2014

Analogical Legal Reasoning: Theory and Evidence

Joshua C. Teitelbaum
Georgetown University Law Center, jct48@law.georgetown.edu

Analogical Legal Reasoning: Theory and Evidence

Joshua C. Teitelbaum[*]

Georgetown University

This version: March 10, 2014

Prior version: June 2013

Abstract

The paper offers a formal model of analogical legal reasoning and takes the model to data. Under the model, the outcome of a new case is a weighted average of the outcomes of prior cases. The weights capture precedential influence and depend on fact similarity (distance in fact space) and precedential authority (position in the judicial hierarchy). The empirical analysis suggests that the model is a plausible model for the time series of U.S. maritime salvage cases. Moreover, the results evince that prior cases decided by inferior courts have less influence than prior cases decided by superior courts.

1   Introduction

How do judges reason about the law? There are many theories. The canonical theory is that judges reason by analogy from case to case (Levi 1949; Weinreb 2005). This method of reasoning is known as analogical legal reasoning to jurisprudence scholars and case-based legal reasoning to scholars in the field of artificial intelligence and law.[1]

In its purest form, analogical legal reasoning (ALR) involves reasoning directly from prior cases to the case at hand—the judge evaluates the similarities and differences between prior cases and the case at hand and reaches a decision through application of the principle that like cases should be treated alike (Alexander and Sherwin 2008). Notably, ALR operates without invoking a legal rule that governs the decision in the case at hand (Sunstein 1993, 1996).[2] In this way, ALR stands in contrast to rule-based legal reasoning (RLR), which involves reasoning deductively from legal rules (Alexander and Sherwin 2008; Schauer 2009).[3] In its purest form, RLR operates without reference to prior cases—the judge simply applies the governing legal rule to the case at hand.[4]

ALR constitutes a "legalist" theory of judicial behavior (Posner 2008). According to the legalist theory, "judges decide cases through systematic application of the external, objective sources of authority that classically comprise the law" (Cross 2003).[5] Although the legalist theory is the traditional theory of judicial behavior in legal circles, it has many critics. Perhaps the leading criticism of the legalist theory is that it suffers from theoretical and empirical indeterminacy (Cross 2003). ALR has been especially targeted by critics, with one commentator complaining that "it is infrequently described with any rigor or care" (Alexander 1996).[6]

---

[1] Some commentators argue that its use of analogy makes legal reasoning a distinctive form of reasoning (e.g., Fried 1981; Weinreb 2005). The mystical notion that legal reasoning is a distinctive form of reasoning was famously articulated by Sir Edward Coke, the Chief Justice of England, who denied the authority and competence of the King of England to render legal judgments on the grounds that legal questions "are not to be decided by natural reason but by the artificial reason and judgment of law." *Prohibitions Del Roy*, 77 Eng. Rep. 1342 (1607).

[2] On the different forms of analogical legal reasoning, see generally Macagno and Walton (2009).

[3] See also Westen (1982), Eisenberg (1988), Posner (1990, 1995, 2006, 2008), Schauer (1991), and Alexander (1996, 1998).

[4] At most, the judge uses prior cases to infer (perhaps abductively or inductively) the governing legal rule. However, she does not reason directly from case to case.

[5] Of course, there are many other theories of judicial behavior. Posner (2008) identifies no fewer than nine theories, including most notably the legalist theory, the attitudinal theory, which posits that judges decide cases according to their ideological preferences (e.g., Segal and Spaeth 1993, 2002), and the economic and strategic theories, which posit that judges decide cases strategically, taking into account the responses of other actors, to promote their ideology (e.g., Epstein and Knight 1998; Smith and Tiller 2002), enhance their reputation or career prospects (e.g., Miceli and Coşgel 1994; Levy 2005), or further some other specified objective.

[6] Notable exceptions include Sunstein (1993, 1996), Brewer (1996), and Weinreb (2005). Alexander (1996) also asserts that ALR is a "fantasy."

This paper has two objectives. The first objective is to offer a formal model of ALR. The model posits that the outcome in the case at hand is a weighted average of the outcomes of prior cases. The weight placed on the outcome of a prior case in the weighted average captures the *precedential influence* of the prior case and depends on the *fact similarity* (distance in fact space) and *precedential authority* (position in the judicial hierarchy) of the prior case relative to the case at hand.

The ALR model is closely related to the empirical similarity model of Gilboa et al. (2006) and Billot et al. (2005), as well as the wider literature on case-based decision theory.[7] Case-based decision theory is a model of reasoning by analogy to past cases (Gilboa and Schmeidler 2001).[8] Empirical similarity theory is a closely-related model for real-valued assessment problems. Under the empirical similarity model, new assessments are made according to similarity-weighted averages of prior assessments.[9] In most applications of empirical similarity theory, the similarity function (i.e., the function that determines the weights in the weighted average) is symmetric—the influence of a prior case on a new case is the same as the (counterfactual) influence of the new case on the prior case. This is because the similarity function typically is based on a metric, usually a weighted Euclidean metric (e.g., Gayer et al. 2007; Lieberman 2010; Gilboa et al. 2011). In the ALR model, by contrast, the similarity function is asymmetric—the influence of a prior case on a new case is not necessarily the same as the (counterfactual) influence of the new case on the prior case. This is because the similarity function is based on a quasimetric, i.e., a function that satisfies the properties of a metric, apart from symmetry. This allows the ALR model to capture an important feature of reasoning by analogy in law, namely that precedential influence depends not only on fact similarity, which is symmetric, but also on precedential authority, which is not symmetric.[10]

---

[7] In case-based decision theory, the term "case" is used generically; it does not refer to a legal case.

[8] See also Gilboa and Schmeidler (1995, 1996, 1997, 2000, 2002, 2003) and Gilboa et al. (2002). Case-based decision theory was inspired by work on case-based reasoning in artificial intelligence (Riesbeck and Schank 1989) and harkens back to the notion that all human "reasonings concerning matter of fact are founded on a species of Analogy" (Hume 1748).

[9] Empirical similarity theory is related to various methods in computer science, statistics, and related fields, including, most notably: kernel methods (Pagan and Ullah 1999), which are commonly used in nonparametric estimation; nearest neighbor methods (Dasarathy 1991; Devroye et al. 1996), which are commonly used in machine learning and pattern recognition; and conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models (Banerjee et al. 2004), which are commonly used in the analysis of areal and other spatial data. I expand upon the connection between empirical similarity theory and kernel regression in Section 3.2. For discussions of the relationship between empirical similarity theory, on the one hand, and nearest neighbor methods and CAR models, on the other hand, see Lieberman (2010) and Gilboa et al. (2011).

[10] Lieberman (2012) and Argenziano and Gilboa (2012) also feature asymmetric similarity functions, the former in a model of similarity-based autoregression and the latter in a model of history-dependent belief formation in coordination games. In neither paper, however, is the similarity function asymmetric because it depends on the authority of the prior case. In Lieberman (2012), the similarity function is asymmetric because it depends on the direction of the change of the characteristics of a variable from

The second objective of the paper is to take the ALR model to data. The data comprise the time series of reported decisions by federal courts in U.S. maritime salvage cases.[11] The first step of the empirical analysis is to embed the ALR model in a statistical model. The next step is to assess whether the ALR model is a plausible model for the data. I focus on two properties of the ALR model: (i) it is an autoregressive process and (ii) the process has a unit root. To investigate whether an autoregressive process could have generated the data, I estimate a linear regression model and test for autocorrelation in the residuals using parametric and nonparametric tests. To investigate whether the data has a unit root, I employ a "nearly efficient" unit root test. The results suggest that the data are consistent with an autoregressive process that has a unit root.

The final step of the empirical analysis is to estimate the ALR model by maximum likelihood and test the null hypothesis that the similarity function is symmetric. This is the crucial step of the empirical analysis, as the key innovation of the ALR model, and the paper's main contribution, is the asymmetry of the similarity function. I find that the symmetry hypothesis is rejected at the one percent level. The implication is that precedential authority, and not just fact similarity, matters for precedential influence. All else equal (namely, fact similarity), the precedential influence of a prior case that was decided by a inferior court is significantly less than the precedential influence of a prior case that was decided by a superior court.

The remainder of the paper is organized as follows. Section 2 presents the ALR model. Section 3 contains the empirical analysis. In Section 4, I discuss certain limitations of the ALR model and the empirical analysis, as well as directions for future research.

2   The ALR Model

2.1   Legal System

Let $\mathcal{K}$ denote the set of judges or *courts* in the legal system. The courts in $\mathcal{K}$ are ordered in accordance with the hierarchy of courts in the legal system. Accordingly, I sometimes refer to $\mathcal{K}$ as the *authority space*.

Let $\mathcal{Q}$ denote the set of *questions* of law that may be presented to a court. For each question $q \in \mathcal{Q}$, there exists a set of *conclusions* of law $\mathcal{Y}$ that a court may reach with respect to question q. There also exists an array of *issues* of fact that the court must resolve in order to reach a conclusion with respect to question q. For each issue i,

one time to the next. In Argenziano and Gilboa (2012), the similarity function is asymmetric because it depends on the outcome of the prior game.

[11]In Section 3.1, I describe the data and explain why I selected U.S. maritime salvage cases for the empirical analysis.

there exists a set of *findings* of fact Φ_i that the court may make with respect thereto. Accordingly, each question $q \in \mathcal{Q}$ induces a *fact space* $\Phi = \Phi_1 \times \cdots \times \Phi_n$. Each element $\phi = (\phi_1, ..., \phi_n) \in \Phi$ is a *fact pattern*. Given question q, the set of conclusions $\mathcal{Y}$ and the fact space Φ are known and unique.

For example, consider the question of patentability under U.S. patent law. The authority space $\mathcal{K}$ comprises the U.S. district courts (at the bottom of the hierarchy), the U.S. courts of appeals (in the middle), and the U.S. Supreme Court (at the top). The question of law q is whether an invention is patentable. The set of conclusions is $Y = \{0, 1\}$, where zero represents no and one represents yes. The issues of fact are whether the invention is (1) a patentable subject matter (i.e., a process, machine, article of manufacture, or composition of matter, or any improvement thereof), (2) novel, (3) nonobvious, and (4) useful. For each issue i, the set of findings is $\Phi_i = \{0, 1\}$, where again zero represents no and one represents yes. Accordingly, the fact space is $\Phi = \{0, 1\} \times \{0, 1\} \times \{0, 1\} \times \{0, 1\}$, and one example fact pattern is $\phi = (1, 1, 0, 1)$.

A *case* involving question q is a triple $c = (\phi, \kappa, y)$, where $\phi \in \Phi$, $\kappa \in \mathcal{K}$, and $y \in \mathcal{Y}$. Define $x = (\phi, \kappa)$ as the *inputs* of the case and y as the *outcome* of the case. The set of all possible cases involving question q is $\mathcal{C} = (\Phi \times \mathcal{K}) \times \mathcal{Y}$. I assume throughout the paper that the inputs and outcomes of cases are or may be represented by real variables: $\Phi \subseteq \mathbb{R}^n$, $\mathcal{K} \subseteq \mathbb{R}$, and $\mathcal{Y} \subseteq \mathbb{R}$.

At time t, a court is presented with question q and a body of evidence. Based on the evidence, the court makes findings of fact $\phi_t \in \Phi$. The court has access to a q-relevant *case history* $C_t = (c_1, ..., c_{t-1})$, where $c_j = (x_j, y_j) \in \mathcal{C}$ is a prior case involving question q. How the court reaches its conclusion y_t depends on its method of legal reasoning. Under ALR, the outcome of the case at hand is a function of the inputs of the case at hand as well as the history of prior cases, $y_t = Y(x_t, C_t)$.[12]

2.2 ALR Model

I model ALR as similarity-weighted averaging of prior outcomes. Formally:

$$y_t = Y(x_t, C_t) = \sum_{j<t} \left(\frac{s(x_j, x_t)}{\sum_{j<t} s(x_j, x_t)} \right) y_j, \tag{1}$$

where $s : \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}_{++}$ is a function that indexes the similarity between the inputs x_j of the prior case and the inputs x_t of the case at hand. Equation (1) posits that the outcome y_t in the case at hand is a weighted average of the outcomes $y_1, ..., y_{t-1}$ of prior

---

[12]Contrast this with RLR, under which the outcome of the case at hand is a function of the inputs only, $y_t = Y(x_t)$.

cases. The weight placed on the outcome y_j of a prior case depends on the degree to which the inputs x_j of the prior case are similar to the inputs x_t of the case at hand. The degree of input similarity is given by s. The greater is the input similarity s of a prior case, the greater is the weight given to the outcome y_j of the prior case in the determination of the outcome y_t of the case at hand. Hence, I interpret s as measuring the *precedential influence* of a prior case on the case at hand.

I assume that input similarity s—and, therefore, precedential influence—is an exponentially decaying function of the distance μ from the inputs x_j of the prior case to the inputs x_t of the case at hand:

$$s(x_j, x_t) = \exp(-\mu(x_j, x_t)), \tag{2}$$

where $\mu : \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}_+$ with $\mu(x_j, x_t) = 0$ only if $x_j = x_t$. The assumption that influence decays exponentially with distance seems natural and appears in other contexts. For instance, Shepard (1987) derives a law of psychological generalization in which the probability of generalizing a response from one stimulus to another decays exponentially with the distance between the stimuli in psychological space. White (2001) argues that both theory and evidence support exponential decay of memory with distance in time (scaled as $\sqrt{t}$). And Glaeser et al. (2003) introduce a model of socially influenced behavior in which social influence decays exponentially with social distance. What's more, Billot et al. (2008) provide an axiomatic justification for specifying an exponential similarity function in the present context (similarity-weighting averaging as a model of reasoning).[13]

Next, I assume that input distance μ is a proportional function, with proportionality factor v, of the weighted Euclidean distance d between the facts ϕ_j of the prior case and the facts ϕ_t of the case at hand:

$$\mu(x_j, x_t) = v(x_j, x_t)d(\phi_j, \phi_t), \tag{3}$$

where $v : \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}_{++}$ captures the *precedential authority* of the prior case relative to the case at hand and $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$, which captures the *fact similarity* of the prior case relative to the case at hand, is given by

$$d(\phi_j, \phi_t) = \sqrt{\sum_{i=1}^{n} \omega_i(\phi_{ji} - \phi_{ti})^2}, \quad \omega_i > 0 \text{ for all } i. \tag{4}$$

---

[13] More specifically, Billot et al. (2008) provided an axiomatization of an exponential similarity function based on a metric induced by norm. The key axiom for the exponential form is ray shift invarance, which requires that if the facts of all prior cases lie along a ray emanating from the facts of the case at hand, then an equal shift along this ray of the facts of all prior cases does not change the outcome of the case at hand.

Note that the weights $\omega_1, ..., \omega_n$ in the weighted Euclidean distance d reflect the relative importance of the n issues of fact that the court must resolve in order to reach a conclusion with respect to the legal question at issue.

In specifying v, I am guided by three criteria. First, all else equal, prior cases decided by inferior courts should have less influence than prior cases decided by superior courts. At the same time, prior cases decided by coequal courts should have no less influence than prior cases decided by inferior courts and no more influence than prior cases decided by superior courts. Second, the influence penalty (resp. bonus) for prior cases decided by inferior (resp. superior) courts should diminish as the distance in fact space to the case at hand becomes large. This is motivated by the notion that the lesser is the factual similarity of a prior case (i.e., the greater is d), the less important is relative position of the deciding court in the judicial hierarchy. Third, the input distance $\mu = vd$ should satisfy all the properties of a metric, apart from symmetry. As stated previously, the key innovation of the ALR model, and the paper's main contribution, is the asymmetry of the similarity function. This is accomplished by relaxing the symmetry of the distance measure on which the similarity function is based. However, it is neither necessary nor proper to relax any of the other properties. Hence, the specification of v should preserve these other properties, including the triangle inequality.[14]

Guided by these criteria, I assume that v is given by

$$v(x_j, x_t) = \sec\theta_{jt} + \tan\theta_{jt}, \tag{5}$$

where

$$\theta_{jt} = \arctan\left(\frac{\beta(\kappa_t - \kappa_j)}{d(\phi_j, \phi_t)}\right), \quad \beta \geq 0.$$

Below I show that with v specified by equation (5), the ALR model satisfies the first and second criteria set forth above. Moreover, in the Appendix I prove that the third criterion is satisfied as well.

Before proceeding, however, I want to supplement the instrumental motivation for v with a geometric motivation. To draw an analogy between a prior case and the case at hand, a court has to traverse (metaphorically) the distance in input space from the prior case to the case at hand. The greater is this distance, the more strained is the analogy.

---

[14]The third criterion rules out many otherwise desirable specifications of v. For instance, a simple way to impose asymmetry would be to define v as follows:

$$v(x_j, x_t) = \begin{cases} 1 & \text{if } \kappa_j > \kappa_t \\ \beta & \text{if } \kappa_j \leq \kappa_t \end{cases}, \quad \beta \geq 1$$

(cf. Lieberman 2012; Argenziano and Gilboa 2012). Under this specification of v, however, $\mu = vd$ generally does not satisfy the triangle inequality on $\mathbb{R}^{n+1}$ $(n > 0)$.

In input space (or fact-authority space), θ_{jt} is direction from the prior case (the origin) to the case at hand, where the distance in fact space is given by $d(\phi_j, \phi_t)$, the distance in authority space is given by $\beta(\kappa_t - \kappa_j)$, and the radial distance is given by

$$f(x_j, x_t) = \sqrt{(d(\phi_j, \phi_t))^2 + (\beta(\kappa_t - \kappa_j))^2}$$

(see Figure 1). Note that $\sec\theta_{jt} = f(x_j, x_t)/d(\phi_j, \phi_t)$ and $\tan\theta_{jt} = \beta(\kappa_t - \kappa_j)/d(\phi_j, \phi_t)$. It follows that

$$\mu(x_j, x_t) = v(x_j, x_t)d(\phi_j, \phi_t) = f(x_j, x_t) + \beta(\kappa_t - \kappa_j).$$

That is, input distance equals radial distance plus an adjustment that takes into account both the direction and distance in authority space. The adjustment is positive if the prior case has inferior authority ($\kappa_j < \kappa_t$) and negative if the prior case has superior authority ($\kappa_j > \kappa_t$). It is harder to travel uphill than downhill. If the prior case has coequal authority ($\kappa_j = \kappa_t$), there is no adjustment and $\mu = d$. The court only has to overcome any factual dissimilarity to draw the analogy.[15]

Figure 2 provides a visualization of v. Observe that $v = 1$ if $\beta = 0$. Thus, $\beta = 0$ implies that precedential influence is symmetric, depending only on fact similarity and not on precedential authority. However, if $\beta > 0$, then $v > 1$ if the prior case has inferior authority ($\kappa_j < \kappa_t$), $v = 1$ if the prior case has coequal authority ($\kappa_j = \kappa_t$), and $v < 1$ if the prior case has superior authority ($\kappa_j > \kappa_t$). Hence, $\beta > 0$ begets influence penalties and bonuses for prior cases with inferior and superior authority, respectively, and thus implies that precedential influence is asymmetric.

Figure 3 displays the relationship in the ALR model between precedential influence (s), fact similarity (d), and precedential authority (v).[16] The precedential influence of a prior case is greatest when the facts of the prior case are identical to the facts of the case at hand ($\phi_j = \phi_t \Leftrightarrow d = 0$), and it decays exponentially at rate v as fact similarity decreases (i.e., as d increases). Both the precedential influence at $d = 0$ and the rate of decay for $d > 0$ differ depending on the precedential authority of the prior case. If the prior case was decided by a superior court ($\kappa_j > \kappa_t$), the precedential influence at $d = 0$ is the highest possible ($s = 1$) and the rate of decay for $d > 0$ is the lowest possible ($v < 1$).

---

[15]In Teitelbaum (2013), I provide an axiomatic motivation for v. In brief, I show that the input distance $\mu = vd$ is a quasimetric induced by a skewed norm. A skewed norm is a positive definite function $F(N, p)(x) = N(x) - \langle p, x \rangle$, where N is a norm on $\mathbb{R}^n$, $p \in \mathbb{R}^n$, and $\langle \cdot, \cdot \rangle$ denotes the scalar product (Plastria 1992). I then provide a "skewness" axiom that, when imposed in lieu of the symmetry axiom in the main result of Billot et al. (2008), characterizes an exponential similarity function based on a skewed norm. The skewness axiom essentially postulates exponential discounting of the influence of prior cases with inferior authority relative to equidistant prior cases with superior authority.

[16]Note that fact similarity and precedential authority are negatively related to d and v, respectively.

If the prior case was decided by a coequal court ($\kappa_j = \kappa_t$), the precedential influence at $d = 0$ is equally high ($s = 1$) but the rate of decay for $d > 0$ is higher ($v = 1$). If the prior case was decided by an inferior court ($\kappa_j < \kappa_t$), the precedential influence at $d = 0$ is lower ($s < 1$) and the rate of decay for $d > 0$ is even higher ($v > 1$). All else equal (namely, d), therefore, the influence of prior cases with inferior authority is less than the influence of prior cases with superior authority, and the influence of prior cases with coequal authority is greater than the influence of prior cases with inferior authority and not less than the influence of prior cases with superior authority. Moreover, the size of the influence penalty (resp. bonus) for prior cases with inferior (resp. superior) authority increases with the degree of fact similarity (i.e., as d increases) at a rate determined by (and positively related to) the parameter β.

Importantly, Figure 3 illustrates that with v specified by equation (5), the ALR model satisfies the first and second criteria set forth above. In the Appendix, I prove that the third criterion is satisfied as well. That is, I prove that $\mu = vd$ is a quasimetric, i.e., a function that satisfies the properties of a metric, apart from symmetry.

3   Empirical Analysis

In this section, I take the ALR model to data. After describing the data, I embed the ALR model in a statistical model and assess whether it is a plausible model for the data. I then estimate the ALR model by maximum likelihood and test the null hypothesis that the similarity function is symmetric (i.e., $\beta = 0$) against the alternative hypothesis that it is asymmetric (i.e., $\beta > 0$).

3.1   Data

The data comprise the time series of reported decisions by federal courts in U.S. maritime salvage cases. Under U.S. maritime law, a salvor of imperiled maritime property on navigable waters is entitled to a monetary award from the owner.[17] There are two forms of maritime salvage: "contract" salvage and "pure" salvage. Contract salvage is rendered pursuant to a prior agreement. Pure salvage is rendered voluntarily in the absence of a contract. The data include only pure salvage cases.

In the United States, the federal courts have exclusive jurisdiction in pure salvage cases. There are three elements of a valid pure salvage claim: (i) a marine peril; (ii) service voluntarily rendered; and (iii) success in whole or in part. In the case of a valid pure salvage claim, the court determines the award according to six factors enumerated

---

[17] The following is a bare bones description of U.S. maritime salvage law. For more detailed overviews, see, e.g., Force (2004) or Schoenbaum (2011, ch. 16).

by the Supreme Court in *The Blackwall*, 77 U.S. 1 (1869): (1) the labor expended by the salvors in rendering the salvage service; (2) the promptitude, skill, and energy displayed in rendering the service and saving the property; (3) the value of the property employed by the salvors in rendering the service, and the danger to which such property was exposed; (4) the risk incurred by the salvors in securing the property from the impending peril; (5) the value of the property saved; and (6) the degree of danger from which the property was rescued. The law of salvage provides no precise formula or rule for computing a salvage award on the basis of the *Blackwall* factors. The court has broad discretion in fashioning the award based on its findings with respect to the *Blackwall* factors, though it is bound to apply all of the *Blackwall* factors and the award generally may not exceed the value of the property saved.

There are several reasons why I selected maritime salvage cases for the empirical analysis. First, the outcome (the salvage award) is a continuous variable (a dollar amount) and the inputs (the *Blackwall* factors) are well defined and stable over time.[18] Second, awards in maritime salvage cases arguably are apolitical legal questions. Moreover, it is hard to imagine that a maritime salvage case is an opportunity for a federal judge to advance strategic goals such as career advancement. Thus, if there is any setting in which we should expect a legalist model of judicial behavior to be operative (and other models such as attitudinal or strategic models to be inoperative), it is maritime salvage cases. Third, the law of maritime salvage is federal law, and federal courts have exclusive jurisdiction in cases involving claims for salvage awards. Accordingly, state variation in law or courts is not an issue. Fourth, it seems reasonable to treat the federal courts as a single adjudicative body for purposes of maritime salvage cases: there is no split among the circuits (*The Blackwall* is controlling precedent for all circuits); there are no specialty courts for maritime cases; and it generally is believed that federal courts are reasonably uniform in quality. Lastly, although the criteria for determining a maritime salvage award are well defined and stable through time, the law provides no precise formula or rule. This leaves open the possibility that courts are engaging in ALR.

The data comprise 684 pure salvage cases from 1799 to 2007. The cases were identified using seven search methods. The first search method was "KeyCiting" and "Shepardizing" *The Blackwall* in Westlaw and LexisNexis, respectively. The second search method was performing keyword searches in three databases: Westlaw's Federal Maritime Law – Cases (FMRT-CS); Lexis' Admiralty Cases, Federal and State (MEGA); and American Maritime Cases (AMC), which is available on Westlaw and Lexis. The third search

---

[18] In the words of the U.S. Court of Appeals for the Ninth Circuit, the *Blackwall* factors "have weathered the storms of the past century." *Saint Paul Marine Transp. Corp. v. Cerro Sales Corp.*, 505 F.2d 1115 (9th Cir. 1974).

method was consulting the salvage digests in West's federal digest. The fourth search method was consulting the salvage award tables in the quinquennial digests of American Maritime Cases. The fifth search method was consulting leading treatises on admiralty and maritime law. The sixth search method was consulting early American digests and reporters. The final search method was examining each case, however identified, for two purposes: (i) find additional cases cited therewithin and (ii) cull cases that, upon closer inspection, did not apply the *Blackwall* factors to determine an award for pure salvage.[19]

For each case, the data record the date of the decision, the court,[20] the award (in 1980 U.S. dollars), and the court's finding on each *Blackwall* factor. The position of the court in the judicial hierarchy is coded as follows: district court = 0, circuit court = 1, and Supreme Court = 2. The *Blackwall* factors (other than the value of the property saved) are coded as binary variables: low = 0 or high = 1. This is for two reasons. The first reason is that courts routinely characterize salvage operations as "high order" or "low order." The second reason is that binary coding minimizes subjectivity and, therefore, disagreement/error. The value of the property saved is recorded in 1980 U.S. dollars.[21]

Of the 684 cases, 545 cases (79.7 percent) were decided by district courts, 134 cases (19.6 percent) were decided by circuit courts, and five cases (0.7 percent) were decided by the Supreme Court. Table 1 displays summary statistics for the award and the six *Blackwall* factors. For instance, it shows that the awards range from $240 to $1,866,000, with a mean award of $74,000; the value of the property saved ranges from $1,200 to $42,133,000, with a mean value of $1,386,000; and the labor expended by the salvors was high in 39 percent of the cases.[22] Table 2 displays the mean award conditional on different findings on the *Blackwall* factors. For instance, it shows that the mean award for cases in which the labor expended by the salvors was low is $36,000, whereas the mean award for cases in which the labor expended by the salvors was high is $134,000.

---

[19] These searches yielded 881 pure salvage cases from 1779 to 2007. Of these cases, 197 were excluded from the final data set because they were missing either a clear statement of the salvage award or clear findings with respect to one or more of the *Blackwall* factors.

[20] The court is the court of final adjudication, and the data record the award and findings of fact as determined by the court of final adjudication.

[21] The following procedures were followed in coding the cases. After receiving instructions from me, a research assistant read every case and hand coded every variable. In addition, I met regularly with the research assistant to review his progress and discuss any coding issues or questions. Finally, I audited his work by independently reading and shadow coding 15 percent of the cases. Our disagreement rate was zero with respect to the awards and less than one percent with respect to the *Blackwall* factors.

[22] Though not shown in the table, it is worth noting that the award percentage (the salvage award expressed as a fraction of the value of the property saved) ranges from less than one percent to 85 percent, with a mean of 14 percent.

3.2 Statistical ALR Model

The first step of the empirical analysis is to embed the ALR model in a statistical model. Following Gilboa et al. (2006) and its progeny,[23] I assume $y_1 = \varepsilon_1$ and

$$y_t = \sum_{j<t} \left(\frac{s(x_j, x_t; \Omega)}{\sum_{j<t} s(x_j, x_t; \Omega)} \right) y_j + \varepsilon_t, \quad t = 2, ..., T, \tag{6}$$

where $\varepsilon_t \overset{iid}{\sim} N\left(0, \sigma^2\right)$ for $t = 1, ..., T$. (Note that I include $\Omega \equiv (\omega_1, ..., \omega_n)$ in (6) to make explicit the dependence of s on the weights $\omega_1, ..., \omega_n$ in the weighted Euclidean distance d on which s is based.) Model (6) posits that the outcome y_t of the case at hand is normally distributed around a similarity-weighted average of the outcomes $y_1, ..., y_{t-1}$ of prior cases. To highlight two key properties of the model, rewrite (6) as

$$y_t = (\alpha_{1,t})y_{t-1} + (\alpha_{2,t})y_{t-2} + \cdots + (\alpha_{t-1,t})y_1 + \varepsilon_t, \tag{7}$$

where $\alpha_{i,t} = s\left(x_{t-i}, x_t\right) / \sum_{i<t} s\left(x_i, x_t\right)$. From (7), we can see that (i) model (6) is an autoregressive process of order $(t-1)$ and (ii) because the $\alpha_{i,t}$'s sum to one for each t, the process has a unit root (Lieberman 2010).

Before proceeding to the next step of the empirical analysis, let me say a few words about the relationship between model (6) and kernel regression.[24] Kernel regression assumes a data generating process of the form

$$y_t = g(x_t) + \epsilon_t, \quad t = 1, ..., T, \tag{8}$$

where $\epsilon_i \overset{iid}{\sim} \left(0, \sigma^2\right)$ and g is an unknown function. A standard estimator for g is the Nadaraya-Watson estimator

$$\hat{g}(x_t) = \sum_{j=1}^{T} \left(\frac{K(x_t - x_j; H)}{\sum_{j=1}^{T} K\left(x_t - x_j; H\right)} \right) y_i,$$

where K is a kernel function and H is a diagonal matrix of bandwidth parameters $h_1, ..., h_n$. Note the connection between (6) and (8). Each generates a new/predicted value of y by taking a weighted average of the observed values of y where the weights are a function the distance between the new/hypothesized x and the observed values of x. Notwithstanding this connection, however, there is an important distinction between

---

[23] See, e.g., Gayer et al. (2007), Lieberman (2010), and Gilboa et al. (2011).
[24] On kernel regression, see, e.g., Pagan and Ullah (1999). My comments below echo the comments by Gilboa et al. (2011) on the relationship between the empirical similarity model and kernel regression. For more on this relationship, see Gilboa et al. (2006, 2011) and Lieberman (2010).

(6) and kernel regression. Kernel regression is a statistical technique that uses weighted averaging to estimate (8), which assumes that the data are generated by an unknown function g, whereas (6) assumes that that the data are generated by weighted averaging. In other words, (8) specifies a *rule* relating x_t to y_t, and thus assumes that the distribution of y_t depends only on x_t, whereas (6) assumes that the distribution of y_t depends not only on x_t but also on the history of prior cases, $C_t = \{(x_j, y_j) : j < t\}$.

3.3   Plausibility of ALR Model

The next step of the empirical analysis is to assess whether the ALR model is a plausible model for the data. I focus on the two key properties of the model highlighted above: (i) model (6) is an autoregressive process and (ii) the process has a unit root.

To investigate whether an autoregressive process could have generated the data, I estimate a linear regression model and test for autocorrelation in the residuals. Because salvage awards are bounded below by zero and skewed to the right,[25] the dependent variable is the log-transformed award. The regressors comprise a constant and the six *Blackwall* factors, where the value of the property saved is log-transformed.[26] The model is estimated by ordinary least squares.

I test for autocorrelation in the residuals using the Ljung-Box test and the Runs test. The Ljung-Box test is a portmanteau test which considers the null hypothesis that the residuals are not autocorrelated against the alternative hypothesis of an autoregressive process of order p.[27] The test statistic is $Q = T(T+2) \sum_{j=1}^{p} \rho_j^2/(T-j)$, where ρ_j the jth autocorrelation coefficient of the residual series. Under the null hypothesis, Q is asymptotically distributed $\chi^2(p)$. The Runs test is a nonparametric test that considers the null hypothesis that the residual process is random against the alternative of a nonrandom process.[28] The test counts the number of runs r above and below zero and compares it to the expected number of runs $\bar{r}$. Under the null hypothesis, $\bar{r} = (2n_0 n_1/T) + 1$, where n_0 and n_1 are the number of values above and below zero, respectively. The test statistic is $Z = (r - \bar{r})/s_r$, where $s_r = \sqrt{2n_0 n_1 (2n_0 n_1 - T)/T^2(T-1)}$. Under the null hypothesis, Z is approximately distributed $N(0, 1)$.

The results of both tests are reported in Table 3. For the Ljung-Box test, results are presented for six and twenty lags, which correspond to the values suggested by Box et al.

---

[25] More than 72 percent of the awards are less than the mean award ($74,000).

[26] The log-transformation of the value of the property saved is not an arbitrarily imposed assumption; rather, it is the specification selected by the multivariable fractional polynomial procedure of Sauerbrei and Royston (1999). For details, see the Appendix.

[27] The Ljung-Box test was proposed by Ljung and Box (1978). For a textbook treatment, see, e.g., Johnston and DiNardo (1997).

[28] For a textbook treatment of the Runs test, see, e.g., Bradley (1968, ch. 11).

(1994) (who suggest $\min\{20, T-1\}$) and Tsay (2005) (who suggests $\ln T$), respectively, and also for forty lags. In each test, the null hypothesis is rejected at the five percent level, suggesting that the data are consistent with an autoregressive process.

To investigate whether the data generating process has a unit root, I employ the ADF-GLS test proposed by Elliott et al. (1996). The ADF-GLS test is an augmented Dickey-Fuller (ADF) test in which the time series is detrended via generalized least squares (GLS).[29] The ADF-GLS test is "nearly efficient" in the sense that its asymptotic local power functions are virtually indistinguishable from the Gaussian power envelope, and has greater power than the standard ADF test (which is not "nearly efficient") (Haldrup and Jansson 2006). The test proceeds in two steps. First, it estimates γ in the regression $y_t - \overline{\rho} y_{t-1} = (1-\overline{\rho})\gamma + v_t$, where $\overline{\rho} = 1 - 7/T$. Second, it estimates δ in the regression $\Delta \widetilde{y}_t = \delta \widetilde{y}_{t-1} + \sum_{j=1}^{p} \eta_j \Delta \widetilde{y}_{t-j} + u_t$, where $\Delta \widetilde{y}_t = \widetilde{y}_t - \widetilde{y}_{t-1}$, $\widetilde{y}_t = y_t - \widehat{\gamma}$, and $\widehat{\gamma}$ is the estimate of γ obtained in the first step. The test statistic is the ordinary t statistic for $\delta = 0$. The results of ADF-GLS test are reported in Table 3. Results are presented for forty lags, which corresponds to the value suggested by both the sequential t criterion proposed by Ng and Perron (1995) and the modified Akaike information criterion (MAIC) proposed by Ng and Perron (2000). The test fails to reject the null hypothesis of a unit root at the five percent level, suggesting that the data are consistent with a unit root process.

Of course, the results of the autocorrelation and unit root tests reported in Table 3 provide only indirect, negative assurance that the ALR model is a plausible model for the data. The autocorrelation tests reject the null hypothesis that the residuals from the linear regression model are not autocorrelated, and the unit root test fails to reject the null hypothesis that the data generating process has a unit root. These results are consistent with a data generating process that is autoregressive and has a unit root, but they do not affirmatively establish that the data were generated by such a process, let alone by the ALR model. Indeed, it is important to highlight two limitations of the analysis. First, the Ljung-Box test and the ADF-GLS test contemplate a fixed order autoregressive process,[30] and thus can provide only oblique evidence with respect to the ALR model, which is an autoregressive process of increasing order.[31] Second, although the ADF-GLS test is "nearly efficient" and has greater power than the standard ADF test, it has low power in the present context. To get a sense of the test's power, I performed 1,000 Monte Carlo simulations in which I first generated simulated data under the alternative hypothesis that the data generating process is the linear regression model

---

[29] For a textbook treatment, see, e.g., Davidson and MacKinnon (2004, ch. 14).

[30] More specifically, the Ljung-Box test tests against the alternative hypothesis of a pth order autoregressive process, and the ADF-GLS test tests for a unit root in a pth order autoregressive process.

[31] I am not aware of any autocorrelation or unit root tests that contemplate an increasing order autoregressive process.

estimated above and then tested for a unit root using the ADF-GLS test.[32] At the five percent level, the test correctly rejects the null hypothesis of a unit root 195 times out of 1,000, implying a power of 19.5 percent.

That said, the Runs test is more generally applicable, as it is nonparametric and makes no assumptions about the data generating process. Indeed, not only does the Runs test on the residuals of the linear regression model reject the null hypothesis of randomness, which provides negative assurance on the plausibility of the ALR model, but a Runs test on the residuals of the fitted ALR model (see Section 3.4) fails to reject the null hypothesis of randomness, which provides positive assurance.[33]

3.4   Testing the Symmetry of the Similarity Function

The final step of the empirical analysis is to estimate the model by maximum likelihood and test the hypothesis $\beta = 0$ against the alternative $\beta > 0$, i.e., test the hypothesis that the similarity function s is symmetric against the alternative that s is asymmetric. As stated previously, this is the crucial step, as the key innovation of the ALR model, and the paper's main contribution, is the asymmetry of the similarity function.

The loglikelihood function is

$$l\left(\Theta\right) = -\frac{T}{2}\ln\left(2\pi\right) - \frac{T}{2}\ln\left(\sigma^2\right) - \frac{y'S'Sy}{2\sigma^2},$$

where $\Theta = (\beta, \omega_1, ..., \omega_n, \sigma^2)$ is the vector of model parameters, $y = [y_1 \cdots y_T]'$, and

$$\underset{(T\times T)}{S} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ -\frac{s(x_1,x_3)}{\sum_{j<3}s(x_j,x_3)} & -\frac{s(x_1,x_3)}{\sum_{j<3}s(x_j,x_3)} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{s(x_1,x_T)}{\sum_{j<T}s(x_j,x_T)} & -\frac{s(x_2,x_T)}{\sum_{j<T}s(x_j,x_T)} & \cdots & -\frac{s(x_{T-1},x_T)}{\sum_{j<T}s(x_j,x_T)} & 1 \end{bmatrix}.$$

Recall that β is the shape parameter for the precedential authority function v defined in equation (5), $\omega_1, ..., \omega_n$ are the weights in the Euclidean distance function d defined in equation (4), and σ^2 is the variance of the error term ε_t in equation (6). For the derivation

---

[32]To generate each simulated data set, I performed two steps. First, I sampled 684 residuals from a normal distribution with a mean of zero and a standard deviation set equal to the root mean squared error of the residuals from the linear regression model estimated above. Second, I used the fitted regression line and the simulated residuals to generate 684 simulated awards, one for each observation in the data. To select the number of lags for each ADF-GLS test, I took the greater of (i) the value suggested by the sequestial t criterion and (ii) the value suggested by the MAIC.

[33]The test statistic is -1.543, whereas the five percent critical value is -1.645.

of the loglikelihood function, as well as an explication of the asymptotic theory of model (6), which establishes a theoretical basis for simple hypothesis tests involving the model parameters, see Lieberman (2010).

Table 4 presents the maximum likelihood estimates of the model parameters. Note that in the estimation both the award (y) and the value of the property saved (x_5) are log-transformed. The estimates for $\omega_1, ..., \omega_6$ suggest that each of the *Blackwall* factors, save only the skill displayed by the salvors (x_2), is statistically significant to the determination of the award. They also suggest that the factors which receive the greatest weight are the labor expended by the salvors (x_1), the value of the property saved (x_5), and the danger to the property saved (x_6), and that the factors which receive the least weight are the skill displayed by the salvors (x_2), the danger to the salvors' property (x_3), and the risk incurred by the salvors (x_4). In addition, the estimate for σ^2 suggests that unobserved heterogeneity in salvage awards has unit variance.

Most importantly, the estimate for β is 0.036 with a standard error of 0.017, and the hypothesis $\beta = 0$ is rejected in favor of the alternative $\beta > 0$ at the one percent level. That is, I find that the similarity function is asymmetric, with significant influence penalties and bonuses for cases decided by inferior and superior courts, respectively (see Figure 4). The implication is that precedential authority, and not just fact similarity, matters for precedential influence. All else equal, the precedential influence of a prior case that was decided by an inferior court is significantly less than the precedential influence of a prior case that was decided by a coequal court, which in turn is significantly less than the precedential influence of a prior case that was decided by a superior court.

4   Discussion

The use of analogical reasoning in law is a central topic in the jurisprudence and artificial intelligence and law literatures. Contributing to these literatures, this paper presents a formal model of analogical legal reasoning and takes the model to data. The ALR model posits that the outcome of the case at hand is a weighted average of the outcomes of prior cases, where the weights are a function of the fact similarity and precedential authority of the prior cases. The results of the empirical analysis suggest that the ALR model is a plausible model for the time series of reported decisions by federal courts in U.S. maritime salvage cases. What's more, the results indicate that the similarity function is asymmetric, affirming that precedential influence indeed depends not only on fact similarity, which is symmetric, but also on precedential authority, which is not.

The ALR model and the empirical analysis, however, are subject to several important limitations. First, the ALR model is a stylized representation of analogical legal reasoning

in its purest form—the judge reasons directly from case to case without invoking a governing legal rule. A model that combines elements of analogical and rule-based legal reasoning may be more realistic. In future research, it would be interesting to explore a hybrid ALR-RLR model, perhaps along the lines of a mixed SAR model (Anselin 1988).

Second, the ALR model specifies a particular method of assessment (similarity-weighted averaging of all prior cases), as well as a specific notion of similarity (exponentially decaying function of asymmetric weighted Euclidean distance). Although I would argue that any model of ALR must involve some similarity-weighted statistic of the outcomes of prior cases, statistics other than the mean—e.g., the median or the mode—are plausible alternatives. Another plausible alternative is a similarity-weighted statistic of selected prior cases (as opposed to all prior cases)—e.g., the k-nearest cases.[34] Furthermore, one could specify other similarity functions (e.g., $s = 1/(1+\mu)$),[35] other precedential authority functions,[36] or other distance functions (e.g., $d(a,b) = \sum_{i=1}^{n} \omega_i |a_i - b_i|$).[37]

Third, the ALR model takes a representative agent approach and assumes that all judges are equipped with the same similarity function. Allowing for heterogeneous judges surely would be more realistic. However, tractability would require making strong assumptions about the structure of such heterogeneity.

Fourth, the empirical analysis relies on data that records the inputs and outcomes of legal cases. Such data can provide only indirect evidence regarding the method of legal reasoning.[38] Nevertheless, it arguably is the best available evidence. In many cases, a court's written opinion offers no direct evidence regarding the method of legal reasoning. Even in cases in which the court's opinion offers some direct evidence, it rarely is definitive and, in any event, it arguably is of little probative value.[39]

Lastly, the empirical analysis speaks only to whether the ALR model is a plausible model for the time series of outcomes in U.S. maritime salvage cases. It says nothing

---

[34] A prior study that uses nearest neighbor methods for predicting judicial decisions is Mackaay and Robillard (1974).

[35] This specification of s, which exhibits subexponential decay, was suggested by Gilboa et al. (2006) and used by Gayer et al. (2007) in a study of case-based reasoning about real estate prices.

[36] For instance, Hodgson et al. (1987) and Drezner and Wesolowsky (1989) suggest other quasimetrics from which one could derive alternative specifications for v.

[37] This specification of d is the weighted L_1 distance. The L_1 distance is also known as the Manhattan distance or the the taxicab metric.

[38] Moreover, although the coding scheme was guided by doctrinal considerations and the coding procedures were designed to minimize disagreement/error, one can always quible with the way that legal cases are coded in any study (or even argue that the enterprise of coding legal cases is inherently flawed).

[39] "As a rule, we conceive of the judge's writing of an opinion as a procedure in which he justifies his decision. The writing coincides neither necessarily nor realistically with the process by which he reaches his decision, the process of discovery" (Murray 1982). There are (at least) two reasons to think that a court might use the language of RLR to justify its decision even if it engages in ALR in reaching its decision. First, "the language of 'rules' is much more efficient and parsimonious than that of 'cases'" (Gilboa and Schmeidler 2000). Second, "[r]ules are excellent justification mechanisms" (Hunter 2001).

about whether it is a plausible model for case outcomes in other areas of law.[40] Furthermore, the fact the ALR model is a legalist model of judicial behavior suggests that it may not be well suited to other areas of law, including, in particular, politically charged areas (to which we might expect attitudinal or strategic models to be better suited).[41]

All of that said, I believe that taking a formal modeling approach to ALR helps sharpen ideas not just about ALR but also about RLR, and suggests ways to distinguish them theoretically. For instance, the model of the legal environment in Section 2.1 suggests a way to theoretically distinguish ALR and RLR: under ALR the outcome of the case at hand is a function of the inputs of the case at hand as well as the history of prior cases, $y_t = Y(x_t, C_t)$, whereas under RLR the outcome of the case at hand is a function of the inputs only, $y_t = Y(x_t)$. Stated another way, under RLR the outcome depends on a bounded number of parameters, whereas under ALR the number of parameters increases with the prior case history (cf. Gayer et al. 2007). In addition, the mathematical kinship between the ALR model and kernel regression highlighted in Section 3.2 suggests a theoretical connection between ALR and RLR: under ALR, although the judge "does not explicitly resort to general rules and theories," she "can be viewed as someone who believes in a general rule of the form $Y = f(X^1, ..., X^m)$ but does not know the functional form of f and therefore attempts to estimate it by nonparametric techniques" (Gilboa et al. 2006). It also suggests a theoretical distinction: ALR posits that the data are generated by weighted averaging, whereas RLR posits that the data are generated by a rule and uses weighted averaging as a statistical technique to estimate the rule. In future research, it would be valuable to explore whether these connections and distinctions could be leveraged to develop a way to empirically distinguish ALR and RLR.

---

[40]In future research, it would be interesting to take the ALR model to data on case outcomes in other areas of law. One area in which potentially suitable data already have been collected is U.S. criminal confession cases. See Benesh (2002) and Kastellec (2010).

[41]In future research, it would be interesting to probe the extent to which non-legalist theories of judicial behavior could be formalized using statistical models. For example, I believe one could profitably model an attitudinalist judge as a Bayesian nonparametric statistician.

Appendix

A Proof that μ is a quasimetric

Here I prove that the input distance function $\mu : \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}_+$, on which the similarity function $s : \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}_{++}$ is based, is a quasimetric on $\mathbb{R}^{n+1}$.

Recall the definition of a quasimetric (Wilson 1931):

Definition 1 *A function $\xi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a* quasimetric *on $\mathbb{R}^n$ if for all $x, y \in \mathbb{R}^n$:*

(i) $\xi(x, y) \geq 0$;

(ii) $\xi(x, y) = 0$ *if and only if $x = y$;*

(iii) $\xi(x, z) \leq \xi(x, y) + \xi(y, z)$ *for any $z \in \mathbb{R}^n$ (triangle inequality).*

Note that a metric is a quasimetric that also satisfies symmetry: $\xi(x, y) = \xi(y, x)$. A quasimetric is not necessarily symmetric, i.e., in general $\xi(x, y) \neq \xi(y, x)$.

Theorem 1 *For all $\phi_j, \phi_t \in \mathbb{R}^n$ and $\kappa_j, \kappa_t \in \mathbb{R}$, with $x_j = (\phi_j, \kappa_j)$ and $x_t = (\phi_t, \kappa_t)$, let*

$$\mu(x_j, x_t) = v(x_j, x_t) d(\phi_j, \phi_t),$$

where

$$v(x_j, x_t) = \sec \theta_{jt} + \tan \theta_{jt},$$

$$\theta_{jt} = \arctan \left(\frac{\beta(\kappa_t - \kappa_j)}{d(\phi_j, \phi_t)} \right), \quad \beta \geq 0,$$

and d is a metric on $\mathbb{R}^n$. Then μ is a quasimetric on $\mathbb{R}^{n+1}$.

Proof. Recall from Section 2.2 that

$$\mu(x_j, x_t) = v(x_j, x_t) d(\phi_j, \phi_t) = f(x_j, x_t) + \beta(\kappa_t - \kappa_j),$$

where

$$f(x_j, x_t) = \sqrt{(d(\phi_j, \phi_t))^2 + (\beta(\kappa_t - \kappa_j))^2}.$$

Observe that f is a metric on $\mathbb{R}^{n+1}$. Specifically, it is the weighted Euclidean metric.

(i) Observe that $f(x_j, x_t) \geq |\beta(\kappa_t - \kappa_j)| \geq 0$. It follows that $\mu(x_j, x_t) \geq 0$.

(ii) If $x_j = x_t$ then $f(x_j, x_t) = 0$ and $\beta(\kappa_t - \kappa_j) = 0$, and hence $\mu(x_j, x_t) = 0$. Now suppose $\mu(x_j, x_t) = 0$ but $x_j \neq x_t$. If $x_j \neq x_t$ then $f(x_j, x_t) > 0$ and $\beta(\kappa_t - \kappa_j) \neq 0$. However, because $f(x_j, x_t) \geq |\beta(\kappa_t - \kappa_j)|$, this implies $\mu(x_j, x_t) > 0$, which contradicts $\mu(x_j, x_t) = 0$.

(iii) Take any $x \in \mathbb{R}^{n+1}$. To prove that μ satisfies the triangle inequality, we must show that $\mu(x_j, x_t) \leq \mu(x_j, x) + \mu(x, x_t)$. The condition holds if and only if

$$
\begin{aligned}
f(x_j, x_t) \quad \leq \quad & f(x_j, x) + f(x, x_t) \\
& + \beta(\kappa - \kappa_j) + \beta(\kappa_t - \kappa) - \beta(\kappa_t - \kappa_j).
\end{aligned}
$$

Observe that $\beta(\kappa - \kappa_j) + \beta(\kappa_t - \kappa) - \beta(\kappa_t - \kappa_j) = 0$. Observe further that

$$
f(x_j, x_t) \leq f(x_j, x) + f(x, x_t)
$$

holds because f is a metric on $\mathbb{R}^{n+1}$. Hence, the condition holds. ∎

B Selection of $\ln x_5$ by multivariable fractional polynomial regression

As stated in footnote 26, the log-transformation of the value of the property saved (x_5) is the specification selected by the multivariable fractional polynomial (MFP) procedure of Sauerbrei and Royston (1999).[42] The following is a brief summary of the MFP procedure. For a textbook treatment, see Royston and Altman (1994, ch. 6).

The standard MFP regression model may be expressed as

$$
y_t = Y(x_t; \theta) = b_0 + \sum_{i=1}^{h} b_i x_{it} + \sum_{i=h+1}^{n} \sum_{j=1}^{m} b_{ij} x_{it}^{(p_j)} + \varepsilon_t, \quad t = 1, ..., T,
$$

where $\varepsilon_t \overset{iid}{\sim} N\left(0, \sigma^2\right)$ and $\theta = (b_0, b_1, ..., b_h, b_{h+1,1}, ..., b_{h+1,m}, ..., b_{n1}, ..., b_{nm}, \sigma^2)$. The first h covariates, $x_1, ..., x_h$, are binary, categorical, or ordinal, and the remaining covariates, $x_{h+1}, ..., x_n$, are continuous. The round bracket notation signifies the Box-Tidwell transformation,

$$
x_{it}^{(p_j)} = \begin{cases} x_{it}^{p_j} & \text{for } p_j \neq 0 \\ \ln x_{it} & \text{for } p_j = 0 \end{cases}.
$$

The MFP algorithm selects the covariates and the powers $p_1, ..., p_m$ for the continuous covariates. The researcher predefines the set of potential covariates, the set of potential powers, denoted $\mathcal{P}$, and the maximum degree of the fractional polynomial, denoted M. The researcher also predefines two significance levels: α_1, which determines the critical value for variable selection; and α_2, which determines the critical value for power selection. The covariates are selected using a backward elimination procedure in which the potential covariates are iteratively removed and added based a sequence of significance

---

[42] A fractional polynomial is an extension of a conventional polynomial that allows for noninteger and negative powers.

tests at level α_1. The powers are selected using a closed test procedure in which an M degree fractional polynomial is tested at level α_2 against a linear model and then, if and as necessary, against increasingly complex fractional polynomials. Once the covariates and powers are selected, the parameter vector θ is estimated by maximum likelihood.

To select the specification for the value of the property saved (x_5), which is the only continuous covariate, I ran an MFP regression in which the dependent variable is the log-transformed salvage award, the set of potential covariates comprises the six *Blackwall* factors, the set of potential powers is $\mathcal{P} = \{-4, -3, -2, -1, -\frac{1}{2}, 0, -\frac{1}{2}, 1, 2, 3, 4, 5, 6, 7, 8\}$, the maximum degree of the fractional polynomial is $M = 5$, the significance level for covariate selection is $\alpha_1 = 1$, and the significance level for power selection is $\alpha_2 = 0.05$. Note that setting $\alpha_1 = 1$ forces the MFP algorithm to select all the covariates into the model, which is justified here by the doctrinal principle that courts are bound to apply all of the *Blackwall* factors in determining salvage awards. Although the model allows for a five-degree fractional polynomial in x_5 with powers ranging from -4 to 8, the MFP algorithm selects a one-degree fractional polynomial with power zero, which corresponds to a simple log transformation. For further details, see Teitelbaum (forthcoming).

References

Alexander, Larry. 1996. Bad beginnings. *University of Pennsylvania Law Review* 145: 57–87.

———. 1998. The banality of legal reasoning. *Notre Dame Law Review* 73:517–533.

Alexander, Larry, and Emily Sherwin. 2008. *Demystifying legal reasoning.* Cambridge: Cambridge University Press.

Anselin, Luc. 1988. *Spatial econometrics: Methods and models.* Dordrecht: Kluwer Academic Publishers.

Argenziano, Rossella, and Itzhak Gilboa. 2012. History as a coordination device. *Theory and Decision* 73:501–512.

Banerjee, Sudipto, Bradley. P. Carlin, and Alan E. Gelfand. 2004. *Hierarchical modeling and analysis for spatial data.* Boca Raton, FL: Chapman & Hall/CRC.

Benesh, Sara C. 2002. *The U.S. court of appeals and the law of confessions: Perspectives on the hierachy of justice.* New York: LFP Scholarly Publishing.

Billot, Antoine, Itzhak Gilboa, Dov Samet, and David Schmeidler. 2005. Probabilities as similarity-weighted frequencies. *Econometrica* 73:1125–1136.

Billot, Antoine, Itzhak Gilboa, and David Schmeidler. 2008. Axiomatization of an exponential similarity function. *Mathematical Social Science* 55:107–115.

Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. 1994. *Time series analysis: Forecasting and control*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Bradley, James V. 1968. *Distribution-free statistical tests*. Engelwood Cliffs, NJ: Prentice-Hall.

Brewer, Scott. 1996. Exemplary reasoning: Semantics, pragmatics, and the rational force of legal argument by analogy. *Harvard Law Review* 109:923–1028.

Cross, Frank B. 2003. Decisionmaking in the U.S. circuit courts of appeals. *California Law Review* 91:1459–1515.

Dasarathy, Belur V., ed. 1991. *Nearest neighbor (NN) norms: NN pattern classification techniques*. Los Alamitos, CA: IEEE Computer Society Press.

Davidson, Russell, and James G. MacKinnon. 2004. *Econometric theory and methods*. New York: Oxford University Press.

Devroye, Luc, László Györfi, and Gábor Lugosi. 1996. *A probabilistic theory of pattern recognition*. New York: Springer.

Drezner, Zvi, and George O. Wesolowsky. 1989. The assymetric distance location problem. *Transportation Science* 23:201–207.

Eisenberg, Melvin Aron. 1988. *The nature of the common law*. Cambridge, MA: Harvard University Press.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64:813–836.

Epstein, Lee, and Jack Knight. 1998. *The choices justices make*. Washington, DC: CQ Press.

Force, Robert. 2004. *Admiralty and maritime law*. Washington, DC: Federal Judicial Center.

Fried, Charles. 1981. The artificial reason of the law or: What lawyers know. *Texas Law Review* 60:35–58.

Gayer, Gabrielle, Itzhak Gilboa, and Offer Lieberman. 2007. Rule-based and case-based reasoning in housing prices. *B.E. Journal of Theoretical Economics* 7:art. 10.

Gilboa, Itzhak, Offer Lieberman, and David Schmeidler. 2006. Empirical similarity. *Review of Economics and Statistics* 88:433–444.

———. 2011. A similarity-based approach to prediction. *Journal of Econometrics* 162: 124–131.

Gilboa, Itzhak, and David Schmeidler. 1995. Case-based decision theory. *Quarterly Journal of Economics* 110:605–639.

———. 1996. Case-based optimization. *Games and Economic Behavior* 15:1–26.

———. 1997. Act similarity in case-based decision theory. *Economic Theory* 9:47–61.

———. 2000. Case-based knowledge and induction. *IEEE Transactions on Systems Man, and Cybernetics—Part A: Systems and Humans* 30:85–95.

———. 2001. *A theory of case-based decisions*. Cambridge: Cambridge University Press.

———. 2002. Cognitive foundations of probability. *Mathematics of Operations Research* 27:65–81.

———. 2003. Inductive inference: An axiomatic approach. *Econometrica* 71:1–26.

Gilboa, Itzhak, David Schmeidler, and Peter P. Wakker. 2002. Utility in case-based decision theory. *Journal of Economic Theory* 105:483–502.

Glaeser, Edward L., Bruce I. Sacerdote, and Jose A. Scheinkman. 2003. The social multiplier. *Journal of the European Economic Association* 1:345–353.

Haldrup, Niels, and Michael Jansson. 2006. Improving size and power in unit root testing. In *Palgrave handbook of econometrics, volume 1: Econometric theory*, ed. Terence C. Mills and Kerry Patterson, 252–277. New York: Palgrave Macmillan.

Hodgson, M. John, Richard T. Wong, and John Honsaker. 1987. The p-centroid problem on an inclined plane. *Operations Research* 35:221–233.

Hume, David. 1748. *An enquiry concerning human understanding*. Oxford: Oxford University Press.

Hunter, Dan. 2001. Reason is too large: Analogy and precedent in law. *Emory Law Journal* 50:1197–1264.

Johnston, Jack, and John DiNardo. 1997. *Econometric methods*. 4th ed. New York: McGraw-Hill.

Kastellec, Jonathan. 2010. The satistical analysis of judicial decisions and legal rules with classification trees. *Journal of Empirical Legal Studies* 7:202–230.

Levi, Edward H. 1949. *An introduction to legal reasoning.* Chicago: University of Chicago Press.

Levy, Gilat. 2005. Careerist judges and the appeals process. *RAND Journal of Economics* 36:275–297.

Lieberman, Offer. 2010. Asymptotic theory for empirical similarity models. *Econometric Theory* 26:1032–1059.

———. 2012. A similarity-based approach to time-varying coefficient non-stationary autoregression. *Journal of Time Series Analysis* 33:484–502.

Ljung, G. M., and G. E. P. Box. 1978. On a measure of lack of fit in time series models. *Biometrika* 65:297–303.

Macagno, Francis, and Douglas Walton. 2009. Argument from analogy in law, the classical tradition, and recent theories. *Philosophy and Rhetoric* 42:154–182.

Mackaay, Ejan, and Piere Robillard. 1974. Predicting judicial decisions: The nearest neighbor rule and visual representation of case patterns. *Datenverarbeitung im Recht* 3:302–331.

Miceli, Thomas J., and Metin M. Coşgel. 1994. Reputation and judicial decision-making. *Journal of Economic Behavior and Organization* 23:31–51.

Murray, James R. 1982. The role of analogy in legal reasoning. *UCLA Law Review* 29: 833–871.

Ng, Serena, and Pierre Perron. 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90:268–281.

———. 2000. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69:1519–1554.

Pagan, Adrian, and Aman Ullah. 1999. *Nonparametric econometrics.* Cambridge: Cambridge University Press.

Plastria, Frank. 1992. On destination optimality in asymmetric distance Fermat-Weber problems. *Annals of Operations Research* 40:355–369.

23

Posner, Richard A. 1990. *The problems of jurisprudence.* Cambridge, MA: Harvard University Press.

———. 1995. *Overcoming law.* Cambridge, MA: Harvard University Press.

———. 2006. Reasoning by analogy. Review of *Legal reason: The use of analogy in legal argument*, by Lloyd L. Weinreb. *Cornell Law Review* 91:761–774.

———. 2008. *How judges think.* Cambridge, MA: Harvard University Press.

Riesbeck, Christopher K., and Roger C. Schank. 1989. *Inside case-based reasoning.* Hillsdale, NJ: Lawrence Erlbaum.

Royston, Patrick, and Douglas G. Altman. 1994. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics* 43: 429–467.

Sauerbrei, Willi, and Patrick Royston. 1999. Building multivariable prognostic and diagnostic models: Transformation of the predictors using fractional polynomials. *Journal of the Royal Statistical Society, Series A* 162:71–94.

Schauer, Frederick. 1991. *Playing by the rules: A philosophical examination of rule-based decision making in law and in life.* Oxford: Clarendon Press.

———. 2009. *Thinking like a lawyer.* Cambridge, MA: Harvard University Press.

Schoenbaum, Thomas J. 2011. *Admiralty and maritime law.* 5th ed. St. Paul, MN: Thomson/West.

Segal, Jeffrey A., and Howard J. Spaeth. 1993. *The Supreme Court and the attitudinal model.* Cambridge: Cambridge University Press.

———. 2002. *The Supreme Court and the attitudinal model revisited.* Cambridge: Cambridge University Press.

Shepard, Roger N. 1987. Toward a universal law of generalization for psychological science. *Science* 237:1317–1323.

Smith, Joseph L., and Emerson H. Tiller. 2002. The strategy of judging: Evidence from administrative law. *Journal of Legal Studies* 31:61–82.

Sunstein, Cass R. 1993. On analogical reasoning. *Harvard Law Review* 106:741–791.

———. 1996. *Legal reasoning and political conflict.* New York: Oxford University Press.

Teitelbaum, Joshua C. 2013. Asymmetric empirical similarity. *Mathematical Social Sciences* 66:346–351.

———. Forthcoming. Inside the *Blackwall* box: Explaining U.S. marine salvage awards. *Supreme Court Economic Review.*

Tsay, Ruey S. 2005. *Analysis of financial time series.* 2nd ed. Hoboken, NJ: John Wiley.

Weinreb, Lloyd L. 2005. *Legal reason: The use of analogy in legal argument.* Cambridge: Cambridge University Press.

Westen, Peter. 1982. On "confusing ideas": Reply. *Yale Law Journal* 91:1153–1165.

White, K. Geoffrey. 2001. Forgetting functions. *Animal Learning and Behavior* 29: 193–207.

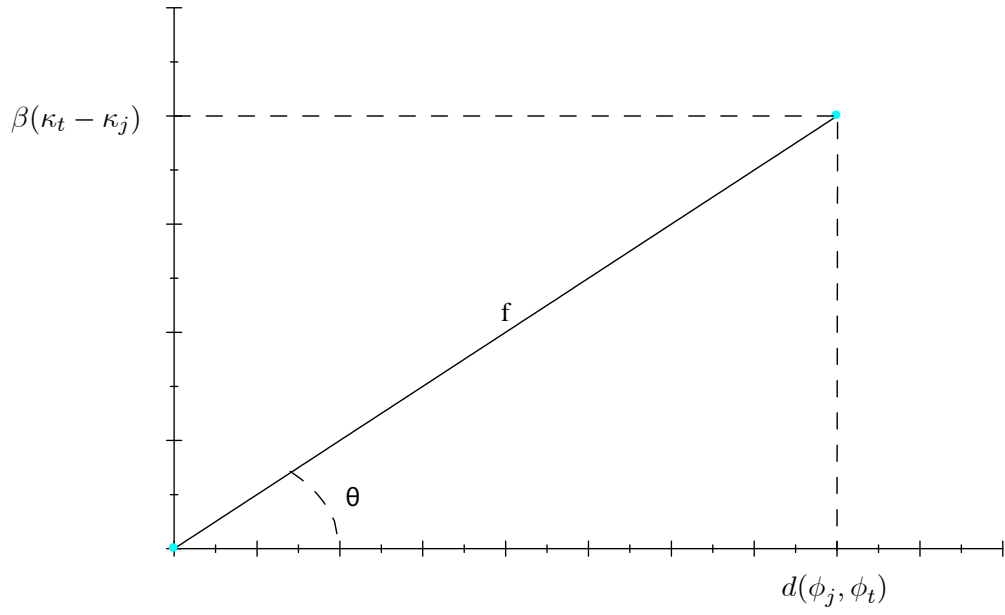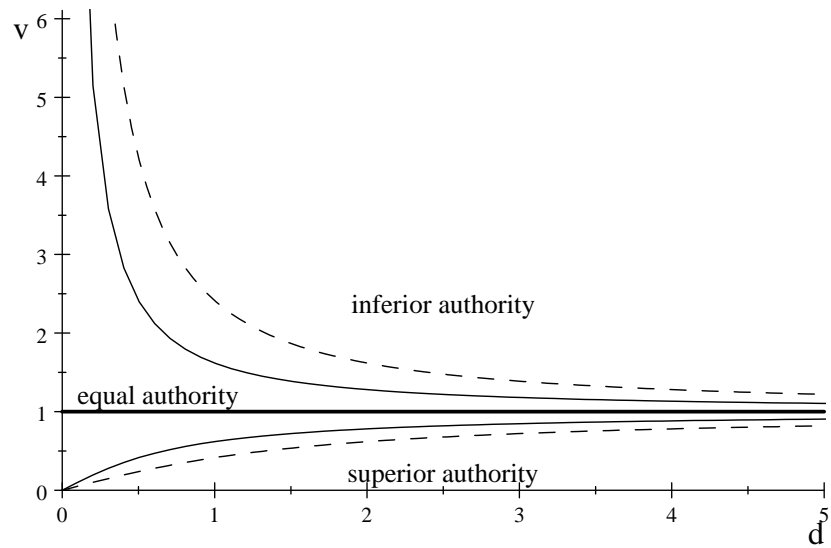Wilson, W. A. 1931. On quasi-metric spaces. *American Journal of Mathematics* 53: 675–684.

Figure 1: Geometry of v.

$$v = \sec\theta + \tan\theta,$$
$$\theta = \arctan(\beta(\kappa_t - \kappa_j)/d)$$

Thick:	$\beta	\kappa_t - \kappa_j	= 0$
Thin:	$\beta	\kappa_t - \kappa_j	= \frac{1}{2}$
Dash:	$\beta	\kappa_t - \kappa_j	= 1$

Figure 2: Visualization of v.

$$s = \exp(-vd)$$

Thick:	$\beta|\kappa_t - \kappa_j| = 0$
Thin:	$\beta|\kappa_t - \kappa_j| = \frac{1}{2}$
Dash:	$\beta|\kappa_t - \kappa_j| = 1$
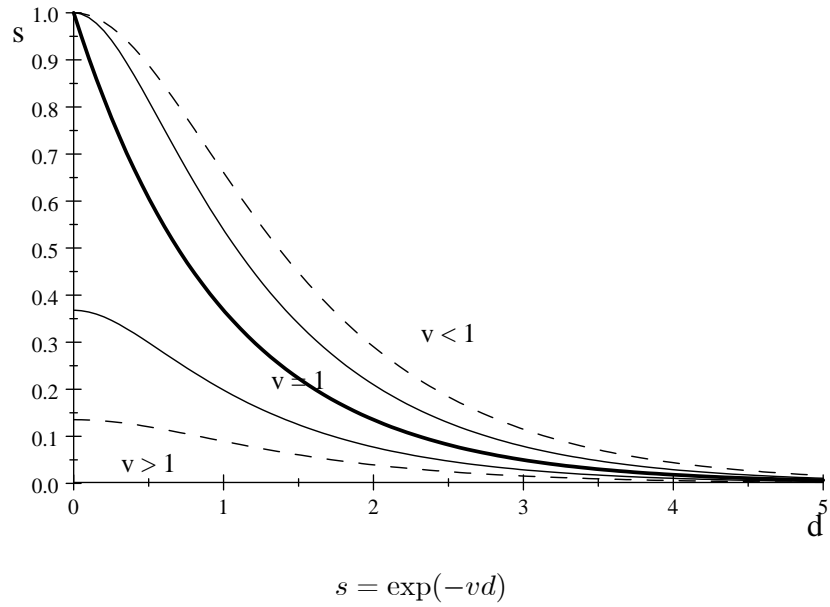
Figure 3: Precedential influence in the ALR model.

$$s = \exp(-vd)$$

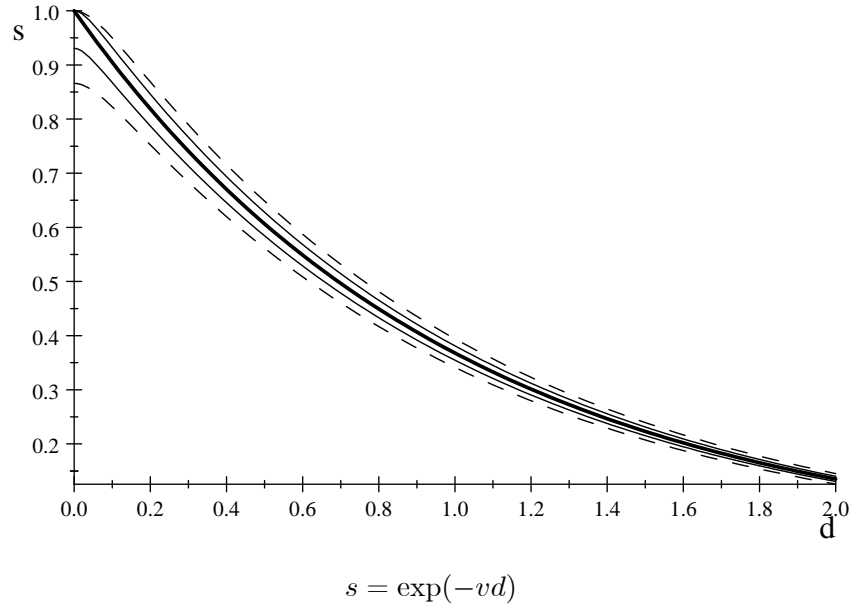| Thick:	$\beta = 0.036, |\kappa_t - \kappa_j| = 0$
Thin:	$\beta = 0.036, |\kappa_t - \kappa_j| = 1$
Dash:	$\beta = 0.036, |\kappa_t - \kappa_j| = 2$

Figure 4: Influence penalty/bonus implied by $\beta = 0.036$.

Table 1: Summary Statistics

	Variable	Mean	Standard deviation	Minimum	Maximum
y	Salvage award	74.44	153.35	0.24	1,865
x_1	Labor expended by salvors	0.39	0.49	0	1
x_2	Skill displayed by salvors	0.44	0.50	0	1
x_3	Danger to salvors' property	0.27	0.45	0	1
x_4	Risk to salvors	0.18	0.39	0	1
x_5	Value of property saved	1,385.71	2,913.98	1.20	42,133
x_6	Danger to property saved	0.51	0.50	0	1

Notes: 684 cases from 1799 to 2007. y and x_5 in thousands of 1980 U.S. dollars.

Table 2: Conditional Salvage Awards

Variable	Cases	Mean
y	684	74.44
y if $x_1 = 0$	415	35.67
y if $x_1 = 1$	269	134.24
y if $x_2 = 0$	382	51.07
y if $x_2 = 1$	302	103.99
y if $x_3 = 0$	498	57.57
y if $x_3 = 1$	186	119.60
y if $x_4 = 0$	559	58.41
y if $x_4 = 1$	125	146.09
y if $x_5 \leq \overline{x}_5$	504	37.98
y if $x_5 > \overline{x}_5$	180	176.50
y if $x_6 = 0$	337	45.18
y if $x_6 = 1$	347	102.85

Notes: y in thousands of 1980 U.S. dollars. $\overline{x}_5$ denotes the mean of x_5.

Table 3: Autocorrelation and Unit Root Tests

Test	Test statistic	Five percent critical value
Ljung-Box test:		
$p = 6$	52.604	12.592
$p = 20$	119.360	31.410
$p = 40$	205.136	55.758
Runs test	-1.817	-1.645
ADF-GLS test $(p = 40)$	-1.520	-1.950

Notes: For the Runs test, the critical value reflects a one-sided test against the alternative of positive autocorrelation. For the ADF-GLS test, the critical value is interpolated from tables presented by Elliott et al. (1996).

Table 4: ALR Model – Maximim Likelihood Estimates

Parameter		Estimate	Standard error
β	Shape parameter for v	0.036*	0.017
ω_1	Weight on x_1	14.822*	2.602
ω_2	Weight on x_2	0.785	0.843
ω_3	Weight on x_3	3.419†	2.081
ω_4	Weight on x_4	6.101*	3.092
ω_5	Weight on $\ln x_5$	9.232*	1.777
ω_6	Weight on x_6	11.823*	2.888
σ^2	Variance of ε_t	1.005*	0.019
	Loglikelihood	-974.097	

Notes: Dependent variable is $\ln y$. 684 cases from 1799 to 2007.

*Significant at the one percent level (one-tailed test).

†Significant at the ten percent level (one-tailed test).