1 - (1 - p + p\psi^2)^s \right] \sigma^2 + \left[ (1 - p + p\psi^2)^s - (1 - p + p\psi)^{2s} \right] (x_t - x^*)^2
\]

Thus case law converges at rate \(-\ln(1 - p + p\psi)\) to the asymptotic distribution

\[
x \sim N \left( x^*, \frac{1 - \psi}{1 + \psi} \sigma^2 \right)
\]

Hence, the stylized assumption of a deterministic timing of decisions is essentially equivalent to a more realistic stochastic formulation with a constant hazard rate.\(^4\)

\(^4\)This equivalence can be made absolutely rigorous from the point of view of social welfare. In a stochastic model, the welfare loss equals

\[
L(x_t; \beta_S, \psi_S, p) = \frac{1}{1 - \beta_S (1 - p + p\psi_S^2)} \left[ (x_t - x^*)^2 + \frac{\beta_S}{1 - \beta_S} p (1 - \psi_S)^2 \sigma^2 \right]
\]

and in the deterministic version

\[
L(x_t; \beta_D, \psi_D) = \frac{1}{1 - \beta_D \psi_D^2} \left[ (x_t - x^*)^2 + \frac{\beta_D}{1 - \beta_D} (1 - \psi_D)^2 \sigma^2 \right]
\]
A more substantial extension would be the inclusion of a demand-side influence on the timing of precedent-setting judicial decisions. In particular, the law-and-economics literature has suggested that an crucial efficiency-promoting force in a common law system is the endogeneity of cases: when the law is more distant from the efficient rule, discontented citizens are more likely to challenge it in court, thereby giving judges more numerous and more frequent occasions to revise the rule. Priest (1977), Rubin (1977), and Landes and Posner (1979) develop alternative frameworks in which the evolutionary process of the law toward efficiency is driven by the fact that inefficient rules are more likely to be litigated by initiative of the parties whose interests are most affected and, as a consequence, more likely to be subsequently improved.

We can capture this hypothesis in very general terms by letting the probability that the court has a chance to refine the law at time $t$ be defined by a function $p(x_{t-1}) > 0$ that is symmetric around $x^*$ and monotone increasing away from it. For such a state-contingent process, the transition path is no longer explicitly computable, but convergence still occurs and we can compute exactly the asymptotic distribution when the distribution of judges’ preferences is normal. This yields:

**Proposition 3** Suppose that judges’ preferences are normally distributed $\hat{x} \sim N(x^*, \sigma^2)$ and that in every period the high court has probability $p(x_t) > 0$ of being able to refine the existing law $x_t$. Then, from any starting point case law converges to an asymptotic distribution with continuous density on $\mathbb{R}$:

$$
\mu(x) = \frac{1}{p(x)} \varphi(x) \int_{-\infty}^{\infty} \frac{1}{p(x')} \varphi(x')
$$

where $\varphi(x)$ denotes the density of $N(x^*, \frac{1-p}{1+p}\sigma^2)$.

Identity between the two is established by

$$
\begin{align*}
\beta_D \psi_D^2 &= \beta_S (1-p+p\psi_S^2) \\
\frac{\beta_B(1-\psi_B^2)}{(1-\beta_D)(1-\beta_D\psi_D^2)} &= \frac{\beta_S(1-\psi_S^2)^2}{(1-\beta_S)[1-\beta_S(1-p+p\psi_S^2)]}
\end{align*}
$$

namely by defining

$$
\begin{align*}
\psi_D^2 &= \frac{\beta_S(1-p+p\psi_S^2)p(1-\psi_B^2)+(1-\beta_B)(1-\beta_S)(1-p+p\psi_S^2)}{p(1-\psi_B^2)+(1-\beta_B)(1-p+p\psi_S^2)} \in (0, 1)
\end{align*}
$$

$$
\beta_D = \frac{\beta_S(1-\psi_S^2)+(1-\beta_S)\beta_B(1-p+p\psi_S^2)}{\beta_S(1-\psi_S^2)+(1-\beta_S)} \in (0, 1)
$$
If \( p(x - x^*) = p(x^* - x) \forall x \in \mathbb{R} \land p'(x) > 0 \forall x > x^* \) then the asymptotic common law has expectation

\[
E(x) = x^*
\]

and variance

\[
Var(x) = \frac{\int_{-\infty}^{x} \frac{(x-x^*)^2}{p(x)} \varphi(x) \, dx}{\frac{1}{\int_{-\infty}^{\infty} \frac{1}{p(x)} \varphi(x) \, dx}} < \frac{1 - \psi}{1 + \psi}^2
\]

This proposition provides rigorous support to the intuition that a positive responsiveness of caseload to the sub-optimality of the current rule will be a force toward pushing towards greater asymptotic efficiency of case law.

A potential trade-off derives again from the speed of convergence; however, unlike the case of \textit{stare decisis}, this framework allows for some slackness of the constraint. Indeed, since faster jurisprudential evolution is always desirable \textit{ex ante}, a higher average probability of legal innovation is certainly preferable. But for any given average probability smaller than one, the system is asymptotically more efficient if the probability is decreasing as the law approaches the optimum: and while this makes the speed of convergence non-constant, it need not have an expected cost \textit{ex ante}.

This mechanism, therefore, could be an important further element driving case law towards efficiency: it is not, however, a necessary element; and for the sake of analytical tractability we will continue our analysis with our baseline assumption of a constant maximum caseload.

3 Case law vs. statutes

Naturally, the alternative to judge-made law is legislation: rather than relying on the evolution of jurisprudence, society can entrust a legislator, or more realistically a legislature, with the authority to write its laws. In its purest form, as theorized for instance by an intellectual tradition that both inspired and followed the French Revolution, this system of civil law interprets sovereignty as the authority to enact by statute any legal rule deemed appropriate, without regard for previous legal conventions and indeed perhaps (as in the case of the Revolution itself) with the precise intent of breaking away from them. Rousseau
(1762), for instance, goes as far as to state that “there is in the State no fundamental law which cannot be revoked” by the assembled people manifesting the general will. This idea, which is not peculiar to French thought but rather belongs to the common heritage of the Enlightenment, can be found already in Hobbes’s (1681) dismissal of the picture of the English common law as embodying a supra-personal “reason” to which even the sovereign will should bow. For him such “reason” could be nothing but “the natural reason of this or that Judge,” the only legitimate law being therefore “the command of him or them that have the sovereign power.”

Probably the definitional feature of civil law systems is its reliance on written statutes: laws are created by some combination of the executive and the legislative power who set them out in writing. The primary source of law is the “enacted law,” including rules established by legislatures, executives and administrative agencies (Glendon et al. 1999). Even before the era of extensive codification in the Continent, starting in the 17th century, the civil-law tradition conceived of itself as a continuation of and a commentary on Justinian’s Codex Iuris Civilis. Hence, at least according to the traditional view, the judiciary is not by itself an independent source of law: its main task is the faithful application and implementation of the legislative text in particular cases. As the Justinian Code put it: Non exemplis, sed legibus, indicandum est. The emphasis on codification rather than judicial precedent results also in differences in the roles of judges and in the style of legal reasoning in the two main legal families (Apple and Deyling 1995). Civil law has a deductive nature: it goes from general rules to particular cases. Moreover, the preexistence of the rule is a condition for its applicability “after the fact,” so that statute-law is, by its very nature, anticipatory.

Legislation, therefore, simply sets at the beginning of each period

\[ x_t = \hat{x}_t \]

and therefore statute law has no intrinsic evolutionary property: its improvement over time

---

5Hayek (1973) traces the understanding of the legal process to the origins of the civil law system of continental Europe: “from the thirteenth century onwards, and mainly on the European continent, law-making slowly and gradually came to be regarded as an act of the deliberate and unfettered will of the ruler (...) From the detailed studies of this process, it appears to be closely connected with the rise of absolute monarchy when the conceptions which later governed the aspirations of democracy were formed.”
cannot be attributed to an endogenous mechanism, but to exogenous improvements in democratic institutions that make the legislature more likely to enact a statute closely approximating the optimum, i.e. formally that reduce $\hat{\sigma}^2$. The role of judges, in this context, is the mechanical application of the written statute, ensured by some appropriate incentive scheme.\textsuperscript{6}

### 3.1 The short-run vs. the long-run certainty of the law

Social stability and predictability are among the basic aims that a legal system is expected to promote, and one of the most salient consequences of the structural differences between the two main legal families is the extent to which each is able to achieve them (Glendon et al. 1999). Law contributes a great deal to the reduction of uncertainty about the consequences of future actions. Oliver W. Holmes went so far as to define law itself as a prediction: “a legal duty so called is nothing but a prediction that if a man does or omits certain things he will be made to suffer in this or that way by judgment of the court.” (Holmes 1897).

Formally, our model captures directly a desire for certainty, since the conditional variance of future legal rules is included in the social welfare function. Yet the very association of the certainty of the law with its variance might appear to run counter to the commonplace association of the term with precisely worded formulae in a written text: one could argue that certainty means that the law is deterministic, perfectly known beforehand.

From the latter point of view, statutes would seem to provide more certainty than case law: the statute that will be applied is already written \textit{ex ante} and an individual agent can perfectly foresee, if the text is exhaustive and clear enough, the legal consequences that his actions would entail. On the other hand, the common law could be more subject to the whims of potentially biased judges and, specially in a society with high heterogeneity of judges’ preferences, individuals might be \textit{ex ante} quite uncertain about the ruling that would be applied to their case in the eventuality of a conflict, and this imposes additional

\textsuperscript{6}This seems to conform to at least some traditional understandings of the role of the judiciary in civil law countries. In post-Revolutionary France “the judge was to be only \textit{la bouche} (the mouthpiece), not \textit{le cerveau} (the brains), of the law.” (Damaska 1986)
costs on their decision-making. In his incensed criticism of the common law, Bentham took up both sources of concerns stressing that

“uncertainty is the inherent disease of that wretched substitute to law, which is called *unwritten* law, and which, in plain truth, is not law at all,” since “from the *uncertainty* comes not only *insecurity* but *corruption: insecurity*, in the situation of the *non-lawyer–corruption* (...) in that of the judge.” (1998)

In a sense, given that its exact content is known precisely by the individuals only after judicial adjudication, case law has one of the defining features of *standards* as opposed to legal rules. Kaplow’s (1992, 1999) comparative analysis of the optimality of rules versus standards, which finds that rules, i.e. laws which purport to provide advance guidance through *ex-ante* determination of their content, are more likely to be preferred to standards, whose precise content is given *ex post* (after individuals act), in contexts where there is a great expected frequency of application of the law, and there is a high value associated with individuals becoming better informed. Although rules can be more costly to promulgate, standards are more costly to predict for the agents and their legal advisors and also in terms of the effort required for the exact determination of their content.

All these considerations pushing in favour of statutes lose sight of the evolutionary aspect of the legal process: for a broad range of decisions what is essential is not the predictability of the immediate legal effects of one’s actions in the light of a given law, but rather the persistence of the current legal framework over time. Fundamental decisions for a country’s economic performance, such as investment in physical and human capital, turn crucially on the stability of the law itself in the medium and long run. It is, therefore, important to keep in mind the distinction made by Leoni (1961, chap. IV), who labelled the predictability of the law given by written statutes as its *short-run* (or Greek) certainty, and remarked that:

“the idea of the certainty of the law has not only the above-mentioned sense in the history of the political and legal systems of the West. It has also been understood in a completely different sense. [...]

The Romans accepted and applied a concept of the certainty of the law, that could be described as meaning that the law was never to be subjected to sudden
and unpredictable changes. Moreover, the law was never to be submitted, as a rule, to the arbitrary will or to the arbitrary power of any legislative assembly or of any one person, including senators or other prominent magistrates of the state. This is the long-run concept, or, if you prefer, the Roman concept, of the certainty of the law."

It is precisely this distinction that is captured mathematically by our observations about variance: indeed, a stark visualization is possible. At any moment $t$, statute law gives short-run certainty because the rule $x_t$ is perfectly known, having been committed in advance to the statute books; what is more, we are willing to concede (despite Leoni’s admittedly reasonable reservations) that citizens can safely assume that the law will not be changed until a new election has changed the composition of the legislature itself: e.g. most clearly every four years in a parliamentary system such as Britain’s. However, after $N$ periods of certainty the law becomes indeed open to “sudden and unpredictable changes”, namely its variance jumps discontinuously to $\hat{\sigma}_L^2$.

On the contrary, case law can be said never to provide complete short-run certainty: at time $t$ citizens will know the settled precedent $x_{t-1}$, but only ex post will the courts rule exactly on $x_t$: thus the variance of current law will be

$$Var(x_t|x_{t-1}) = (1 - \psi)^2 \hat{\sigma}_J^2 > 0$$

Yet the conditional variance of future law increases continuously but ever more slowly as the horizon lengthens, and asymptotes to

$$\frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \ll \hat{\sigma}_J^2$$

which confirms Leoni’s point that the long-run certainty of judge-made law is much higher than that of statute law, even if $\hat{\sigma}_J^2$ were several times larger than $\hat{\sigma}_L^2$. For instance, the graph below is drawn for $\hat{\sigma}_J^2 = 5\hat{\sigma}_L^2 \land \psi = 0.8$. 

19
It is visually immediate that the preference for either legal system will be influenced by the patience or foresight embedded in the evaluation: an emphasis on the long run tends to favour common law, while short run concerns press for statutes. However, certainty is only half of the picture: welfare depends no less importantly from the average deviation of the law from its optimum. This, in turn, is determined both by the evolutionary properties of case law, and by the preferences of the decision-makers entrusted with creating legal rules.

3.2 Legal origins

Legislation as a prerogative of sovereignty embodies the rule that the sovereign deems most appropriate and desirable, but even in the absence of democratic institutions it would be incorrect to regard statutes as the expression of a dictator’s arbitrary whims. Even a lunatic tyrant would find it easier to satisfy his extemporaneous desires by placing himself above all laws, rather than by minutiously rewriting the statute books. Furthermore, any ruler, no matter how formally absolute and untrammelled his authority, will still have some regard for social welfare: first, he will need the support of at least some social groups to remain in power, and his ability to ignore public opinion is likely to be limited; second, while this force may diminish the more secure a ruler’s grip on the reins of power, stability will lead to a concern for efficiency by another channel, namely the ability to extract rents from the
subjects in the long-term (cf. the literature on “stationary bandits”, pioneered by Olson 2000).

Indeed, it is far from unthinkable that the preferences of an unelected legislator could be more closely aligned with social welfare than those of a judge whose concern for efficiency is also far from perfectly assured. A closely related point has been raised precisely concerning comparative legal institutions by Glaeser and Shleifer (2002): in their analysis, the driving factor leading England to a common-law system and France to a civil-law system is the different degree to which local magnates could sway court decisions in either country during the Middle Ages. In England juries were relatively unbefehlten to feudal lords, and therefore more efficient than a magistrate controlled by the sovereign; on the other hand in France powerful noblemen would have been able to manipulate jury verdicts to such an extent that it was preferable to entrust the administration of justice to the king’s emissaries, even if the monarch’s own interests would occasionally colour their decisions.

It should be noted that Glaeser and Shleifer focus on a difference between common law and civil law that is not apparently related to the one we have been considering. It is easily possible to conceive of a system of trial by jury where legal norms are determined by statutes rather than precedents, and conversely of a purely jurisprudential system where judges both decide the case and develop the law based on previous decisions by their peers: indeed a persuasive case could be made that the former is an accurate description of contemporary British and American criminal justice, and the latter of ancient Roman courts.

Nevertheless, our model helps explain why the two European legal traditions developed simultaneously the jury trial and stare decisis on the one hand, and the inquisitive trial and written statutes on the other: the same force, namely the power of local magnates to sway judicial decisions, can be seen to affect both choices. Within our framework, both the king when issuing his edicts and the judges when ruling from the bench can be induced to favour certain powerful special interests: therefore, the problem can essentially be framed in terms of comparative opportunities for buying influence.

For simplicity, we will abstract from the fact that judges almost surely differ in their preferences, and assume, loosely following Posner, that all share a fundamental concern for efficiency. However, every judge is also subject to pressure from a local magnate whose
personal interest is at stake in the case before the judge: the extent of this pressure is expressed by \( b \in (0, 1) \), which measures to what extent pleasing the magnate replaces achieving efficiency in the judge’s objective function. Thus the judge’s preference is given by

\[
\hat{x}_J = (1 - b) x^* + b \hat{x}_m
\]

where \( \hat{x}_m \) is a random variable denoting the idiosyncratic preference of each magnate.

The king has some private interest, which may or may not include an explicit concern for the welfare of his subjects; furthermore \( M \) magnates compete for the king’s ear, trying to extract from the sovereign patronage and favours, including special provisions in royal edicts. Assuming for simplicity that all magnates have identical resources, we can model the king’s decision-making process as a Nash bargaining game, and let his preference be

\[
\hat{x}_K = k \hat{x}_0 + (1 - k) \sum_{m=1}^{M} \frac{\hat{x}_m}{M}
\]

where \( k \in \left[ \frac{1}{M+1}, 1 \right] \) measures the degree of absolutism of the monarchy. As Goldberg and Maggi (1999) point out, this bargaining game yields the same conclusion as the more complex model of influence for sale presented by Grossman and Helpman (1995): we can therefore identically think of magnates competing in a menu auction offering tributes and services (e.g. military support) to the sovereign in exchange for favourable legislation.

The king can be seen as the most powerful of magnates: literally a *primus inter pares* if \( k = \frac{1}{M+1} \); in general, his preferences are drawn from the same distribution as those of magnates, and this distribution has mean \( x^* \) and variance \( \sigma_m^2 \): this could reflect the concern that not only the king but every feudal lord has for efficiency within his fiefdom.

This entails a result closely mirroring Glaeser and Shleifer (2002):

**Lemma 1** Conferring legislative authority to the monarch is unambiguously more attractive when the magnates are more powerful, because this simultaneously

1. increases the distortion they impose on the courts by exerting undue influence on judges (\( \partial \hat{\sigma}_J^2 / \partial b > 0 \)); and

2. reduces the degree of absolutist discretion embedded in the king’s edicts (\( \partial \hat{\sigma}_K^2 / \partial k > 0 \)).
A clear economic intuition supports this result. All oligarchs, including the king himself in his private capacity, compete for legislative influence; competition among their conflicting interests is a force driving towards social efficiency, and the more level the playing field the more efficient the outcome will be: while we have only considered how the king could be more powerful than every other lord, the same argument could easily extend to heterogeneity across barons. On the contrary, every case involves only one magnate (or at most two, which would not qualitatively change the result), and therefore the variance of their influence on judges is greater.

Needless to say, however, the convergence mechanism of case law alters the picture: in a sense, we could say that, through the role of precedent, common law forces magnates to compete against each other over time, although each is only trying to influence the judge deciding his own case. So it may well be optimal to adopt a case law system even when \( \hat{\sigma}_J^2 > \hat{\sigma}_K^2 \).

As we already suggested, \textit{ex ante} efficiency can be precisely evaluated from the point of view of a Rawlsian benevolent planner judging legal systems at a hypothetical time 0 (in the “original position,” cf. Rawls 1999). It is straightforward to compute that in a statute-law system the expected loss is constant

\[
E[l_K(x_t)] = \hat{\sigma}_K^2 \forall t \geq 1
\]

while we have already shown that in a case-law system it is a decreasing function of time

\[
E[l_J(x_t)] = \frac{1 - \psi + 2\psi^{2t-1}}{1 + \psi} \hat{\sigma}_J^2 \forall t \geq 1
\]

As a consequence, we can prove the following

**Proposition 4** Consider a Rawlsian social planner designing \textit{ex ante} the optimal institutional system: he will choose case law whenever

\[
\frac{\hat{\sigma}_K^2}{\hat{\sigma}_J^2} \geq \sqrt{1 - \beta}
\]

setting optimally the institutional importance of precedent to

\[
\psi^* = \frac{1 - \sqrt{1 - \beta}}{\beta}
\]
The intuition behind these results is straightforward. The optimal initial choice is dictated by two considerations: the relative quality of the two decision-making bodies, and the patience of the social planner. The more patient society is (the higher $\beta$), the more will it accept a slow convergence towards a very efficient rule (the higher the optimal $\psi$): since this is a property of case law that statute law does not replicate, a more forward-looking institutional designer will also be more likely to choose a common-law system.

If judges and lawmakers were equally concerned about social welfare ($\hat{\sigma}_J^2 = \hat{\sigma}_K^2$) it would be unambiguously preferable to set up a common-law system, regardless of the time preferences of a society. Yet, as our previous discussion has highlighted, legislators can be more attuned than judges to the popular will: this does not require them to be Platonic philosopher kings, but merely to be convincing petitioned by a broader range of actors than would influence a magistrate. Under such conditions, a social planner who only cares about the remote asymptotic conditions of the system ($\beta \to 1$) will still choose case law for its convergence properties; but a reasonable concern for the short run may lead instead to statutes being considered more efficient.

A simple calibration conveys plausible results for the historical origins of European legal families. A credible definition of a period, i.e. the interval between significant cases decided by the high court, would be no longer than one year: then the discount factor would be of the order of magnitude of $\beta \approx 0.9$. This implies a threshold of $\hat{\sigma}_J^2 \approx 3.2\hat{\sigma}_K^2$, which entails the optimality of statute law in a kingdom with a dozen of equally powerful feudal lords whose influence on judges accounts for about 51% of their objective function.

Needless to say, no legal system was literally designed by a Rawlsian social planner: yet, like Glaeser and Shleifer (2002), we believe that considerations of social efficiency must have helped to tilt the balance in favour of the emergence of common law in medieval England, where the the power of feudal lords was relatively modest, and civil law in medieval France where the king was at best the first among equals.

### 3.3 Democracy and the law

The quality of statutes is entirely determined by the importance of social welfare in the legislature’s preferences: indeed, throughout history supporters of *ius scriptum* have been
wont to argue that the legislator would be wiser and more impartial than the courts: a point notably made by Hobbes and the French revolutionaries, but also much earlier by Aristotle, for whom as few decisions as possible should be left to the discretion of the judges, since:

“they will often have allowed themselves to be so much influenced by feelings of friendship or hatred or self-interest that they lose any clear vision of the truth and have their judgement obscured by considerations of personal pleasure or pain.”

Such a line of reasoning becomes much more forceful in a modern, democratic polity: if both the legislature and the judiciary are democratically selected, then the presumption must be that judges are no less biased than legislators—a view that seems borne out by recent political commentary.

The argument can be backed by a stylized formal model that is in fact remarkably similar to the one presented above; after all special-interest groups continue playing an important role in democratic politics: cf. the fundamental contribution by Grossman and Helpman (2001). $S > 1$ special-interest groups are exogenously formed from the general population, and each group’s preferences are a random variable $x_s$ with mean $x^*$ and variance $\sigma^2$. These groups could simply bargain with an elected government just like barons with their king, but we can also offer an alternative interpretation that focuses purely on electoral competition: if members of these groups have a keener interest in politics than the whole population, and specifically they watch more closely the platform choices of political parties, then the median-voter theorem fails even in a single-issue majoritarian election, because politicians will cater to the preferences of the groups supporting them in order to ensure turnout at the polls, as shown by Glaeser, Ponzetto and Shapiro (2005).

As a result, democratic politics have a degree of imperfection $\iota \in (0,1)$ such that the social optimum only carries a weight $(1 - \iota)$ in legislative preferences, which can be expressed as

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

---

Judges share a priori the preferences of the legislature, being either directly elected by the same voters, or appointed by the representatives themselves. However, they may also be influenced by a special-interest group that is directly involved in the case being discussed. Hence their preferences are

\[ \hat{x}_J = (1 - b) \hat{x}_L + b \hat{x}_s \]

where \( b \in (0, 1) \) again measures outside influence on judicial decisions: this can result simply from the greater resources available to more powerful litigants, from better lawyers to amicus curiae briefs, or also from outright intimidation and bribing, which unfortunately have not completely disappeared since the Middle Ages.

This leads to the following:

**Lemma 2** If both the legislature and the judiciary are democratically elected by a process with a degree \( \iota \in (0, 1) \) of imperfection, and judges are subject to a degree \( b \in (0, 1) \) of partiality to special-interest groups, then

1. legislators’ preferences \( \hat{x}_L \) have mean \( x^* \) and variance

\[ \hat{\sigma}^2_L (\iota) \in (0, \sigma^2) : \frac{\partial \hat{\sigma}^2_L}{\partial \iota} > 0 \]

2. judges’ preferences \( \hat{x}_J \) have mean \( x^* \) and variance

\[ \hat{\sigma}^2_J (\iota, b) \in (\hat{\sigma}^2_L, \sigma^2) : \frac{\partial \hat{\sigma}^2_J}{\partial \iota} > 0 \land \frac{\partial \hat{\sigma}^2_J}{\partial b} > 0 \]

3. the relative attractiveness of case law can be measured by

\[ \frac{\partial^2 \hat{\sigma}^2_L}{\partial \hat{\sigma}^2_J} (\iota, b) \in (0, \iota) : \frac{\partial \left( \frac{\hat{\sigma}^2_L}{\hat{\sigma}^2_J} \right)}{\partial \iota} > 0 \land \frac{\partial \left( \frac{\hat{\sigma}^2_L}{\hat{\sigma}^2_J} \right)}{\partial b} < 0 \]

The crucial result is the third, which captures a key disadvantage of creating the law from decisions about particular cases rather than from general, abstract rules. When all branches of government are democratically elected, all equally suffer from the inevitable imperfections of democratic processes; but the judges are subject to additional pressure from the groups that are affected directly from their ruling, rather than merely by the general principle the ruling will help develop. As a consequence, unless judges are perfectly impartial, their
opinion will be more variable than the legislature, because it will depend not only on the society-wide balance of general and special interests, but also on the varying power of the parties appearing before them in court.

A better court system (lower $b$) will of course improve judicial decision-making but will not affect legislation; while greater democratic representativeness (lower $\iota$) will improve the operation of both branches: however, it will make any remaining judicial bias $b$ relatively starker, thereby enhancing the profile of the legislature.

Thus a limit case of the model perfectly captures an argument famously proposed by the theorists of the French Revolution, albeit surprisingly naïve and contradicted by empirical evidence (not to mention common sense): if one truly believes that the people “could not be divided in its will” (Palmer 1941), then the legislature as a perfect ($\iota = 0$) manifestation of the people’s will must necessarily have preferences $\hat{x}_L = x^*$, and assigning any role to judges can have no benefit, while it has at least some potential to dilute the perfect optimality of statutes.

Short of this dangerous utopia, however, we must keep in mind the counter-argument always made by proponents of judge-made law: tradition is the source of the strength of jurisprudence, despite the individual shortcomings of each judge. In other words, o long as $\iota, \psi > 0$, the convergence properties of case law may well make it preferable to statutes even if $\hat{\sigma}_L^2 < \hat{\sigma}_J^2$.

### 3.4 The superiority of case law in a stationary environment

To some extent, the Rawlsian analysis of Proposition 4 already captured the appeal of the convergence property; however, it would be fundamentally inappropriate to suggest welfare considerations for contemporary societies based on the perspective we have adopted to analyze the medieval origins of European legal families. In the modern era, legal rules and institutions are not created from scratch, and in particular all case-law systems are rooted in, and indeed typically transplants of, the English common law.

Furthermore, an emphasis on ex-ante welfare judgements would betray the very arguments made by the philosophical supporters of case law: we could hardly think, for instance, of a conception further removed from Hayek’s thought than a benevolent institutional designer
rationally planning the creation of common law. The earliest supporters of judge-made law invariably stressed the importance of the evolution that had already occurred over centuries, not the one that could yet take place in the future. For instance, according to Cicero, Cato the Censor proudly proclaimed that:

“the reason why our political system was superior to those of all other countries was this: the political systems of other countries had been created by introducing laws and institutions according to the personal advice of particular individuals [...] Our state on the contrary is not due to the personal creation of one man, but of very many; it has not been founded during the lifetime of any particular individual, but through a series of centuries and generations.”

Similarly, when in seventeenth-century England Hobbes attacked the common law as irrational, the Lord Chief Justice, Sir Matthew Hale, replied that:

“It is a reason for me to prefer a law by which a kingdom hath been happily governed four or five hundred years than to adventure the happiness and peace of a kingdom upon some new theory of my own.”

Formally, the importance of an uninterrupted tradition can be grasped by considering that the same expression that was given for the \( \text{ex-ante} \) expected loss at time \( t \) is also the unconditional expectation of the loss suffered by a society that has been governed by a given legal system for \( t \) previous periods. Equivalentely we can state the following:

**Condition 1** After \( t \) periods of evolution case law is expected to achieve a more efficient rule than statute law if and only if

\[
\frac{\sigma_L^2}{\sigma_J^2} > \frac{1 - \psi}{1 + \psi} + \frac{2\psi^{2t-1}}{1 + \psi}
\]

Asymptotically a case-law system is expected to be more efficient when \( \psi \to 1 \) for any value of \( \sigma_L^2/\sigma_J^2 \); the striking result is the speed of convergence: in the graph below with \( \psi \) on the horizontal and \( \sigma_L^2/\sigma_J^2 \) on the vertical axis, the curves display the lower boundary of the

---

8 *De Re Publica* II.1.2 quoted in Leoni (1961, p. 88)
9 In Holdsworth (1924, p. 504).
region of expected superiority of common law for horizons of a mere 10 and 50 periods, with
the asymptotic boundary $\frac{1 - \psi}{1 + \psi}$ depicted by the dashed curve. After centuries of evolution,
as claimed by Cato and Hale, judge-made law with binding precedent (i.e very high. $\psi$) must be expected to be more efficient than statute law almost regardless of the individual
deficiencies of the magistrates.

A patient society will not be concerned exclusively about current efficiency, but also
about the future: this will make common law even more attractive, because the benefits of
its future evolution will be added to the benefits of its past evolution.

**Lemma 3** After $t$ periods of evolution case law is expected to achieve greater social welfare
than statute law if and only if

$$\frac{\hat{\sigma}_L^2}{\hat{\sigma}_J^2} > \frac{1 - \psi}{1 + \psi} \frac{(1 - \beta) 2^t - 1}{(1 + \psi)(1 - \beta\psi^2)}$$

Hence case law is expected to achieve greater social welfare than statute law if, but not only
if, it is expected to be currently more efficient.

While the conditions on the expected outcomes of different legal systems are already
strongly suggestive of the optimality of case law, a more complete analysis only requires
specifying the distribution of decision-makers’ preferences. In particular in the normal case
we can prove the following:
Proposition 5 If judges’ preferences are normally distributed $\hat{x}_J \sim N(x^*, \sigma_J^2)$ and legislators’ preferences are normally distributed $\hat{x}_L \sim N(x^*, \sigma_L^2)$ then the ratio of the efficiency losses imposed respectively by case law evolved over $t$ periods and by statute law has an $F_{1,1}$ distribution

$$\frac{1 + \psi \sigma_L^2 (\hat{x}_J - x^*)^2}{1 - \psi + 2\psi^{\frac{1}{2}} \sigma_J^2 (\hat{x}_L - x^*)^2} \sim F_{1,1}$$

Though somewhat abstract in itself, this result allows us to compute exactly the probability that case law attains a more efficient rule than statute law. This also provides another way of visualizing the convergence of case law, and its speed.

The graph below shows, as a function of $\sigma_L^2 / \sigma_J^2$, the probability that a case-law system with $\psi = 0.95$ has achieved a more efficient rule than a statute-law system after respectively 10, 20, 40 and 80 periods; the dashed line at 0.5 provides a reference point indicating when the two systems are statistically indistinguishable.

![Graph showing the probability of case law achieving a more efficient rule than statute law](image)

After 80 periods, the system has virtually reached its asymptotic condition, namely increases in the horizon no longer determine any noticeable difference in the curve; after Hale’s four or five centuries, even as persistent a system as $\psi = 0.99$ would have essentially converged.

In general, it is thus worth focusing on the asymptotic case: in the graph below, the curves are drawn or values of respectively $\psi = 0.99, 0.95, 0.9$ and 0.8 from top to bottom. The probabilities thus computed for asymptotic case law overestimate the actual probability.
by less than one percentage point for histories longer than, respectively, $t = 401, 64, 28$ and merely 12 periods.\footnote{In other words, the functions are considered to have converged when their distance from the asymptote is less than 0.01 according to the sup norm.}

Within the limits of a stylized model, this analysis is certainly consistent with the empirical findings of La Porta et al. (1999) who find an overwhelming statistical significance of common-law systems in predicting a variety of economic and political variables that we construe as reflecting our overarching concept of efficiency.

But once again, current realized efficiency is not society’s only concern, as the welfare function itself emphasizes. It remains possible (see the Appendix for the derivation) to compute exactly the probability that after $t$ periods of evolution common law yields higher social welfare than civil law; but this probability does not coincide with the one we have just shown.

We previously highlighted Leoni’s distinction between the short-run and long-run certainty of the law: this is precisely the dichotomy that emerges at this stage of the analysis. Case law may deliver \textit{ex post} efficiency but not \textit{ex ante} welfare due to the cost of uncertainty about the judge’s ruling. For a society that is exclusively concerned with the present ($\beta = 0$), this represents a pure disadvantage of case law: the probability that it provides greater \textit{ex post} efficiency but lower \textit{ex ante} welfare is presented in the graph below for values of $\psi = 0.8$, \ldots
0.9, 0.95, and 0.99 respectively from top to bottom, with the ratio $\sigma_L^2/\sigma_J^2$ on the horizontal axis\textsuperscript{11}.

![Graph showing the probability of common law providing greater welfare even if yielding lower current efficiency when $\psi = 0.9$ and for values of $\beta = 0.95, 0.8, 0.6, 0.4, 0.2$ and 0.05 respectively from the top.]

Unless $\sigma_L^2/\sigma_J^2$ is very small, implying that most of the uncertainty in the legal system is coming from the strong pressure of special-interests over judicial decisions, this cost is rather low: the case of $\sigma_L^2/\sigma_J^2 = 1$ can be seen as representing the intrinsic cost of judge-made law in a society with efficient courts.

Furthermore, focusing on $\beta = 0$ fails to take into account the offsetting role of long-run certainty: according to Leoni’s intuition, this is in fact the dominant factor in welfare comparisons. The graph below shows the probability that common law provides greater welfare even if yielding lower current efficiency when $\psi = 0.9$ and for values of $\beta = 0.95, 0.8, 0.6, 0.4, 0.2$ and 0.05 respectively from the top.

\textsuperscript{11}The asymptotic functions are drawn: convergence in the case of welfare occurs in respectively $t = 12, 28, 64,$ and 399 periods.
As a consequence, the probability that common law delivers greater welfare to a patient society is extremely high, as shown by the graph below for $\psi = 0.9$ and for values of $\beta = 0.99$, $0.75$, $0.5$, $0.25$, and $0$ respectively from the top. As $\beta \to 1$, society is entirely focused on asymptotic efficiency, and therefore social preferences tend to a step function at

$$\frac{\sigma_j^2}{\sigma_f^2} = \frac{1-\psi}{1+\psi} \approx 0.05.$$

Overall, we cannot escape the conclusion that, in a stationary environment (i.e. absent any changes in the exogenous parameters) case law is preferable to statute law: the strength of this conclusion for plausible parameter values is starkly captured by a final, single graph: the likelihood that a society with $\beta = \psi = 0.9$ prefers common law (evolved through $28$ or more rulings) to statutes is described as a function of $\frac{\sigma_j^2}{\sigma_f^2}$ by
and it equals 0.5, which implies no statistical difference between the two systems, if $\sigma_J^2 \approx 22.1\sigma_L^2$.

Such a theoretical result fits well with recent empirical evidence on the benefits associated with common law and with a long-standing philosophical tradition; given the remarkable robustness of the prediction, however, it is worth considering why civil law prevails in so many countries.

One possibility is of course that case law is in fact suboptimal according to our preceding analysis because the judiciary is extremely inefficient and corrupt compared to the legislature; this explanation strikes us as implausible, given that corruption displays a tendency to permeate analogously the three branches of government. Although ultimately the issue can only be settled by detailed empirical analysis, skepticism is warranted concerning the hypothesis that competition between special-interest groups in lobbying law-makers but not judges is sufficient to determine preferences whose variances differ by more than an order of magnitude.

A more convincing, if no less bleak, explanation is that there remains a discrepancy between normative and positive analysis. Countries could simply have sub-optimal institutions because their designers are not primarily motivated by the pursuit of efficiency or social welfare; for instance, it is wholly unsurprising that no communist country has ever had a system of judge-made law: revolutionary, authoritarian one-party rule is incompatible with a slowly evolved legal system even if—or perhaps precisely because—the latter is likely to be socially
preferable to socialist institutions.

Yet historical experience also shows that the English common law has invariably been retained by the British colonies and dominions where it had been fully introduced: therefore, its absence need not denote the intervention of a malevolent designer, but simply the result of institutional inertia. It must be stressed that this need not be inefficient: while this is not explicitly presented in our model, converting a legal system to a different mode of operation is certainly socially disruptive in the short run, and entails an undoubted cost in the medium run as well, as the human capital of lawyers and magistrates needs to adapt. The per-period efficiency gain, on the other hand, are of the order of magnitude of $\sigma^2_L$, and precisely:

$$(1 - \beta) E [L_L (t) - L_J (t)] < \sigma^2_L - \frac{1 - \psi}{1 + \psi} \delta^2_J$$

If the cost of change are sufficiently high, and unless society is not extremely focused on the very long run (if $\beta \to 1$ any short-run cost is worth bearing to obtain any permanent benefit) it is therefore inefficient to move from statute law to case law when a country’s democratic institutions are sufficiently good. Suggestive evidence supporting this consideration comes from the observation that reforms explicitly tending to bring common-law mechanism into civil-law systems have recently been proposed not in the advanced democracies of Western Europe, but in the more fragile institutional setting of Latin America.

Finally, a crucial consideration is that the analysis has so far been conducted, as we have explicitly acknowledged, in a stationary environment. A usual argument in favor of legislation, as Bentham, Hobbes, Austin, and even Hayek himself have noticed, is the faster adaptation to social changes that it makes possible: an efficiency-maximizing legislator could, by the enactment of a new statute, adapt the law perfectly to new circumstances, that is, to a situation in which the optimal ruling for society has shifted.

4 Social change

We assume a simple process for the change of preferences over time: when change occurs the whole distribution shifts uniformly (or identically, from a mathematical point of view, it is the horizontal axis that is translated), but the relative preferences of individuals are
unchanged. Hence a summary statistic for the evolution of preferences is the evolving social optimum $x_t^*$.

The fundamental determinant of any comparative optimality judgement following a change in the underlying social conditions is the in-built inertia of the evolutionary process of case law: sudden, radical changes are only possible through statutes. While this is, of course, the source of the unwelcome long-run uncertainty of statute law, it is also a rationale for preferring a legislative system whenever the social optimum has moved too far from the rule inherited from the past. Formally, it is straightforward to show the following:

**Lemma 4** If the social optimum exogenously changes from $x_0^*$ to $x^*$, statute law is expected to yield higher social welfare than case law if

$$\frac{\psi^2}{1 - \beta \psi^2} (x_0^* - x^*)^2 \geq \frac{\hat{\sigma}_L^2}{1 - \beta} - \frac{1 - \psi}{1 + \psi} \frac{\hat{\sigma}_J^2}{1 - \beta}$$

and this condition is necessary as well as sufficient when case law had time to converge to its asymptotic distribution before the social change.

This is the more likely the greater the social change ($|x_0^* - x^*|$), the less patient society (the lower $\beta$), the better the legislature (the lower $\hat{\sigma}_L^2$) and the worse the judges (the greater $\hat{\sigma}_J^2$).

The right-hand side measures the steady-state cost of civil law, namely the difference in the asymptotic expected loss from the two systems. The left-hand side is the cost of the sluggish adjustment of precedent to changed social conditions. Given the dynamic nature of the trade off, a more patient society cares more about long-run convergence than short-run reform, and is therefore more likely to prefer case law to statutes despite the initial inefficiency. Nonetheless, unless society is infinitely patient ($\beta = 1$), there can always be social changes sufficiently disruptive to make statutes preferable.

However, it is arguably incorrect to consider the dichotomy so starkly. The common-law tradition has always admitted the occasional introduction of statutes whereby a legislator dealt with novel circumstances that pre-existing precedents were not suited to accommodate. Although he was a passionate defender of the common law, Hayek (1973) acknowledges that the gradual evolutionary process of case-law can become “too slow to bring about the
desirable rapid adaptation of the law to wholly new circumstances” and that legislation
would be the natural response to correct for this flaw; and, being a great admirer of Leoni’s
emphasis on the long-run properties of common law, he explicitly points out a disagreement
with him on that we cannot entirely dispense with legislation.

As early a phenomenon as the emergence of the Court of Equity in the 14th century has
been seen as reaction to the unresponsiveness of the common law to the circumstances of
society (Scheb and Scheb 2002). Later, by the end of the 18th century, Jeremy Bentham
attacked Sir William Blackstone’s defense of tradition in law and complained about the
inordinate slowness of common law in responding to social needs (cf. Glendon et al. 1999).
In a way, the abiding question of the common law is, in Dawson’s words, “does law derive from
authority or reason, and if it derives from both, how do we resolve their contradictions?” The
principle of stare decisis prevents the common law from quickly responding to wide-ranging
or sudden social change, and slows down the internalization of the new circumstances into the
long-lived stock of traditional legal rules. We argue that this technological feature of case law
helps explain the ample spreading of written legislation in recent history: from a short-run
perspective, the long-term benefits of common law in terms of certainty can be more than
offset by the need for immediate adaptation to a changing world. This technological story
complements the explanation Glaeser and Shleifer (2003) offer for the “Rise of the Regulatory
State” that followed industrialization in 20th century America in terms of an increase in the
incentives to subvert the legal system: the driving force in their model is the dramatic
increase in economic activity and in firms’ size, which pushes away from adjudication by
common law judges.

We capture this structure of limited legislative intervention by admitting that, after
a change in the social optimum, the legislature is allowed to write a statute providing a
foundation for judicial decisions; but afterwards, no further legislative intervention can occur,
and the process of gradual evolution of case law proceeds again to refine the rule. This system
provides the same short-run adaptive capability of pure statute law, but it also benefits from
the convergence property of case law; we can thus prove the following:

**Proposition 6** If the social optimum exogenously changes from $x_0^*$ to $x^*$, then a mixed
system is preferable to statute law if and only if

\[ \sigma_L^2 \geq \frac{1 - \psi}{1 + \psi} \sigma_J^2 \]

The mixed system is expected to be preferable to case law if

\[(x_0^* - x^*)^2 + \frac{1 - \psi}{1 + \psi} \sigma_J^2 \geq \sigma_L^2 \]

and this condition is necessary as well as sufficient when case law had time to converge to its asymptotic distribution before the social change.

Remarkably, the whole condition is governed exclusively by the steady-state parameters of the system: in particular, it is independent of the social discount factor \(\beta\).

For the first part of the proposition, the underlying intuition is that the mixed system begins as statute law and then gradually becomes case law as the original legislation is “digested” and thereby slowly revised by jurisprudence. The rate at which this happens (governed by \(\psi\)) and the time preference of society (governed by \(\beta\)) only affect quantitatively how beneficial or detrimental the evolution is. Qualitatively, the optimality of either system does not depend on the speed of change, but merely on the direction of change: the evolution resulting from a mixed system is beneficial if the asymptotic loss is lower under pure case law, and detrimental if it is lower under pure statute law.

For the second part of the proposition, the lack of a role for \(\beta\) is intuitive because the only difference concerns the new starting point. The right-hand side is the time-invariant expected loss from statutes: they are always, on average, appropriate to the current optimum; but their imperfection is captured by their variance. The left-hand side is the initial loss expected from case law: this is impacted in full by the change in social conditions \((x_0^* - x^*)^2\), which precedent has not yet reacted to; it also incorporates the expected imperfection of precedents compared to the old optimum, namely \(\frac{1 - \psi}{1 + \psi} \sigma_J^2\) after a sufficiently long evolution.

### 4.1 Ongoing change

It is of course more realistic to assume that social change is an ongoing process, and that the new social optimum is not expected to be forever unchanged. In such a stochastically
evolving world, statute law becomes even more appealing: in fact, if progress is sufficiently fast and innovation sufficiently great, case law can be suboptimal regardless of society’s patience.

Formally suppose that social change occurs over time according to a Poisson process with arrival rate \( \lambda \) and i.i.d. increments \( \zeta \) with mean zero and variance \( \sigma_\zeta^2 \). We can then prove the following

**Lemma 5** *After a sufficiently long history of evolution, case law is expected to yield lower social welfare than statute law if and only if*

\[
\Lambda(\beta, \psi) \lambda \sigma_\zeta^2 \geq \hat{\sigma}_L^2 - \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2
\]

*where*

\[
\Lambda(\beta, \psi) \geq 0
\]

*and*

\[
\frac{\partial \Lambda}{\partial \psi} > 0 \land \Lambda(\beta, 0) = 0 \land \lim_{\psi \to 1} \Lambda(\beta, \psi) = \infty \forall \beta \in (0, 1)
\]

The lemma shows that when the process of social change is sufficiently intense, statute law will become preferable: this is because its ability to adapt faster to changing circumstances becomes more important, while the convergence of case law loses its appeal, since the social optimum changes faster than jurisprudence can approximate it. The role of \( \psi \) in determining the threshold is therefore clear: the more precedent-bound judicial decisions, the lower their ability to respond to exogenous change, and therefore the more likely that case law is outdated.

Once again, however, a mixed system is likely to be preferable. In fact, we can prove that under some conditions it is unambiguously optimal:

**Proposition 7** *A mixed system provides higher social welfare than pure statute law if and only if*

\[
\hat{\sigma}_L^2 \geq \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2
\]

*For every \( \lambda, \beta, \psi \in (0, 1) \) and \( \hat{\sigma}_J^2, \hat{\sigma}_L^2 > 0 \) there exists a finite threshold \( \Theta(\lambda, \beta, \psi, \hat{\sigma}_J^2, \hat{\sigma}_L^2) \) such that for*

\[
\sigma_\zeta^2 \geq \Theta(\lambda, \beta, \psi, \hat{\sigma}_J^2, \hat{\sigma}_L^2)
\]
the mixed system provides higher social welfare than pure case law with probability one.

The first part of the proposition mirrors our result for a once and for all change in the social optimum: the mixed system provides all the advantages of legislation in adapting to changes, and furthermore it adds the convergence properties of case law, assuming these exist at all (namely that the asymptotic case-law loss is lower than under statute law).

The second part shows captures the long-run importance of recurring shocks: even if case law could achieve great efficiency to a static optimum, and legislation could be rather unpredictable, letting the legislator intervene is always optimal if shocks are frequent and large enough.

Quantitatively, we would in fact expect the mixed system to be optimal: for instance, if $\beta = \psi = 0.9$, even for $\hat{\sigma}_L^2 = \hat{\sigma}_J^2$ its optimality is assured by a probability of social change $\lambda \geq 3\%$ and variance of shocks $\sigma_\xi^2 \geq 2\hat{\sigma}_L^2$.

Furthermore, there is substantial support, in the literature as well as in the data, for the hypothesis that both major legal families are in fact converging towards a mixed system that contemplates both legislative intervention and the gradual accumulation of precedent.

### 4.2 Convergence

A considerable legal literature documents the convergence of common-law and civil-law systems in terms of substantive outcomes (Coffee 2000) and the evolution of a hybrid of the two law-making technologies, as presented above (Zweigert and Kötz 1998, Glendon et al. 1999 references).

To be sure, there has never been a pure specimen of either legal system. Ever since the Magna Carta and Edward I’s statutes, common-law countries have used legislation regularly if not extensively: suffice it to mention the statutes enacted by the English Parliament or by federal and state legislatures in the U.S., let alone the American federal Constitution and those of the states.

There is, however, agreement that before the 20th century statutes were used sparingly, almost looked upon as “a necessary evil which disturbed the harmony of the Common Law” (Zweigert and Kötz 1998). This view has been progressively abandoned in recent history,
so that in today common law systems, “legislation and regulations (...) have made case law less central to the common law” (Glendon et al. 1999 147).

This evolution of common law over time is entirely consistent with the presence of a mixed system as we have described it: in ancient England great social upheavals were sufficiently rare that the prevailing view was surely the one expressed in the famous motto of the barons convened at Merton in 1236: *Nolumus leges Angliae mutari*.

In the United States, since the late nineteenth century the pace of socio-economic progress has been much faster, and therefore legislative activity more intense. Already in 1848 New York adopted the first Code of Civil Procedure in the U.S., that soon after was emulated by other states. The “Progressive Era” (1887-1917) saw a dramatic increase in explicit regulation at the federal and state level (cf. Glaeser and Shleifer 2003) to deal with a variety of issues like antitrust, competition, pricing, etc. that had been traditionally considered as matters to be solved by private litigation.

Civil-law countries, on the other hand, have witnessed an increasing awareness of the decisive role of judges in the interpretation of the legal texts: “While there thus seems to have been a certain diminution in the role of precedent in the common law, court decisions have increasingly come to be treated as important sources of law in the civil law systems” (Glendon 1984). The idea of judicial activity as a merely passive application of a clear-cut pre-especified statute, although revealing of a legal mindset is, in practice, by and large an illusion. By their very nature, texts are ambiguous, incomplete and open to multiple readings in the light of a particular instance of conflict resolution: the gap between the text and the actual case has to be actively filled by the judge.

Although interpretation has always been necessary, it is historically true that its role was played down until recently, especially so at the time of the first attempts of massive codification in the Continent. Thus, for instance, Article 4 of the French Civil Code forbids a judge to refuse to decide a case “on the pretext that the law is silent, unclear or incomplete” (Glendon et al. 1999 142) and in the period of the promulgation of the German Civil Code of 1874, legal thinking in Germany was dominated by the Pandectist School, according to whose understanding the application of law was in terms of ‘logical necessity,’ issuing from the principles contained in the Roman law, and had nothing to do with the analysis of social

Contrary to those historical intentions, there is a growing consensus that statutory interpretation in civil law countries has drifted away from literalism (Dale 1977, Germain 2003). More recently, statutes have been losing their privileged position and a more or less covert use of case-law has become widespread (Zweigert and Kötz 1998), although civil law theory does not explicitly recognize adherence to a formal doctrine of *stare decisis*\textsuperscript{12}. As David and Brierley (1985) put it, in civil law countries “the creative role of judicial decisions is always, or nearly always, hidden behind the screen of an ‘interpretation’ of legislation” (134). In the field of corporate law, Enriques (2002) emphasizes that even in countries with strict bright line rules in their statutes “on the books,” the existence of general clauses in codes and their subsequent application and interpretation by good judges often does the job of shaping better law “off the books” and provided more investor-friendly legal environments\textsuperscript{13}.

Simultaneously, the theory of unfettered power of the legislature widely held in the eighteenth and nineteenth century on the European continent has been gradually replaced by an emphasis on checks and balances that owes much to the American constitutional experience.

Therefore, a case can be made that our analysis of legal system in a dynamically evolving framework yields not merely a normative, but also a positive conclusion: most advanced legal systems seem to be evolving towards institutions that combine the flexibility of legislation in responding to mutable conditions with the ability of case law gradually to improve legal rules through the action of a multitude of different decision-makers over time.

\textsuperscript{12}As Glendon et al. put it, “In view of the de facto importance of case law as a civil law authority, one might expect the differences between the common law and civil law systems in this area to diminish over time. Certainly, the presence or absence of a formal doctrine of stare decisis is not of crucial significance.” Merryman (1985) also acknowledges that “whatever the ideology of the revolution may say about the value of precedent, the fact is that courts do not act very differently toward reported decisions in civil law jurisdictions than do courts in the United States.” (47)

\textsuperscript{13}Enriques quotes Hertig (in press) saying that: “civil law jurisdictions are showing ‘common law creativeness’ in protecting minority shareholders.”
5 Conclusion

[...]
References


Madison, James, *The Federalist* 37, 1778, in *The Federalist Papers*.


A Appendix

A.1 Proof of Proposition 1

The first-order condition of the judge’s convex optimization problem implies that case law follows an AR(1) process

\[ x_{t+1} = \psi x_t + (1 - \psi) \hat{x}_{t+1} \]

Then, conditional on any starting point \( x_t \), the law at time \( t + s \) is the random variable

\[ x_{t+s} | x_t = \psi^s x_t + (1 - \psi) \sum_{\tau=1}^{s} \psi^{s-\tau} \hat{x}_{t+\tau} \]

whose expectation and variance are easily computed, and which converges with a factor \( \psi \) to

\[ \lim_{s \to \infty} x_{t+s} | x_t = (1 - \psi) \sum_{\tau=0}^{\infty} \psi^\tau \hat{x}_{t+\tau} \equiv x \]

Since all \( \psi^\tau \hat{x}_t \) are independent random variables each with finite mean \( \psi^\tau x^* \) and variance \( \psi^{2\tau} \sigma^2 \), the Central Limit Theorem implies that the asymptotic distribution is normal.

A.2 Proof of Proposition 2

The loss to the judge as a function of his ruling becomes

\[ L_j (x_t) = \sum_{s=0}^{\infty} \beta_j^s \left[ (E_t (x_{t+s}) - \hat{x}_t)^2 + Var_t (x_{t+s}) \right] + \frac{\psi}{1 - \psi} (x_t - x_{t-1})^2 \]

where the previous case is encompassed for \( \beta_j = 0 \).

We guess that judges follow a linear rule

\[ x_{t+1} = ax_t + (1 - a) \hat{x}_{t+1} + b (\hat{x}_{t+1} - x^*) = \]

\[ ax_t + (1 - a + b) \hat{x}_{t+1} - bx^* \]

for some \( a \in (0, 1) \) and \( b > 0 \). Then

\[ x_{t+s} | x_t = a^s x_t + (1 - a + b) \sum_{i=1}^{s} a^{s-i-1} \hat{x}_{t+i} - \frac{1 - a^s}{1 - a} bx^* \]

so that

\[ E_t (x_{t+s}) = a^s x_t + (1 - a^s) x^* \]

and

\[ Var_t (x_{t+s}) = \frac{1 - a^{2s}}{1 - a^2} (1 - a + b)^2 \sigma^2 \]

Hence the judge’s ruling is

\[ x_t = \arg \min_x \sum_{s=0}^{\infty} \beta_j^s \left[ a^s x + (1 - a^s) x^* - \hat{x}_t \right]^2 + \frac{\psi}{1 - \psi} (x - x_{t-1})^2 \]
with first-order condition
\[
\sum_{s=0}^{\infty} \beta_j a^s [a^s x + (1 - a^s) x^* - \hat{x}_t] + \frac{\psi}{1 - \psi} (x - x_{t-1}) = 0
\]
\[
\left( \frac{1}{1 - \beta_j a^2} + \frac{\psi}{1 - \psi} \right) x + \frac{\beta_j a (1 - a)}{1 - \beta_j a (1 - \beta_j a^2)} x^* - \frac{1}{1 - \beta_j a} \hat{x}_t - \frac{\psi}{1 - \psi} x_{t-1} = 0
\]
\[
x = \frac{\psi (1 - \beta_j a^2)}{1 - \beta_j \psi a^2} x_{t-1} + \frac{(1 - \psi) (1 - \beta_j a^2)}{(1 - \beta_j a) (1 - \beta_j \psi a^2)} \hat{x}_t - \frac{(1 - \psi) \beta_j a (1 - a)}{(1 - \beta_j a) (1 - \beta_j \psi a^2)} x^*
\]
so that the equilibrium is defined by
\[
a = \frac{\psi (1 - \beta_j a^2)}{1 - \beta_j \psi a^2} \in (0, 1)
\]
\[
b = \frac{(1 - \psi) \beta_j a (1 - a)}{(1 - \beta_j a) (1 - \beta_j \psi a^2)} \geq 0
\]
since indeed
\[
\frac{\psi (1 - \beta_j a^2)}{1 - \beta_j \psi a^2} + \frac{(1 - \psi) (1 - \beta_j a^2)}{(1 - \beta_j a) (1 - \beta_j \psi a^2)} - \frac{(1 - \psi) \beta_j a (1 - a)}{(1 - \beta_j a) (1 - \beta_j \psi a^2)} = 1
\]
We can then simplify
\[
b = \frac{\beta_j a (1 - a)^2}{1 - \beta_j a}
\]
to express the conditional variance as
\[
Var_t (x_{t+s}) = (1 - a^{2s}) \frac{1 - a}{1 + a} \left( \frac{1 - \beta_j a^2}{1 - \beta_j a} \right)^2 \hat{\sigma}^2
\]
The factor a is implicitly defined by
\[
C (a) \equiv \beta_j \psi a^3 - \beta_j \psi a^2 - a + \psi = 0
\]
which has a unique root \( a \in (0, \psi) \): existence follows from
\[
C (0) = \psi > 0 \land C (\psi) = -\beta_j \psi^3 (1 - \psi) \leq 0
\]
and uniqueness from considering the derivative
\[
C' (a) = 3 \beta_j \psi a^2 - 2 \beta_j \psi a - 1 < 0 \forall a \in (0, 1)
\]
since it is a convex quadratic with
\[
C' (0) = -1 < 0 \land C' (a) = \beta_j \psi - 1 < 0
\]
Totally differentiating the implicit definition yields the unambiguous comparative statics.

\[
\frac{\partial a}{\partial \beta_j} = \frac{(1 - a) \psi a^2}{C'(a)} < 0 \\
\frac{\partial a}{\partial \psi} = -\frac{1 - \beta_j a^2 (1 - a)}{C'(a)} > 0
\]

Tedious algebra finally yields

\[
\frac{\partial \text{Var}_t (x_{t+s})}{\partial \beta_j} = 2 (1 - a^2 s) \frac{a (1 - a)^2 (1 - \beta_j a^2)}{(1 + a) (1 - \beta_j a)^3} \hat{\sigma}^2 + \frac{2}{1 + \alpha} \left( \frac{1 - \beta_j a}{1 - \beta_j a^2} \right)^2 \hat{\sigma}^2 \frac{\partial a}{\partial \beta_j}
\]

which is unambiguously positive.

### A.3 Proof of Corollary 1

We have proved that, conditional on exactly \( c \) changes having occurred between \( t \) and \( t + s \), the law has expectation and variance

\[
E (x_{t+s}|x_t, c) = x^* + \psi^c (x_t - x^*) \\
\text{Var} (x_{t+s}|x_t, c) = \frac{1 - \psi}{1 + \psi} (1 - \psi^2 c) \hat{\sigma}^2
\]

and therefore

\[
E (x_{t+s}^2|x_t, c) = \frac{1 - \psi}{1 + \psi} (1 - \psi^2 c) \hat{\sigma}^2 \left[ x^* + \psi^c (x_t - x^*) \right]^2 = (x^*)^2 + \frac{1 - \psi}{1 + \psi} \hat{\sigma}^2 + 2 \psi^c (x_t - x^*) x^* + \psi^2 c \left[ (x_t - x^*)^2 - \frac{1 - \psi}{1 + \psi} \hat{\sigma}^2 \right]
\]

Recalling Newton’s binomial formula

\[
(a + b)^n = \sum_{k=0}^{n} \frac{n!}{k! (n-k)!} a^{n-k} b^k
\]

we can directly compute the moments of \( x_t|x_0 \), namely expectation

\[
E (x_{t+s}|x_t) = \sum_{c=0}^{s} \frac{s!}{c! (s-c)!} (1 - p)^{s-c} p^c [x^* + \psi^c (x_t - x^*)] = x^* + (1 - p + p\psi)^s (x_t - x^*)
\]
and variance
\[ \text{Var}(x_{t+s}|x_t) = E(x_{t+s}^2|x_t) - [E(x_{t+s}|x_t)]^2 = \]
\[ = \sum_{c=0}^{s} \frac{s!}{c!(s-c)!} (1-p)^{s-c} p^c E(x_{t+s}^2|x_t, c) - [E(x_{t+s}|x_t)]^2 = \]
\[ = \frac{1-\psi}{1+\psi} [1 - (1-p + p\psi^2)^s] \sigma^2 + [(1-p + p\psi^2)^s - (1-p + p\psi)^2] (x_t - x^*)^2 \]

A.4 Proof of Proposition 3

Under this generalization the evolution of case law is defined by the Markov chain
\[ x_{t+1} = \begin{cases} 
  x_t & \text{w/ Pr } 1-p(x_t) \\
  \psi x_t + (1-\psi) \tilde{x}_{t+1} & \text{w/ Pr } p(x_t)
\end{cases} \]
whose transition probability function has point mass \(1-p(x_t)\) at \(x_t\), and density on \(\mathbb{R}\)
\[ p(x_{t+1}|x_t) = p(x_t) \hat{f}\left(\frac{x_{t+1} - \psi x_t}{1-\psi}\right) \]
where \(\hat{f}(\tilde{x})\) denotes the density of the distribution of judges’ preferences.\(^{14}\)

This Markov chain is irreducible (the transition probability function has full support \(\mathbb{R}\) for every \(x_t \in \mathbb{R}\), aperiodic (there is no state to which the process necessarily returns within a fixed time period) and positive recurrent (the expected time for the process to return to any state is finite: a less obvious property that is well-known for autoregressive Markov chains with normally distributed innovations). As a consequence it has a unique stationary distribution which coincides with its asymptotic distribution. This is a continuous distribution whose density \(\mu(x)\) is implicitly defined by the steady-state equation
\[ p(x) \mu(x) = \int_{-\infty}^{\infty} \varphi\left(\frac{x - \psi y}{1-\psi}\right) p(y) \mu(y) dy \]

\(^{14}\)The conditional distribution of \(x_{t+s}|x_t\) has point mass \([1-p(x_t)]^s\) at \(x_t\), and density on \(\mathbb{R}\) that can be expressed recursively as
\[ p^s(x_{t+s}|x_t) = \begin{cases} 
  [1-p(x_t)] p^{s-1}(x_{t+s}|x_t) + p(x_{t+s}|x_t) [1-p(x_{t+s})]^{s-1} + \\
  + \int_{-\infty}^{\infty} p(x_{t+1}|x_t) p^{s-1}(x_{t+s}|x_{t+1}) dx_{t+1}
\end{cases} \]
or explicitly
\[ p^s(x_{t+s}|x_t) = \sum_{J=1}^{s} \sum_{J=0}^{\binom{s}{J}} \int_{J-1}^{\infty} \int_{-\infty}^{\infty} [1-p(x_t)]^{n_0} [1-p(x_{t+s})]^{n_J} p(x_{t+s}|x_{t+J-1}) \\
\sum_{j=0}^{\binom{s}{J}} \int_{J=0}^{\infty} [1-p(x_{t+j})]^{n_j} p(x_{t+J}|x_{t+j-1}) dx_{t+j} \]
where \(J\) counts the number of rulings between \(t\) and \(t+s\); by convention the multiple integral is to be read as \([1-p(x_t)]^{n_0} [1-p(x_{t+s})]^{n_J} p(x_s|x_0)\) for \(J = 1\); and the second summation is over all possible multisets of length \(s-J\) (i.e. the number of periods without changes) over \(J+1\) elements (i.e. the number of different states of the law observed from \(t\) to \(t+s\) inclusive), so that the number of its terms equals the multinomial coefficient \([J, s-J] = \binom{s}{J}\). Computation in closed form is not possible unless we revert to the case of constant \(p\).
For ease of notation, define \( \lambda(x) \equiv p(x) \mu(x) \). Since \( \hat{x} \sim N(x^*, \sigma^2) \) implies

\[
\hat{f}(\hat{x}) = \frac{1}{2\sqrt{\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left( \frac{\hat{x} - x^*}{\sigma} \right)^2 \right\}
\]

the stationarity condition can be rewritten as

\[
\lambda(x) = \int_{-\infty}^{\infty} \frac{1}{2\sqrt{\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left( \frac{x - x^* - \psi(y - x^*)}{(1 - \psi)\sigma} \right)^2 \right\} \lambda(y) \, dy
\]

whose solution is, up to an arbitrary scale constant

\[
\lambda(x) = \exp\left\{ -\frac{(1 + \psi)(x - x^*)^2}{2(1 - \psi)\sigma^2} \right\}
\]

Therefore the asymptotic density is exactly computed as

\[
\mu(x) = \frac{\kappa}{p(x)} \frac{1}{2\sqrt{\pi\sigma^2}} \exp\left\{ -\frac{(1 + \psi)(x - x^*)^2}{2(1 - \psi)\sigma^2} \right\} = \frac{\kappa}{p(x)} \varphi(x)
\]

where \( \varphi(x) \) is of course a normal density with mean \( x^* \) and variance \( \frac{1}{1 + \psi}\sigma^2 \), while the scale constant is \( \kappa = \left[ \int_{-\infty}^{\infty} \frac{1}{p(x)} \varphi(x) \, dx \right]^{-1} \) ensuring that \( \mu(.) \) is properly normalized as a probability density.

The expectation of the asymptotic distribution is

\[
E(x) = \int_{-\infty}^{\infty} \frac{\kappa}{p(x)} \varphi(x) \, dx = x^*
\]

since both \( p(x) \) and \( \varphi(x) \) are symmetric around \( x^* \); and the variance is

\[
Var(x) = \frac{1}{\int_{-\infty}^{\infty} \frac{1}{p(x)} \varphi(x) \, dx} \int_{-\infty}^{\infty} \frac{1}{p(x)} \varphi(x) \, dx = E_{\varphi}\left[ \frac{1}{p(x)} \right] \left[ (x - x^*)^2 \right] + \frac{\text{Cov}_{\varphi}\left[ \frac{1}{p(x)}, (x - x^*)^2 \right]}{E_{\varphi}\left[ \frac{1}{p(x)} \right]}
\]

\[
= \frac{1 - \psi}{1 + \psi} \sigma^2 + \frac{\text{Cov}_{\varphi}\left[ \frac{1}{p(x)}, (x - x^*)^2 \right]}{E_{\varphi}\left[ \frac{1}{p(x)} \right]} < \frac{1 - \psi}{1 + \psi} \sigma^2
\]

since both \( p(x) \) and \( (x - x^*)^2 \) are positive and their derivatives have opposite sign everywhere on \( \mathbb{R} \).

### A.5 Proof of Lemma 1

Both \( \hat{x}_J \) and \( \hat{x}_K \) have expectation \( x^* \), so the only consideration concerns their variances, respectively

\[
\hat{\sigma}^2_J = b^2\hat{\sigma}^2_m
\]
and
\[ \sigma^2_K = \left[ k^2 + \frac{(1 - k)^2}{M} \right] \sigma^2_m \]

\( b \) is a direct measure of magnates’ power, so that the first result follows from
\[ \frac{\partial \sigma^2_J}{\partial b} = 2b\sigma^2_m > 0 \forall b \in (0, 1) \]

while \( k \) is an inverse measure of magnates’ power, so that the second result follows from
\[ \frac{\partial \sigma^2_K}{\partial k} = \frac{2(M + 1)k - 1}{M} \sigma^2_m > 0 \forall k \in \left( \frac{1}{M + 1}, 1 \right) \]

### A.6 Proof of Proposition 4

The total loss under statute law and case law is respectively

\[ L^K_0 = \sum_{t=0}^{\infty} \beta^t E [L_K (x_{1+t})] = \frac{\dot{\sigma}^2_K}{1 - \beta} \]

and

\[ L^J_0 = \sum_{t=0}^{\infty} \beta^t E [L_J (x_{1+t})] = \sum_{t=0}^{\infty} \beta^t \frac{1 - \psi + 2\psi^{2t+1}}{1 + \psi} \sigma^2_J = \frac{1 - 2\beta\psi + \beta\psi^2}{1 - \beta^2} \sigma^2_J \]

The \emph{ex-ante} socially optimal choice of \( \psi \) is therefore
\[ \psi^* = \arg \min_{\psi \in (0, 1)} \frac{1 - 2\beta\psi + \beta\psi^2}{1 - \beta^2} = \frac{1 - \sqrt{1 - \beta}}{\beta} \]

implying
\[ L^J_0 = \frac{\dot{\sigma}^2_J}{\sqrt{1 - \beta}} \]

### A.7 Proof of Lemma 2

The variance of \( \hat{x}_L \) is
\[ \sigma^2_L = \nu^2 \sigma^2_s \in (0, \sigma^2) \Rightarrow \frac{\partial \sigma^2_L}{\partial \nu} = 2\nu \frac{\sigma^2}{S} > 0 \]

and the variance of \( \hat{x}_J \) is
\[ \sigma^2_J = (1 - b)^2 \sigma^2_L + 2b (1 - b) \text{Cov}(\hat{x}_L, \hat{x}_s) + b^2 \sigma^2 = \]
\[ = \{ [\nu + (1 - \nu) b]^2 + b^2 (S - 1) \} \sigma^2_s \in (\sigma^2_L, \sigma^2) \]
\[ \Rightarrow \frac{\partial \sigma^2_J}{\partial \nu} = 2 (1 - b) [\nu + (1 - \nu) b] \frac{\sigma^2_s}{S} > 0 \]
\[ \Rightarrow \frac{\partial \sigma^2_J}{\partial b} = 2 \{ (1 - \nu) [\nu + (1 - \nu) b] + b (S - 1) \} \frac{\sigma^2_s}{S} > 0 \]

54
so that
\[
\frac{\hat{\sigma}_L^2}{\hat{\sigma}_J^2} = \frac{\theta^2}{[\theta + (1 - \theta) b]^2 + b^2 (S - 1)} \in (0, \theta)
\]
\[
\Rightarrow \frac{\partial (\hat{\sigma}_L^2/\hat{\sigma}_J^2)}{\partial \theta} = 2b \frac{(1 - b) \theta + bS}{\{[\theta + (1 - \theta) b]^2 + b^2 (S - 1)\}^2} > 0
\]
\[
\Rightarrow \frac{\partial (\hat{\sigma}_L^2/\hat{\sigma}_J^2)}{\partial b} = -2b^2 (1 - \theta) [\theta + (1 - \theta) b] + b (S - 1) \frac{\theta}{\{[\theta + (1 - \theta) b]^2 + b^2 (S - 1)\}^2} < 0
\]

A.8 Proof of Lemma 3 and Proposition 5

The unconditional expectation of the total loss under statute law and case law is respectively
\[
E[L_L(t)] = \sum_{s=0}^{\infty} \beta^s E[L_L(x_{t+s})] = \sum_{s=0}^{\infty} \beta^s \hat{\sigma}_L^2 = \frac{\hat{\sigma}_L^2}{1 - \beta}
\]
and
\[
E[L_J(t)] = \sum_{s=0}^{\infty} \beta^s E[L_J(x_{t+s})] = \sum_{s=0}^{\infty} \beta^s \frac{1 - \psi + 2\psi^{2(t+s)-1}}{1 + \psi} \hat{\sigma}_J^2 =
\]
\[
= \frac{1 - \psi}{1 + \psi} \frac{\hat{\sigma}_J^2}{1 - \beta} + \frac{2\psi^{2t-1}}{(1 + \psi)(1 - \beta \psi^2)} \hat{\sigma}_J^2
\]

The second part of the condition follows from the observation that
\[
\frac{1 - \beta}{1 - \beta \psi^2} \leq 1 \forall \beta, \psi \in [0, 1]
\]

Under the normality assumption, in period \(t\) since the creation of a legal system, the unconditional distribution of statute law is
\[
x_L(t) = \hat{x}_{L,t}
\]
\[
\Rightarrow x_L \sim N(x^*, \hat{\sigma}_L^2)
\]
while the unconditional distribution of case law
\[
x_J(t) = \psi^t \hat{x}_{J,t} + (1 - \psi) \sum_{s=1}^{t} \psi^{t-s} \hat{x}_{J,s} = \psi^{t-1} \hat{x}_{J,1} + (1 - \psi) \sum_{s=2}^{t} \psi^{t-s} \hat{x}_{J,s}
\]
\[
\Rightarrow x_J(t) \sim N\left(x^*, \frac{1 - \psi + 2\psi^{2(t-1)-1}}{1 + \psi} \hat{\sigma}_J^2\right)
\]

Thus the inefficiency from either system has a chi-square distribution with one degree of freedom
\[
\frac{(\hat{x}_{L,t} - x^*)^2}{\hat{\sigma}_L^2} \sim \chi_1^2 \quad \text{and} \quad \frac{(\hat{x}_{J,t} - x^*)^2}{\hat{\sigma}_J^2} \sim \chi_1^2
\]
and the ratio of the two has an \(F\) distribution with \(1,1\) degrees of freedom
\[
\frac{1 + \psi}{1 - \psi + 2\psi^{2t-1}} \frac{\hat{\sigma}_L^2}{\hat{\sigma}_J^2} \frac{(\hat{x}_{J,t} - x^*)^2}{(\hat{x}_{L,t} - x^*)^2} \sim F_{1,1}
\]
A.9 Derivation of Figure 8

The total loss under statute law and case law is respectively

\[ L_L (t) = \sum_{s=0}^{\infty} \beta^s E [l_L (x_{t+s}) | x_t] = (x_{L,t} - x^*)^2 + \frac{\beta \sigma_L^2}{1 - \beta} \]

and

\[ L_J (t) = \sum_{s=0}^{\infty} \beta^s E [l_J (x_{t+s}) | x_{t-1}] = \sum_{s=0}^{\infty} \beta^s \left[ \left( E (x_{t+s} | x_{t-1}) - x^* \right)^2 + Var (x_{t+s} | x_{t-1}) \right] = \]

\[ = \sum_{s=0}^{\infty} \beta^s \left[ \psi^{2s+2} (x_{t-1} - x^*)^2 + \frac{1 - \psi}{1 + \psi} (1 - \psi^{2s+2}) \sigma_J^2 \right] = \]

\[ = \frac{\psi^2}{1 - \beta \psi^2} (x_{J,t-1} - x^*)^2 + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \sigma_J^2 \]

As previously shown

\[ \frac{(x_{L,t} - x^*)^2}{\sigma_L^2} \sim \chi^2_1 \]

\[ \frac{1 + \psi}{1 - \psi + 2 \psi^{2t-3}} \frac{(x_{J,t-1} - x^*)^2}{\sigma_J^2} \sim \chi^2_1 \]

and therefore

\[ \Pr \{ L_L (t) \geq L_J (t) \} = \]

\[ = \Pr \left\{ (x_{L,t} - x^*)^2 + \frac{\beta \sigma_L^2}{1 - \beta} \geq \frac{\psi^2}{1 - \beta \psi^2} (x_{J,t-1} - x^*)^2 + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \sigma_J^2 \right\} = \]

\[ = \Pr \left\{ \chi^2_1, \chi^2_{1,J} \leq \left[ \frac{\psi^2}{1 - \beta \psi^2} \frac{\sigma_L^2}{\sigma_J^2} \frac{\beta}{1 - \beta} - \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \right] \left( \frac{1 - \beta \psi^2}{(1 - \beta \psi^2)} \right) \right\} = \]

\[ = \int_0^\infty F_{\chi^2_{1,J}} \left( \left[ \frac{\sigma_L^2}{\sigma_J^2} + \frac{\beta}{1 - \beta} \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \right] \right) f_{\chi^2_{1,J}} (L) dL \]

A.10 Proof of Lemma 4

For statute law

\[ L_L (x^*) = E \left[ (x_{L,t} - x^*)^2 \right] + \frac{\beta \sigma_L^2}{1 - \beta} = \frac{\sigma_L^2}{1 - \beta} \]

since the new legislature has preferences \( \hat{x}_L \) with expectation \( x^* \) and variance \( \sigma_L^2 \) and

But for case law, considering the inherited precedent \( x_{J,t-1} \)

\[ L_J (x^*) = \frac{\psi^2}{1 - \beta \psi^2} (x_{J,t-1} - x^*)^2 + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \sigma_J^2 \]

\[ \frac{1}{1 - \beta \psi^2} (x_{J,t-1} - x^*)^2 + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \sigma_J^2 \]
and the unconditional expectation is

\[ E [L_J (x^*)] = \frac{\psi^2}{1 - \beta \psi^2} E [(x_{J, t-1} - x^*)^2] + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \hat{\sigma}_J^2 = \frac{\psi^2}{1 - \beta \psi^2} \left[ (x^*_0 - x^*)^2 + \frac{1 - \psi + 2 \psi^{2t-3}}{1 + \psi} \hat{\sigma}_J^2 \right] + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \hat{\sigma}_J^2 \]

Hence after a change in the optimum from \( x^*_0 \) to \( x^* \) statute law is expected to be preferable to case law if and only if

\[ \frac{\psi^2}{1 - \beta \psi^2} (x^*_0 - x^*)^2 + \frac{2 \psi^{2t-1}}{(1 - \beta \psi^2) (1 + \psi)} \hat{\sigma}_J^2 \geq \frac{\hat{\sigma}_L^2}{1 - \beta} - \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \]

which converges asymptotically to, and is always stricter than

\[ \frac{\psi^2}{1 - \beta \psi^2} (x^*_0 - x^*)^2 \geq \frac{\hat{\sigma}_L^2}{1 - \beta} - \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \]

The only non-trivial result in the second part of the proposition is the role of \( \beta \); rewrite:

\[ (x^*_0 - x^*)^2 \geq \xi (\beta, \psi, \hat{\sigma}_L^2 \hat{\sigma}_J^2) \equiv \frac{1 - \beta \psi^2}{\psi^2} \left[ \frac{\hat{\sigma}_L^2}{1 - \beta} - \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \right] \]

and differentiate

\[ \frac{\partial \xi}{\partial \beta} = \frac{1 - 2 \beta \psi^2 + \beta^2 \psi^2}{(1 - \beta)^2} \left[ \frac{\hat{\sigma}_L^2}{\psi^2} - \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \right] = \frac{(1 - \beta \psi^2)^2 + \beta^2 \psi^2 (1 - \psi)}{(1 - \beta)^2} \left[ \frac{\hat{\sigma}_L^2}{\psi^2} - \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \right] \]

So

\[ \hat{\sigma}_L^2 > \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \Rightarrow \frac{\partial \xi}{\partial \beta} > 0 \]

and if \( \hat{\sigma}_L^2 < \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \) statute law is always expected to be preferable.

### A.11 Proof of Proposition 6

The loss under the mixed system is

\[ L_M (x^*) = \frac{\psi^2}{1 - \beta \psi^2} E [(x_L - x^*)^2] + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \hat{\sigma}_J^2 = \frac{\psi^2}{1 - \beta \psi^2} \hat{\sigma}_L^2 + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \hat{\sigma}_J^2 \]
which is better than civil law if
\[
\frac{\sigma_L^2}{1 - \beta} \geq \frac{\psi^2}{1 - \beta \psi^2} \sigma_L^2 + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \frac{\sigma_J^2}{1 - \beta}
\]
and it is better than pure common law if
\[
(x_{,t-1}^* - x^*)^2 \geq \sigma_L^2
\]
which implies that in expectation the mixed system is preferable if and only if
\[
(x^* - x)^2 + \frac{1 - \psi + 2 \psi^{2t-3}}{1 + \psi} \sigma_J^2 > (x^* - x)^2 + \frac{1 - \psi}{1 + \psi} \sigma_J^2 \geq \sigma_L^2
\]

**A.12 Proof of Lemma 5**

The common-law loss is
\[
L_J (x_0, x_1^*) = \sum_{t=0}^{\infty} \beta^t E \left[ (x_{1+t} - x_{1+t}^*)^2 | x_0, x_1^* \right] = \sum_{t=0}^{\infty} \beta^t \left[ E (x_{1+t} | x_0, x_1^*) - E (x_{1+t}^* | x_0, x_1^*) \right]^2 + Var (x_{1+t} | x_0, x_1^*)^2 + 2 Cov (x_{1+t}, x_{1+t}^* | x_0, x_1^*) + Var (x_{1+t}^* | x_0, x_1^*)
\]
which includes two new sources of uncertainty: one is the exogenous variance in future optima, while the other is the inability of future law instantly to adapt to preference changes.

Since
\[
x_t | x_0 = \psi^t x_0 + (1 - \psi) \sum_{s=1}^{t} \psi^{t-s} \tilde{x}_s
\]
we compute
\[
E (x_{1+t} | x_0, x_1^*) = \psi^{t+1} x_0 + (1 - \psi) \sum_{s=0}^{t} \psi^{t-s} E (x_{1+s}^* | x_1^*)
\]
\[
Var (x_{1+t} | x_0, x_1^*) = \frac{1 - \psi}{1 + \psi} (1 - \psi^{2t+2}) \sigma_J^2 + (1 - \psi)^2 \sum_{s=0}^{t} \psi^{2(t-s)} Var (x_{1+s}^* | x_1^*)
\]
\[
Cov (x_{1+t}, x_{1+t}^* | x_0, x_1^*) = (1 - \psi) \sum_{s=0}^{t} \psi^{t-s} Cov (x_{1+s}^*, x_{1+t}^* | x_1^*)
\]
Given our assumption about the Poisson process for changes in the social optimum
\[
E (x_{1+s}^* | x_1^*) = x_1^*
\]
\[
Var (x_{1+s}^* | x_1^*) = E \left[ (x_{1+s}^* - x_1^*)^2 | x_1^* \right] = \sum_{j=0}^{s} \binom{s}{j} \lambda^j (1 - \lambda)^{s-j} j \sigma_\zeta^2 = s \lambda \sigma_\zeta^2
\]
\[ \text{Cov}_t(x^*_{1+s}, x^*_{1+t} | x^*_t) = E \left[ (x^*_{1+s} - x^*_t) (x^*_{1+t} - x^*_t) | x^*_t \right] = \]
\[ E \left[ (x^*_{1+t} - x^*_t) (x^*_{1+s} - x^*_t) + (x^*_{1+s} - x^*_t)^2 | x^*_t \right] = \]
\[ E \left[ (x^*_{1+t} - x^*_t) (x^*_{1+s} - x^*_t) \right] + \text{Var} \left( x^*_{1+s} | x^*_t \right) = \]
\[ \text{Var} \left( x^*_{1+s} | x^*_t \right) \forall s \leq t \]

and therefore, recalling that
\[ \sum_{s=0}^{t} sa^{-s} = \frac{(1-a) t - a (1 - a^t)}{(1-a)^2 a^t} \]

we compute
\[ L_J (x_0, x_1^*) = \sum_{t=0}^{\infty} \beta^t \left[ \psi^{2t+2} (x_0 - x_1^*)^2 + \frac{1-\psi}{1+\psi} \left( 1 - \psi^{2t+2} \right) \sigma^2_J + \right. \]
\[ \left. \left( 1 - \psi \right)^2 \sum_{s=0}^{t} \psi^{2(t-s)} s - 2 \left( 1 - \psi \right) \sum_{s=0}^{t} \psi^{t-s} s + t \right] \lambda \sigma^2_\xi = \]
\[ = \sum_{t=0}^{\infty} \beta^t \left[ \psi^{2t+2} (x_0 - x_1^*)^2 + \frac{1-\psi}{1+\psi} \left( 1 - \psi^{2t+2} \right) \sigma^2_J + \right. \]
\[ \left. \left( 2 \psi (1-\psi^t) - \psi^2 (1-\psi^{2t}) \right) \lambda \sigma^2_\xi \right] = \]
\[ = \frac{\psi^2}{1 - \beta \psi^2} (x_0 - x_1^*)^2 + \frac{(1-\psi)^2}{(1 - \beta \psi^2)} \sigma^2_J + \frac{\beta \psi \left( 2 - \psi - \beta \psi^2 \right)}{(1 - \beta \psi) (1 - \beta \psi^2)} \sigma^2_\xi \]

Note that the unconditional expectation given \( T \) previous periods is
\[ E \left[ (x_0 - x_1^*)^2 \right] = E \left[ (x_0 - x_0^*)^2 + (x_0^* - x_1^*)^2 \right] = \]
\[ = E \left[ (x_0 - x_0^*)^2 \right] + \text{Var} \left( x_1^* | x_0^* \right) = \]
\[ = \frac{1 - \psi + 2 \psi^{2T+3}}{1 + \psi} \sigma^2_J + \frac{1 + 2 \psi - (2 + \psi) \psi^T+1}{1 - \psi^2} \lambda \sigma^2_\xi = \]
\[ = \frac{1 - \psi}{1 + \psi} \sigma^2_J + \frac{1 + 2 \psi}{1 - \psi^2} \lambda \sigma^2_\xi + \frac{2 \psi^{2T+3}}{1 + \psi} \sigma^2_J - \frac{(2 + \psi) \psi^{T+1}}{1 - \psi^2} \lambda \sigma^2_\xi \]

because with a history of \( T \) periods of common law before the present
\[ E \left[ (x_0 - x_0^*)^2 \right] = \frac{1 - \psi + 2 \psi^{2T+3}}{1 + \psi} \sigma^2_J + \left[ \frac{2 \psi (1 - \psi^T)}{1 - \psi} - \psi^2 \frac{(1 - \psi^T)}{1 - \psi^2} \right] \lambda \sigma^2_\xi \]

Therefore
\[ E \left[ L_J (x_0, x_1^*) \right] = \frac{1 - \psi}{1 + \psi} \sigma^2_J + \]
\[ + \frac{1}{1 - \beta \psi^2} \left[ \frac{(1 - \beta) (1 + 2 \psi) \psi^2}{(1 - \psi^2)} + \frac{\beta \psi \left( 2 - \psi - \beta \psi^2 \right)}{(1 - \beta \psi^2)} \right] \lambda \sigma^2_ \xi + \]
\[ + \frac{\psi^{T+3}}{(1 - \beta \psi^2) (1 - \psi^2)} \left[ 2 (1 - \psi) \psi^{T+1} \sigma^2_J - (2 + \psi) \lambda \sigma^2_\xi \right] \]

59
and statute law is expected to be better than case law, asymptotically, if

\[
\frac{1}{1 - \beta \psi^2} \left[ \frac{(1 - \beta) (1 + 2 \psi) \psi^2}{(1 - \psi^2)} + \frac{\beta \psi (2 - \psi - \beta \psi^2)}{(1 - \beta \psi^2)} \right] \lambda \sigma_\xi^2 \geq \hat{\sigma}_L^2 - \frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2
\]

Let

\[
\Lambda (\beta, \psi) \equiv \frac{1}{1 - \beta \psi^2} \left( \frac{(1 - \beta) (1 + 2 \psi) \psi^2}{(1 - \psi^2)} + \frac{\beta \psi (2 - \psi - \beta \psi^2)}{(1 - \beta \psi^2)} \right) \geq 0
\]

The function ranges on all \( \mathbb{R}^+ \)

\[
\Lambda (\beta, 0) = 0 \land \lim_{\psi \to 1} \Lambda (\beta, \psi) = +\infty
\]

and is monotone increasing in \( \psi \)

\[
\frac{\partial \Lambda}{\partial \psi} (\beta, \psi^2) = \frac{2 \beta (1 - \beta) (1 + 2 \psi) \psi^3}{(1 - \beta \psi^2)^2 (1 - \psi^2)} + \frac{\beta^2 \psi^2 (3 - 2 \psi - \beta \psi^2) + 2 \beta (1 - \psi) + 2 (1 - \beta) \psi (1 + 3 \psi - \psi^3)}{(1 - \beta \psi^2)^3} > 0
\]

**Remark 1** Instead the sign of \( \frac{\partial \Lambda}{\partial \beta} \) is ambiguous

\[
\frac{\partial \Lambda}{\partial \beta} (\beta, \psi^2) = \frac{\psi}{(1 - \beta \psi^2)} \left[ \frac{(2 - \psi - \beta \psi^3)}{(1 - \beta \psi^2)^2} + \frac{(1 - \beta) (1 + 2 \psi) \psi^3}{(1 - \psi^2) (1 - \beta \psi^2)} - \frac{(1 + 2 \psi) \psi}{(1 - \psi^2)} \right]
\]

*It is positive to the left of the curve in the graph*
Remark 2 If the process social change starts at 0 the unconditional expectation given T previous periods of evolution with a constant optimum is

\[ E \left[ (x_0 - x_1)^2 \right] = E \left[ (x_0 - x_0^*)^2 + (x_1^* - x_1)^2 \right] = E \left[ (x_0 - x_0^*)^2 \right] + \text{Var} (x_1|x_0^*) = \frac{1 - \psi + 2\psi^{2T+3}}{1 + \psi} \sigma_j^2 + \lambda \sigma_\xi^2 \]

and therefore

\[ E[L(x_0, x_1^*)] = \frac{\psi^2}{1 - \beta \psi^2} E \left[ (x_0 - x_1^*)^2 \right] + \frac{(1 - \psi)^2}{(1 - \beta \psi^2)} \frac{\sigma_j^2}{1 - \beta} + \frac{\beta \psi (2 - \psi - \beta \psi^2)}{(1 - \beta \psi)(1 - \beta \psi^2)} \frac{\lambda \sigma_\xi^2}{1 - \beta} = \]

\[ = \frac{1 - \psi}{1 + \psi} \frac{\sigma_j^2}{1 - \beta} + \frac{[2 \beta (1 - \psi) + \psi (1 - \beta \psi)] \psi \lambda \sigma_\xi^2}{(1 - \beta \psi)(1 - \beta \psi^2)} + \frac{2\psi^{2T+5}}{(1 + \psi)(1 - \beta \psi^2)} \sigma_j^2 \]

so that civil law is preferred if, and asymptotically only if

\[ \frac{[2 \beta (1 - \psi) + \psi (1 - \beta \psi)] \psi \lambda \sigma_\xi^2}{(1 - \beta \psi)(1 - \beta \psi^2)} \geq \sigma_j^2 - \frac{1 - \psi}{1 + \psi} \sigma_j^2 \]

Let

\[ \Lambda (\beta, \psi) \equiv \frac{[2 \beta (1 - \psi) + \psi (1 - \beta \psi)] \psi}{(1 - \beta \psi)(1 - \beta \psi^2)} \geq 0 \]

so that

\[ \Lambda (\beta, 0) = 0 \land \Lambda (\beta, 1) = \frac{1}{1 - \beta} \]

\[ \Lambda (0, \psi) \equiv \psi^2 \land \Lambda (1, \psi) \equiv \frac{(2 + \psi) \psi}{(1 - \psi^2)} \]

and

\[ \frac{\partial \Lambda}{\partial \beta} = \psi^2 + 2\beta^2 \psi^4 + \beta^2 \psi^5 - 2\psi - 2\beta^2 \psi^3 - 2\beta \psi^4 > 0 \]

\[ \frac{\partial \Lambda}{\partial \psi} = 2\beta + \psi + 2\beta^2 \psi^2 + \beta^2 \psi^3 + \beta^2 \psi^4 - 2\beta \psi - 2\beta \psi^2 - 2\beta^3 \psi^3 > 0 \]

A.13 Proof of Proposition 7

Since a statute is introduced every time there is social change, after a change has occurred and immediately before the ensuing statute is emanated

\[ L_M (x_t^* \neq x_t^*) = \sum_{t=0}^{\infty} \beta^t (1 - \lambda)^t \left[ \psi^{2t+2} \sigma_L^2 + \frac{1 - \psi}{1 + \psi} (1 - \psi^{2t+2}) \sigma_j^2 \right] + \]

\[ + \sum_{t=1}^{\infty} \beta^t (1 - \lambda)^{t-1} \lambda L_M (x_t^* \neq x_t^*) \]

61
L_M \left( x_i^* \neq x_{i-1}^* \right) = \frac{[1 - \beta (1 - \lambda)] \psi^2}{1 - \beta (1 - \lambda) \psi^2} \frac{\hat{\sigma}_L^2}{1 - \beta} + \frac{(1 - \psi)^2}{1 - \beta (1 - \lambda) \psi^2} \frac{\hat{\sigma}_J^2}{1 - \beta} \quad \text{whence}

This system is better than pure statute law if and only if

\frac{1 - \psi}{1 + \psi} \hat{\sigma}_J^2 \leq \hat{\sigma}_L^2

and it is better than pure case law if and only if

\frac{\psi^2}{1 - \beta \psi^2} (x_0 - x_1^*)^2 + \frac{\beta \lambda \psi^2 (1 - \psi)^2}{[1 - \beta (1 - \lambda) \psi^2] (1 - \beta \psi^2)} \frac{\hat{\sigma}_J^2}{1 - \beta} + \frac{\beta \psi (2 - \psi - \beta \psi^2)}{(1 - \beta \psi) (1 - \beta \psi^2)} \frac{\lambda \sigma_{\xi}^2}{1 - \beta (1 - \lambda) \psi^2} \frac{\hat{\sigma}_L^2}{1 - \beta} \geq \frac{[1 - \beta (1 - \lambda)] \psi^2}{1 - \beta (1 - \lambda) \psi^2} \frac{\hat{\sigma}_L^2}{1 - \beta}

and therefore with certainty if and only if

\sigma_{\xi}^2 \geq \frac{(1 - \beta \psi) \left( 1 - \beta \psi^2 \right) \psi}{\beta \left( 2 - \psi - \beta \psi^2 \right) \left( 1 - \beta \psi^2 + \beta \psi^2 \lambda \right) \lambda} \left[ (1 - \beta + \beta \lambda) \hat{\sigma}_L^2 - \frac{\beta (1 - \psi)^2 \lambda \sigma_{\xi}^2}{1 - \beta \psi^2} \hat{\sigma}_J^2 \right]