Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP04/2020

The (in)efficiency of Justice. An equilibrium analysis of supply policies

Antonio Peyrache and Angelo Zago

Date: March 2020

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

The (in)efficiency of the Market for Justice. An equilibrium analysis of supply policies

Antonio Peyrache* Angelo Zago†

Abstract. This paper develops an equilibrium analysis of the market for justice, explicitly allowing for the interaction of the supply and demand of justice and looking at completion time as the rationing mechanism clearing the market. The model is estimated using OLS and IV regressions coupled with a nonparametric estimate of the supply function. We use detailed data from 165 Italian courts between 2005-2012 because the Italian setting enables us to identify the demand function. In fact, in Italy, citizens cannot vote with their feet by choosing where (in which court) to file their case. Hearings take place in the local district court. This rigidity means that the only way the market for justice can clear is via an increase or decrease in the completion time.

Our main empirical findings are that, due to the rigidity of the demand for justice, improvements in the supply have a significant effect on the overall completion time of the system. Specifically, the introduction of best practices aimed at improving case management, the break-up of large courts, the increase in the number of judges, and the reallocation of judges to different courts could significantly reduce the length of trials. A key finding is that, if all four policies were to be implemented, then the average delay would be more than halved, from the current 14 months to around 6 months. To provide an idea of the opportunity cost of these policies we show that the introduction of best practices is equivalent to an increase of about 25% in the number of judges; this policy of simple expansion of the supply would cost around 100 million euros.

Keywords: Courts; Efficiency; Equilibrium; Processing Times.

*School of Economics, University of Queensland, St. Lucia, QLD 4072, Australia. a.peyrache@uq.edu.au

†Dipartimento di Scienze Economiche, Università degli Studi di Verona, Via Cantarane, 24, Verona 37129, Italy. angelo.zago@univr.it

1 Introduction

The judicial system (and the rule of law) has a very important role in securing property rights and enforcing contracts, thus affecting economic behavior, investment choices and economic growth (Aldashev, 2009). An inefficient judicial sector may negatively impact (among other things) on credit markets (Jappelli et al., 2005; Ponticelli and Alencar, 2016), entrepreneurship (Chemin, 2009), investments (Chemin, 2012), the size and growth of firms (Kumar et al., 2001; Giacomelli and Menon, 2017), labor markets (Ichino et al., 2003), FDI (Nunn, 2007), public procurement (Coviello et al., 2018) and housing markets (Mora-Sanguinetti, 2012). In addition, as understood early on by Adam Smith (1776, in Landes and Posner, 1979), the administration of justice is "one of the few proper functions of government", since "private security and enforcement, while working well in some environments, often degenerate into violence" (Djankov et al., 2003: 454). Furthermore, the judiciary is one sector of the economy where the market system cannot work properly, given the absence of a functioning output price mechanism that could sanction inefficient courts. Moreover, in many countries, citizens cannot "vote with their feet" because cases are allocated to local courts not of the plaintiffs' choosing. Thus, "it is crucial to understand the factors that make courts function more or less effectively" (Djankov et al., 2003: 454), both in developed and developing countries (Palumbo et al., 2013).

In this paper we undertake an equilibrium analysis of the market for justice by focusing on supply policies designed to increase the efficiency of the system. While most of the studies we are aware of (illustrated below) consider the demand for justice, we provide a model that can account for inefficiencies arising from its supply side. A measure of performance often used by practitioners is the average length of trials[1] as a way of spotting problematic countries or courts within a country (see, e.g., CEPEJ, 2016). However, looking only at completion time without considering resource use is not fully informative and may be misleading. By using a rather general production model able to take resource use into account we can study the (steady state) performance of the justice sector as a whole and improve both on the analysis of length of trials and on standard measures of partial productivity, e.g., the

[1]As explained in a recent OECD report, "the focus on length is motivated not only by the importance of a timely resolution of disputes for the correct functioning of the economy, but also by the fact that a reasonable trial length is a necessary (though not a sufficient) condition for good performance in other dimensions [..] Also, as emphasized by the adage *justice delayed is justice denied*, timeliness is a prerequisite for achieving justice. Moreover, the length of trials is also generally associated with other crucial measures of performance such as confidence in the justice system" (Palumbo et al., 2013: p. 9).

number of completed cases per judge. We can then relate these performance measures, i.e., the time needed to complete cases, to the possible causes of excessive trial length, enabling us to provide policy suggestions targeted on the sources of inefficiency and their geographical distribution. In particular, we consider four supply policies: the introduction of best practices at the court level; the break-up of large courts (to exploit scale economies); the increase in the number of judges; and the optimal reallocation of judges to courts. We consider the effect of these policies both on the average completion time of the system and on the distribution of completion times for the different courts.

In order to account for resource use, we consider a production frontier for the courts of justice where the number of pending cases is a *variable* input and the number of judges is a *fixed* production factor (a capacity input). For a given number of judges, when the number of pending cases increases, the number of completed cases first increases, then reaches a maximum and finally decreases due to congestion. For this reason the model can accommodate both variable returns to scale and production congestion.[2]

The supply policy analysis is carried out taking into account possible demand effects and therefore in terms of a notion of steady state equilibrium of the system. We suggest that observed slow processing times may be the outcome of a steady state equilibrium with a high number of pending cases (long queues). Implementing the suggested supply policies would shift the system to a new steady state equilibrium with low processing times and short queues.

Identification of the supply and demand functions can be problematic due to a potential simultaneity bias (indeed supply and demand are used as a classical example of such a bias). In order to identify and separate supply and demand effects, we make use of a recent dataset for the Italian court system. In Italy, cases are assigned where they occur and there is no significant migration between the various districts. These two conditions can be used to identify the demand function and the elasticity of demand in relation to processing time. Hence supply and demand can be identified and a counterfactual analysis carried out.

Italy is interesting not only in relation to this identification strategy, but also because of its poor performance and its heterogeneity. Italian courts are among the most inefficient in OECD countries in terms of trial length. The *average disposition time* for a standard commercial case in 2016 was

[2]Congestion may be due, for instance, to *task juggling* (Coviello et al., 2014). The first paper to acknowledge congestion problems in courts is probably Buscaglia and Dakolias (1999), but to the best of our knowledge few papers have dealt with it, only Dimitrova-Grajzla et al. (2012), Coviello et al. (2015), and Bray et al. (2016).

1,120 days in Italy, against 553 in OECD countries (regional average), 395 in France, 499 in Germany and 510 in Spain. Moreover, among OECD countries, Italy appears to have longer trials at all levels, that is in courts of first and second instance and in the highest court (see, e.g., Palumbo et al., 2013, p. 14, Figure 2). According to the World Bank (Doing Business, 2020 edition), Italy ranks 122^{nd} out of 190 countries in terms of enforcing contracts, compared to Germany (13^{th}), France (16^{th}), Spain (26^{th}), and UK (34^{th}). These lacklustre performances are not a recent phenomenon, dating back at least to the fifties.[3]

The average trial length is quite different across Italian regions, with lengthier processes and larger stocks of pending cases in the South (Carmignani and Giacomelli, 2009). However, given that southern courts are provided with more human resources, it is important "to establish whether and to what extent the larger stock of pending cases is due to lack of resources or to their lower productivity" (Carmignani and Giacomelli, 2009: 21).

It is not surprising therefore that the Italian justice system has been investigated quite extensively. In recent years, there has been a lively discussion of the possible causes of these inefficiencies and, in particular, of *pathological demand* effects (Marchesi, 2003), according to which higher litigation rates are the result of lengthy trials. Delays in delivering justice could lead some economic agents (households, workers and firms) to exploit these inefficiencies by strategically postponing their contractual obligations to other parties, and this is more likely to happen the wider the gap between legal and market interest rates (see, e.g., Marchesi, 2003; Felli et al., 2008; Padrini et al., 2009). Other theories point to *supplier-induced demand*, (see, e.g., Carmignani and Giacomelli, 2010 and Buonanno and Galizzi, 2014), according to which the combination of the increase in the number of lawyers and the minimum fee leads to excessive litigation. However, the empirical evidence regarding these possible demand-side causes is rather ambiguous, and numerous studies have been calling for a complementary supply-side analysis (see, e.g., Bianco and Palumbo, 2007; Felli et al., 2008). Indeed, albeit a country with one of the highest litigation rates, Italy is given as the example where "there is scope for improvements also on the supply side, for instance expanding the use of case-flow management techniques" (Palumbo et al., 2013: 45), a policy aimed at introducing best (management) practices.

To empirically test our model predictions, we collected data for all Italian

[3]In an early comparison of the relative perfomances of the Italian and US systems, for instance, Chase (1988) cites previous complaints about the problems caused by lengthy trials in Calamandrei, who noted in 1956 the "slowness of the judicial process, of which everyone in Italy complains" and argued that its abuses are "deeply rooted in judicial practice".

courts (165) for the period 2005-2012,[4] taking advantage of data now publicly available and collecting additional data from other sources. Overall, we find that technical (best practices), size (break-ups) and reallocation inefficiencies are the major issues at the industry level. Congestion inefficiency does not have a big impact at the aggregate level, but is a major problem in some specific courts, evenly distributed across Italy.

Given these findings, we argue that the most impactful policy would be the introduction of best practices, which would have effects throughout the system, including in the inefficient courts of southern Italy. Another effective policy might be to increase the number of judges, although its cost-effectiveness and hence feasibility might be questioned. Overall, despite the fact that the external validity of our results cannot be taken for granted, in Italy they demonstrate the scope for significant improvements in efficiency by suppy-side reforms aimed at shortening the length of trials.

Section 2 introduces a market justice model. After a brief review of the most significant literature, the supply model is presented. Section 3 introduces the computational models, section 4 the data, and section 5 the empirical results. The final section concludes with some suggestions for further research.

2 The market for justice

Most studies on the "market for justice" deal amply with its demand side (see the reviews in Cooter and Rubinfeld, 1989, for the market for legal disputes and, for instance, Buonanno and Galizzi, 2014 for the market for legal services, i.e., lawyers). This paper focuses on an equilibrium analysis of the market for justice based on both supply and demand. Given the absence of a price mechanism in the market for justice, anticipated completion time plays the role of a rationing mechanism keeping supply and demand in balance.

2.1 The demand for justice

In the market for justice, supply and demand are in play. We can measure the demand for justice by the number of incoming cases, and supply by the number of cases completed, i.e., resolved, over the same period. Cooter and Rubinfeld (1989) summarize various contributions in the literature on the market for justice and propose a 'hybrid' model (of previous contributions)

[4]Before the change in court geography introduced at the end of 2012 by the Monti Government.

that takes into account different actors and different stages of the legal disputes to provide the micro-foundations of the demand for justice. At stage 1, one party may damage or harm another. At stage 2, the damaged party decides whether to take legal action. Again, her decision depends on a comparison between the costs of asserting the claim, e.g., hiring a lawyer, and the benefits, such as the expected proceeds from the settlement or winning the case. At stage 3, the parties decide whether to settle out of court. Finally, at stage 4, the case is brought to, and settled by, the court. These different stages are analyzed in a backward fashion.

The parties' behavior at trial is based on the merit of the case and on the efforts of both parties. Each party decides its own level of effort based on what it expects to achieve (i.e., a win or a loss for the defendant). Different exogenous variables can affect these levels, such as adjustments to damages, the impact on reputation, etc. On a similar note, "there is more scope for settlement when litigation is costly [..], negotiations are inexpensive [..], and the disputants are pessimistic about trial outcomes" (Cooter and Rubinfeld, 1989: 1076). From their hybrid model, Cooter and Rubinfeld (1989) then suggest a structural model for empirical research on the trial/settlement split by assuming that the plaintiff's expected gain (defendant's expected loss) from the trial comprises a systematic component and a randomly distributed error. From this a reduced-form model can be obtained, in which the likelihood of going to trial can be determined by evaluating empirically the probability distribution function of the systematic components empirically. They also survey the papers that empirically implement some variants of such a model.

An important dimension in modeling and estimating the demand for justice is trial length. Gravelle (1990) was probably one of the first to understand this in relation to the courts. Economists have long argued that economic efficiency should lead to a preference for rationing rather than waiting (see Barzel, 1974).[5] However, a trial is not a standard commodity, because its demand is generated by a sequence of decisions,[6] and courts are rationed by waiting lists rather than lines. Therefore, an increase in processing times

[5] The idea is that the full price of a commodity may include both the monetary price and the opportunity cost of the waiting time. In the case of lengthy waiting times, an increase in the monetary price of the commodity leads to a lower waiting time. The market clearing *full* price of the commodity does not change - and so consumers are indifferent between the two different equilibria - but the suppliers are better off. Rationing by waiting times, on the other hand, is inefficient because it imposes a loss on consumers not compensated by any gain to suppliers.

[6] There are at least two sets of decisions (and relevant affecting variables): the decision to commit a crime (or the occurence of an accident), together with actions to possibly prevent or avoid it, and the decision to settle the case out of court.

may be efficient to the extent that it reduces the net costs of a trial. The bottom line is that "Because some of the decisions which lead to trials are inefficient and are not readily correctible by other means, it can be efficient to ration by waiting" (Gravelle, 1990: 270). However, given that a trial occurs when an accident happens and when the same is not settled out of court, the length of the trial affects the likelihood of both events - accident happening and settlement occurring - often in different directions, the overall effect of the expected time to settle a case on trial demand may be ambiguous, and thus needs to be empirically determined.[7]

In fact, various papers have sought to empirically estimate the demand for justice in a number of countries. One is Felli et al. (2008), probably the closest to our contribution in terms of estimating demand. They develop an economic model of a litigant's decision to go to trial or to settle following, among others, Van Wijck and VanVelthoven (2000), basing these choices on the expectations regarding the possible outcome of the case, the costs of proceedings and the costs of negotiation. They then suggest an empirical specification in logarithmic first differences, regressing the demand for justice (incoming cases) as a function of different explicative variables, such as the average length of trials, average real earnings for and the number of lawyers, the real market and legal interest rates, some proxies for the business cycle, and a dummy for the year (1995) in which a significant institutional reform was introduced. Estimating their model with a panel of the 26 districts over the period 1991-2002 using fixed effects and GMM, they find that almost all variables are significant (R^2 between 0.64 and 0.8) and, more importantly for our purposes, that a 1% increase in trial length was associated with a fall in demand of 0.75%.[8]

Other studies have empirically estimated the demand for justice. Ginsburg and Hoetker (2006), for instance, estimate the demand for justice in Japan and test different possible explanations - cultural, institutional, political - for the historical low level and subsequent increase in litigation rates. They take into account the increase in the number of lawyers and judges, the procedural and legal reforms, and the structural changes in the economy. Using data on 47 prefectures for the 1986-2001 period, they find that institutional constraints explain the relatively low rate of litigation in Japan. Indeed, the shortage of lawyers, the shortage of judges, and procedural barriers show results consistent with the lack of institutional capacity in the

[7]Jappelli et al. (2005) is one of the few papers looking at the effect of trial length on the decision to forfeit a contractual obligation by opportunistic borrowers. They find a significant effect of judicial length on these decisions and therefore on the functioning of Italian credit markets.

[8]They also provide a relatively updated survey of this literature.

legal system as a major predictor of a low litigation rate. Moreover, the paper rejects the hypothesis that the availability of alternative methods of dispute resolution may be a key factor suppressing litigation, and also that cultural differences may be important. Lastly, their analysis suggests that absolute levels of wealth increase litigation, but economic decline does so as well. However, they do not explicitly consider whether the time to resolve a case in court can influence the demand for justice.

Related literature has tried to shed light on the causes of excessive Italian trial length focusing on the market for lawyers. Buonanno and Galizzi (2014), for instance, suggest a different explanation for the emergence of long trials, based on the degree of competition in the market for lawyers. The idea is that when the number of lawyers increases and when a minimum fee is imposed (as in the Italian case), lawyers may opportunistically convince their clients to go to court more often (than optimal from the client's point of view). They provide empirical evidence to test this supplier-induced demand hypothesis. Since new lawyers may move to where demand is higher, provoking reverse causality, they use two instruments for the number of lawyers, one geographic and one historical.[9] Their main finding is that the number of lawyers operating in a court exerts a positive and statistically significant effect on the litigation rate. This result is robust to different checks, controlling for the general social and economic conditions of different Italian provinces, their urbanization rate and their level of human capital. Buonanno and Galizzi (2014) find that the demand for justice is inversely related to trial length (lagged length).

Carmignani and Giacomelli (2010) carry out a similar analysis of Italian data for the years 2000-2005. They find a positive correlation between lawyers and litigation. They also use a 2SLS approach to control for endogeneity, using the proximity of provinces to a law school in 1975 as an instrument, confirming that the number of lawyers has a significant positive effect on civil litigation. The results hold with different specifications after controlling for economic conditions and the economic cycle, social capital, urbanization, and crime rates. Using the lagged value of trial length, they find a positive relationship between the trial length and the demand for justice (measured by incoming cases).[10]

[9]Note that Ginsburg and Hoetker (2006) also use an instrumental variable approach for the number of lawyers and judges.

[10]Mora-Sanguinetti and Garoupa (2015) undertake a similar study for Spain, a country in which the litigation rate is even higher than Italy (Palumbo et al., 2013). They use a similar IV approach, finding no clear evidence of endogeneity, but a clear *positive* relationship between the number of lawyers and the induced demand for justice. However, they do not consider the effect of trial length on demand.

2.2 The supply of justice

We now focus on the supply side of the market for justice, and assume that the supply of justice is related to the resources used, to how efficiently they are used and to the available technology. Moreover, the supply may depend on management, legal formalism (Djankov et al., 2003), incentives for the judges, etc. (see, e.g., Palumbo et al., 2013: 9). In general, these relationships and the technology can be expressed by the following production function

$$y = \theta F(g, p), \tag{1}$$

where the 'production' of completed (or processed, resolved, defined) cases y depends on the number of judges g and the number of pending cases p; and the parameter $\theta \in [0, 1]$ represents production efficiency. We assume that the shape of the production function complies with the law of variable proportions as described in Svensson and Färe (1980): the number of judges is considered as a capacity factor processing the number of pending cases (the variable factor of production). For a given number of judges, when the number of pending cases increases, the number of completed cases first increases, then peaks (full capacity) and subsequently declines due to congestion. This is represented in Figure 1, with the number of pending cases on the horizontal axis and the number of completed cases on the vertical axis.

The curve in Figure 1 shows the production possibility set for a given number of judges. Relating the analysis to the number of judges is important for at least two reasons. First, there can be significant differences between small and larger courts.[11] Second, previous studies have found different returns to scale in Italian courts.[12] Congestion also needs to be considered, as documented in other studies.[13] Indeed, "the inability of the system to satisfy the demand for justice (i.e., resolve in each given period the same number of cases equal to as are brought to court) generates congestion and delays" (Palumbo et al., 2013: 10).

Our production function may thus appear as in Figure 1. Note that up to point 1, there are increasing returns, while between 1 and 2 the returns decrease. At point 2, the system reaches full capacity. Beyond 2 congestion starts to kick in. The slope of the CRS closure line is $y/p = 1/t$, the inverse

[11]For instance, in 2012 the largest court in Italy (Rome) had 406 judges, while the smallest (5 courts) had 6 judges.

[12]Marchesi (2003), for instance, found that courts with fewer than 20 judges were too small, while courts with more than 80 judges suffered from decreasing returns to scale.

[13]Coviello et al. (2015), for instance, for the Milan court, found that "if a larger future caseload induced judges to increase task juggling by 1%, [..] the completion hazard would decline approximately by a factor ranging between 1% and 2.1%" (p. 909).

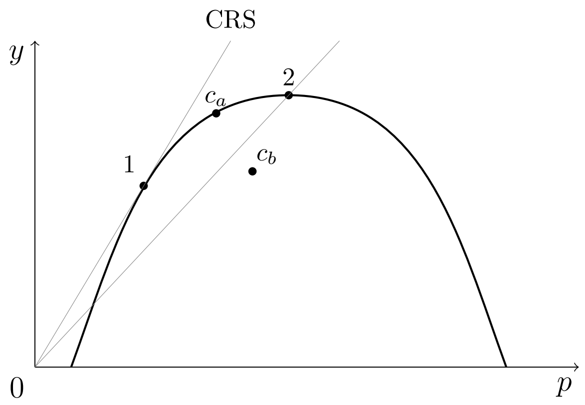


Figure 1: The production of justice

of the time needed to process a case with that capacity, given the flow of incoming cases. Processing time is therefore defined as the ratio of pending cases to the number of completed cases. This is a standard measure in queueing theory and has a very appealing, simple interpretation. Note that a change in 'capacity', in the fixed factor(s) of production, i.e., the number of judges, would shift the production frontier and potentially increase the output potential of courts.

While in our analysis it would be difficult to consider legal formalism, incentives to judges and other factors common in Italian courts, their differing efficiency levels can be taken into consideration. We consider this production model within the framework of efficiency analysis, so that a specific court can be located on the production frontier ($\theta = 1$), such as court c_a, or inside the frontier like court c_b, where the distance to the frontier represents a measure of inefficiency. In this way, we can measure the inefficiencies for each court and for the overall industry following Peyrache and Zago (2016).

An alternative way of representing this production trade-off is in the time-production (t, y) space rather than the introduced pending-completed cases space $(p,\, y)$, where t is trial length and y are the completed cases. Given the assumptions of our production model, we obtain the supply function S shown in Figure 2. Note that the backward bending portion of the supply function represented is derived from the congestion hypothesis.[14] Again, a change in the number of judges would shift the supply function in the time-production space as well as the production function in the production-pending cases space.

[14]Our technology specification is general enough to allow for this possibility. Its relevance to our data is tested in the following section.

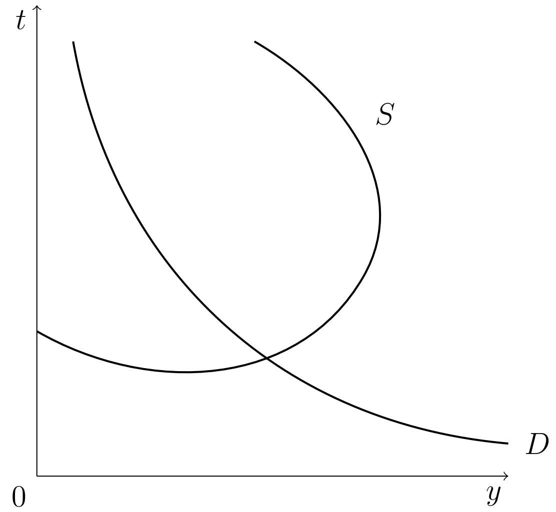


Figure 2: The equilibrium in the market for justice

2.3 Equilibrium

The equilibrium between supply and demand is obtained by considering what we will refer to as a *material balance* condition. This condition states that the variation in the number of pending cases is equal to the discrepancy between the number of incoming cases (inc) and the number of completed cases each year:

$$p_{t+1} - p_t = inc_t - y_t. \tag{2}$$

The system is in steady state if the number of pending cases is constant from one year to the next: $p_{t+1} = p_t$. Or, to put it differently, the system is in steady state if the number of incoming cases is equal to the number of completed cases each year.

Since (unlike normal markets, cleared by prices) the market for justice clears through variations in the length of proceedings (completion time, see Palumbo et al., 2013: 9), the equilibrium conditions for the market for justice involve a steady state condition

$$inc_t = y_t, \tag{3}$$

and a standard equilibrium condition in terms of supply and demand $S(t) = D(t)$. This balance produces a completion time generating demand that is equal to the number of completed cases. The system is in steady state equilibrium if these conditions hold, and this notion of steady state equilibrium is used to assess counterfactual changes in the production efficiency of the courts.

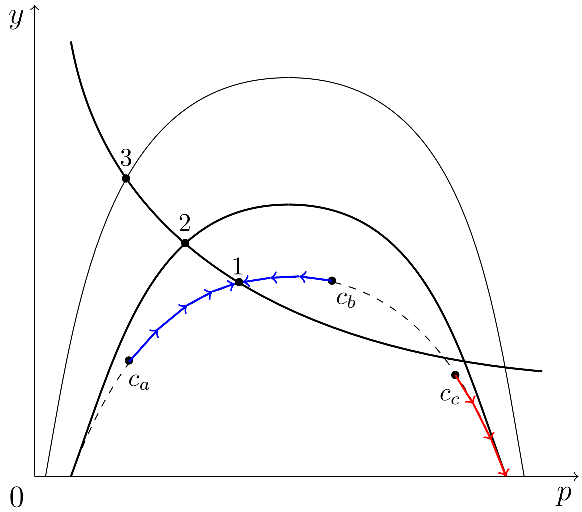


Figure 3: Efficiency and equilibrium

Figure 3 represents the same equilibrium conditions in the pending cases-completed cases space (which is more natural when thinking about production). The demand function can be inverted in order to obtain the combinations of pending-completed cases that are compatible with the demand conditions. In other words, the demand curve in this space determines all the combinations of pending and completed cases making the number of incoming cases (as determined by the demand function) equal to the number of completed cases.

The figure demonstrates a number of effects. First, there are three production functions. Passing through point 1 is the production that can be achieved for any given quantity of pending cases with a given level of efficiency $\theta < 1$. The steady state equilibrium for this level of efficiency is at point 1 where supply and demand meet. If a court is at a point like c_a, then the number of incoming cases is larger than the number of completed cases and the number of pending cases increases, moving the court back towards the equilibrium. For points like c_b the number of completed cases is larger than the number of incoming cases, so the number of pending cases decreases, moving this point towards the equilibrium. This means that equilibrium 1 is a stable steady state equilibrium. There are points like c_c, with a number of incoming cases larger than the number of completed cases: this has the effect of increasing the number of pending cases congesting production even more. The latter set of points converges to a point of full congestion. The three types of court are described below.

One final observation in this picture is that starting at the equilibrium point 1, an increase in production efficiency increases the number of completed cases above the number of incoming cases, reducing the stock of pending cases and moving the system towards a new steady state equilibrium, as shown at point 2. At this new equilibrium the court is fully efficient ($\theta = 1$) and in steady state equilibrium. Similarly, an increase in production capacity, either due to an increase in the number of judges or a better exploitation of scale economies, would shift this frontier even further towards a new steady state equilibrium such as point 3.

This is our basic conceptual framework to analyze improvements in the efficiency of the system through supply side policies aimed at introducing best practices, implementing break-ups, adding judges, and reallocating judges efficiently to the different courts.

3 Methodology

3.1 Demand

The first step is to estimate the demand for justice. For this purpose, the following regression model is used:

$$\log inc_{it} = \alpha_i + \mathbf{X}_{it}\boldsymbol{\beta} - \gamma \log t_{it-1}, \tag{4}$$

where $i = 1, \ldots, n$ stands for the courts and $t = 1, \ldots, T$ is the reference year. The variable inc_{it} is the number of incoming cases in year t and court i and $\mathbf{X}_{it}$ contains control variables, including individual level dummy variables and the population size of each court district (this is exogenously determined, since each district population refers to a given court and movements between courts is forbidden by law). The average completion time for a new case is given by the ratio of the number of pending cases to the number of completed cases $t_{it} = p_{it}/y_{it}$.[15]

The demand equation includes the lagged value for the time to complete trials. This can be interpreted both as a causal relationship between completion time and the number of incoming cases (quantity demanded) on the assumption of adaptive expectations for the plaintiff, or as a prediction equation for the number of incoming cases. To check the robustness of our estimates, we also tested the sign and size of demand elasticity (γ) by using

[15]Since we also have data for the number of judges (g_{it}) and the number of completed (y_{it}) and pending cases (p_{it}), we can compute the average queue completion time for each court in each time period.

the lagged value of the average completion time as an instrument for the following simultaneous demand equation (based on rational expectations):

$$\log inc_{it} = \alpha_i + \mathbf{X}_{it}\boldsymbol{\beta} - \gamma \log t_{it}. \tag{5}$$

The steady state equilibrium condition specifies that the number of processed cases each year must be equal to the number of incoming cases:

$$inc_{it} = y_{it}. \tag{6}$$

This equilibrium condition, together with the material balance condition (described above) means that in equilibrium the queue for the court is in steady state with the number of pending cases constant from one year to the next, which in turn (unless some of the control variables have changed from one year to the next) means that $inc_{it} = inc_{it-1}$.

The equilibrium condition and material balance condition mean that we can derive a relationship between the number of pending cases and the number of completed cases from the demand function:

$$y = p^{-\frac{\gamma}{1-\gamma}} \exp\left(\mathbf{X}\boldsymbol{\beta}/\left(1-\gamma\right)\right).$$

This equation is useful in order to analyze demand trade-offs in the pending-completed cases space. This also implies the following equilibrium relationship between the average completion time and the number of pending cases (dividing the previous equation by p and taking the inverse):

$$t = \frac{p^{\frac{1}{1-\gamma}}}{\exp\left(\mathbf{X}\boldsymbol{\beta}/\left(1-\gamma\right)\right)},$$

where $\mathbf{X}\boldsymbol{\beta}$ is the prediction based on the demand regression estimates.

Unless demand is rigid in terms of completion time ($\gamma = 0$), increasing the efficiency of production (by increasing the number of processed cases) decreases completion time and increases the number of incoming cases. Therefore efficiency gains may be overestimated if the demand side is ignored.[16]

3.2 Supply

We now turn our attention to the supply side of the market for justice where the equilibrium condition is used to determine an efficient configuration of

[16] In the estimations, as will be shown below, we find demand to be quite inelastic at a value of $\gamma = 0.16$. This means that in our dataset a 10% increase in completion time reduces the number of incoming cases by 1.6%.

the system. This will return the *equilibrium* increase in the efficiency of the system rather than a simple *optimal* increase condition (which could be out of equilibrium). Clearly, an equilibrium increase in efficiency will also be optimal, but the reverse is not in general true. For example, in the analysis of Peyrache and Zago (2016) the efficiency of the system is studied from an optimality perspective, by looking at the potential increase in output given the number of incoming and pending cases. This may result in an overestimation of the potential efficiency gains once the equilibrium condition is introduced, i.e., the output oriented efficiency measure is not an equilibrium notion. Since we are looking at equilibrium outcomes, in the following analysis we take the individual average of each variable in the panel, and consider the following data structure:

$$(\mathbf{G}, \mathbf{Y}, \mathbf{P}, \mathbf{X}\boldsymbol{\beta}), \tag{7}$$

where $\mathbf{G}$ is the vector with the average number of judges ($g_i = \frac{1}{T}\sum_t g_{it}$) for each court, $\mathbf{Y}$ the vector with the average number of completed cases ($y_i = \frac{1}{T}\sum_t y_{it}$), $\mathbf{P}$ the vector with the average number of pending cases ($p_i = \frac{1}{T}\sum_t p_{it}$) and $\mathbf{X}\boldsymbol{\beta}$ the vector of predicted incoming cases from the demand regression.

We start with a linear program that defines the level of court i efficiency, which is processing y_i cases with g_i judges and is facing p_i pending cases (decision variables are always represented by Greek letters):

$$\begin{aligned} \min_{\lambda_{ik}} \ & \theta_i \\ st \ & \sum_k \lambda_{ik} g_k \leq g_i \\ & \sum_k \lambda_{ik} p_k = p_i \\ & \sum_k \lambda_{ik} y_k \geq y_i/\theta \\ & \sum_k \lambda_{ik} = 1. \end{aligned} \tag{8}$$

This is known as the data envelopment analysis (DEA) estimator of efficiency (see Coelli et al., 2005). The only difference from a standard DEA model is the imposition of an equality constraint for the number of pending cases. Together with the standard inequality constraints associated with the number of judges and the number of completed cases, this equality constraint allows for the law of variable proportions shown in Figure 1: given the number of judges (our capacity measure) when the number of pending cases increases, the number of completed cases increases at first, peaks and then decreases (due to congestion). The efficiency score θ_i measures the distance from the frontier in terms of the additional number of pending cases that could be processed when the court is benchmarked against other courts of similar size. Therefore it measures the efficiency of best practices.

Neither the observed combination (y_i, p_i, g_i) nor the efficient combination $(y_i/\theta, p_i, g_i)$ are necessarily equilibrium levels of production, since the process time induced by these quantities may be incompatible with our demand equation. In order to define an equilibrium outcome that keeps the efficiency level of the court constant at the observed level θ_i, we use the following program where the relationship derived from the demand function is explicitly included:

$$\begin{aligned} \min_{\lambda_{ik},\pi_i,\tau_i} \quad & t_i = \frac{\pi_i}{\tau_i} \\ st \quad & \sum_k \lambda_{ik} g_k \leq g_i, \\ & \sum_k \lambda_{ik} p_k = \pi_i, \\ & \theta_i \sum_k \lambda_{ik} y_k \geq \tau_i, \\ & \sum_k \lambda_{ik} = 1, \\ & \tau_i = \pi_i^{-\frac{\gamma}{1-\gamma}} \exp\left(\mathbf{X}_i \boldsymbol{\beta} / \left(1-\gamma\right)\right). \end{aligned} \tag{9}$$

In this program, the number of pending cases π_i and the number of completed cases τ_i are decision variables and are chosen in order to minimize the average completion time of the court $t_i = \pi_i/\tau_i$. The last constraint in this optimization program is the demand constraint as explained at the beginning of this section. Since the level of efficiency θ_i is kept constant at the level determined with program (8), the completion time as determined by this program is an equilibrium completion time for the given level of efficiency θ_i. In other words, this program returns the *equilibrium* average completion time for the efficiency level θ_i as opposed to the *observed* non-equilibrium level $t = p_i/y_i$.

In some extreme cases, if the court is operating in the congestion part of the production frontier, the number of pending cases becomes unbounded and the equilibrium is $\tau = 0; \pi = max$. To detect these cases simply check that $inc > y$ and $p > \pi$. In all other cases the equilibrium time is well defined and the system eventually converges on the optimal time. This optimal time is lower or higher than the observed time depending on the starting point and the shape of the estimated production frontier. For courts that are heading towards full congestion the only solution is to increase resources (the number of judges) or increase production efficiency. This is discussed in the Results section.

The above optimization program is non-linear and can be shown relatively easily to be convex; by replacing the demand equation in the third constraint and into the objective function, one obtains:

$$\begin{aligned}
\min_{\lambda_{ik},\pi_i} \quad & \pi_i \\
st \quad & \sum_k \lambda_{ik} g_k \leq g_i, \\
& \sum_k \lambda_{ik} p_k = \pi_i, \qquad (10)\\
& \theta_i \sum_k \lambda_{ik} y_k \geq \pi_i^{-\frac{\gamma}{1-\gamma}} \exp\left(\mathbf{X_i}\boldsymbol{\beta}/\left(1-\gamma\right)\right), \\
& \sum_k \lambda_{ik} = S, \\
& S = 1. \qquad (11)
\end{aligned}$$

It can be shown that the function on the right hand side of the third constraint is convex, thus the program can be solved using standard convex optimization solvers.[17] We substituted the objective function

$$t\left(\theta_i, S=1\right) = \frac{\pi_i^{\frac{1}{1-\gamma}}}{\exp\left(\mathbf{X}_i\boldsymbol{\beta}/\left(1-\gamma\right)\right)},$$

with π_i because completion time monotonically increases in π_i for $0 \leq \gamma < 1$ (therefore this has no effect on the optimal solution). We call the *equilibrium completion time* $t\left(\theta_i, S=1\right)$ to emphasize that it depends on the current level of efficiency of production. The role of the variable S (which is constrained to be equal to one) becomes clear in the next paragraph.

Full efficiency We are now in a position to ask what happens to the equilibrium completion time should we introduce best practices and increase

[17]We use CVX in Matlab, a package for specifying and solving convex programs; CVX Research, Inc. CVX: Matlab software for disciplined convex programming, version 2.0. http://cvxr.com/cvx, April 2011. M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs, Recent Advances in Learning and Control (a tribute to M. Vidyasagar), V. Blondel, S. Boyd, and H. Kimura, editors, pages 95-110, Lecture Notes in Control and Information Sciences, Springer, 2008. http://stanford.edu/~boyd/graph_dcp.html.

the level of efficiency to $\theta_i = 1$. This involves solving the same optimization program (10) with the full efficiency values rather than the observed efficiency values. The comparison of the equilibrium completion time at full efficiency vs. the equilibrium completion time at a given level of efficiency produces an equilibrium measure of the efficiency of the court:

$$TE_i = \frac{t\left(\theta_i = 1, S = 1\right)}{t\left(\theta_i, S = 1\right)}.$$

It should be noted that this concept of efficiency is an equilibrium notion, since it includes a constraint that takes into account the behavior of demand when the average completion time changes. On the contrary, θ_i is a measure of optimality without regard to the level of demand for the service, i.e., it is potential rather than something that can be realized, and should be used to assess if a court is on the frontier or in the interior of the set. It should also be noted that in general the observed time (at the given level of efficiency θ_i) may not comply with our equilibrium concept. Therefore it is the equilibrium efficiency that prevails after all adjustments for the number of pending cases are made and the steady state conditions hold. In other words our measure of time efficiency compares two alternative steady states: one with the observed level of inefficiency and the other with the court lying at the frontier.

The optimal size of courts Technical inefficiency is far from the only component of inefficiency in the system. Two further kinds are explored below and in the following section: inefficiencies from scale economies and inefficiency from the non-optimal allocation of judges to the various courts.

The most convincing argument in terms of the size of courts is suggested by looking at the scatter plot below (Figure 4), showing the number of processed cases and judges. This is consistent with the framework of Fare and Svenson (1980) which considers the maximum capacity of each court. Scale economies are then derived from the production function with the variable factor at the optimal proportion. Clearly, in the scatter, smaller courts are better able to process cases than larger courts. In fact, production is fairly proportional to the number of judges up to a number of 50-100 judges and then declines sharply. This means that one way to increase efficiency in production is to split larger courts into smaller units. For example, the largest court is obs=118 (Rome) with more than 400 judges: by implementing a break-up policy into subunits the production of this court could increase dramatically (indeed our results indicate output more than doubling for the Rome court).

To calculate into how many units a court should be split, we consider the

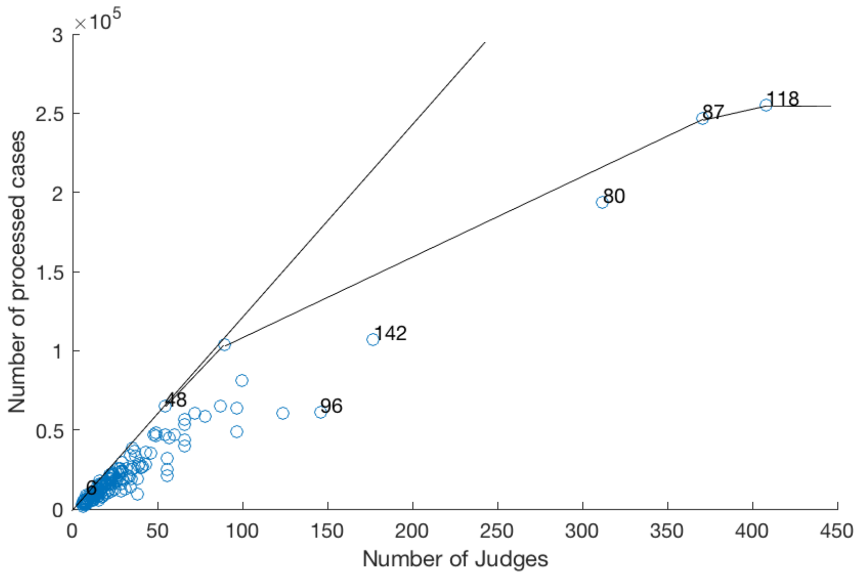


Figure 4: Size and efficiency

following capacity model (based on Maindiratta, 1990; see also Peyrache and Zago, 2016 for a discussion of this effect in the justice sector):

$$\begin{aligned}
\min \quad & \beta_i \\
st \quad & \sum_k \lambda_{ik} g_k \leq g_i, \\
& \sum_k \lambda_{ik} p_k = p_i, \\
& \sum_k \lambda_{ik} y_k \geq y_i/\beta, \\
& \sum_k \lambda_{ik} = S_i, \\
& S_i \in \{1, 2, \ldots\}.
\end{aligned}$$

The optimal values from this program can be used to determine the equilibrium completion time for each court in the scenario of an efficient break-up of large units into an optimal number of smaller units. The optimal values of the integer variable S_i^* provides the optimal number of sub-units into which large courts should be split. Given these optimal values, we can assess the steady state equilibrium completion time for each court by using these values in program (10). Since this is a relaxation of the original problem, the completion time is shorter or equal to the previous optimum; therefore:

$$SE_i = \frac{t\left(\theta_i, S_i^*\right)}{t\left(\theta_i, S_i = 1\right)}.$$

This is the optimum equilibrium size of the court, accounting for the effect of the demand for justice and keeping the level of efficiency at the current level. Best practices could be introduced as courts are split up into smaller units, producing a dual benefit, as follows:

$$SE_i = \frac{t\left(\theta_i = 1, S_i^*\right)}{t\left(\theta_i, S_i = 1\right)}.$$

Reallocation efficiency Another supply policy we consider is the reallocation of judges to courts. This may enhance the efficiency of the system by moving judges from courts which have very fast processing times to courts with much slower processing times; or, similarly, it could consider gains obtainable when the number of judges is found not to constrain production, i.e., there is excess capacity. The reallocation problem for the system can be written in the following way, which can be solved in a single step for all courts:

$$\begin{aligned}
\min_{\lambda_{ik}, \pi_i, \mu_i} \quad & \sum_i \pi_i \\
st \quad & \sum_k \lambda_{ik} g_k \leq \mu_i, \ \forall i \\
& \sum_k \lambda_{ik} p_k = \pi_i, \ \ \forall i \\
& \theta_i \sum_k \lambda_{ik} y_k \geq \pi_i^{-\frac{\gamma}{1-\gamma}} \exp\left(\mathbf{X}_i \boldsymbol{\beta} / \left(1-\gamma\right)\right), \ \forall i \\
& \sum_k \lambda_{ik} = S_i, \ \forall i \\
& \sum_i \mu_i \leq \sum_i g_i, \\
& t_i = \frac{1}{\exp\left(\mathbf{X}_i \boldsymbol{\beta} / \left(1-\gamma\right)\right)} \pi_i^{\frac{1}{1-\gamma}} \leq t_{max}.
\end{aligned}$$

This problem looks at the minimization of the system completion time, given demand constraints for each court, by reallocating judges to the courts. The optimal value of the decision variables μ_{it} provides the optimal reallocation of judges in order to minimize the average completion time of the system.

It should be noted that by setting $\mu_i = g_i$, one obtains the individual court efficiency problem (10) in stacked form (and with identical solution). Therefore the reallocation of judges cannot worsen the current configuration. In this program two parameters can be changed: θ_i and S_i. These parameters are used to determine alternative supply policy scenarios in terms of best practices and break-up policies. In particular it is possible to look at 4 combinations: current efficiency θ_i and no break-ups $S_i = 1$; current efficiency with break-ups S_i; full efficiency without break-ups and full efficiency with break-ups.

4 Data

We consider courts of first instance (*Tribunale Ordinario*), which have juridisction over civil and criminal cases. Generally presided over by one judge, for important cases a panel of three judges presides. Their decisions can be appealed at the *Corte d'Appello* (for reasons of substance, i.e., concerning facts giving rise to the case) or at the *Corte di Cassazione* (i.e., for reasons concerning legitimacy issues or similar). We referred to a panel of 165 courts (the Italian court population) for the years 2005-2012.

The following measures were used for inputs and outputs: for outputs the total number of cases completed in a given year; for inputs, the number of judges and the number of pending cases at the beginning of the year. Using pending cases as an input was first suggested by Lewin et al. (1982), and can be defended on common sense grounds: without pending cases, there are no processed cases and therefore no output. In general, in any system, there is a percentage of pending cases, and these cases can be interpreted as an intermediate input stock (raw material inventory or working capital).

Table 6 provides descriptive statistics for the inputs and output available for the pooled sample of 165 courts over the period 2005-2012 (a total of 1,320 observations). On average, an Italian court completes almost 24,000 cases per year, with quite a wide range between courts (the minimum is fewer than 2,000 cases, the maximum is above 250,000 cases). On average a court has 34 judges, with a range from 6 to above 400. The stock of pending cases at the begining of the year is around 28,000 cases, from a minimum of fewer than 2,000 cases to almost 280,000 cases over the period considered. As a very rough measure, because we consider the average for all the courts over the years, completion time is about 14 months.[18]

[18] Completion time (in years) is calculated as $\frac{pending\ cases}{completed\ cases} = \frac{28{,}040}{23{,}850} = 1.17$, or about 14 months.

The conterfactual analysis that follows is based on the assumption that the system is in steady state (see, e.g., Eq. (3)). From Table 6 we can see that the difference between incoming cases and completed cases is $< 0.3\%$ overall for the period 2005-2012 in the aggregate and for each individual year (apart from 2006, where the difference is $< 2\%$). The stock of pending cases is quite stable over time: compared to the overall average for the period 2005-2012, it changes less than 2% in 2005 and 2006 and even less ($< 1\%$) for the following years.

5 Empirical results

We now set out the results of the estimations and the conterfactual analyses, starting with the estimate of demand, i.e., the models in Eq. (4) and (5). We then discuss the differing types of courts in relation to their equilibrium, followed by the counterfactual analysis of the different policies. Lastly, we look at the geographical distribution of some of the policy scenarios.

5.1 Demand estimation

We estimated a number of different specifications of the demand equation and show the results in Table 1. For the demand equation based on *adaptive expectations*, we use *lagged* completion time to assess the elasticity of demand and population as a proxy for the size of demand. We consider OLS, Fixed Effects, Random Effects and Fixed Effects with time dummy variables. As can be inferred from the table, both the Random Effects and OLS models are rejected in this specification. The estimated elasticity of demand is -0.159 for the Fixed Effects model and -0.119 for the Fixed Effects model that includes time dummy variables. We consider the Fixed Effects model to be the best specification in this context since it enables to control for unobserved heterogeneity and at the same time is more parsimonius than the model including time dummies. Both the OLS and Random Effects models are excluded because they produce inconsistent estimates of demand elasticity.

Table 1: Demand estimation results

	1 (FE)	2 (FE1)	3 (OLS)	4 (RE)	5 (2SLS)	6 (Year FE)	7 (Court FE)	8 (Both FE)
Population	2.125**	0.98**	1.021**	1.023**	1.022**	1.021**	1.871**	1.252**
	(0.264)	(0.348)	(0.009)	(0.022)	(0.009)	(0.009)	(0.217)	(0.299)
Lagged compl. time	−0.159**	−0.119**	0.275**	−0.007	IV	IV	IV	IV
	(0.029)	(0.029)	(0.018)	(0.025)				
Current compl. time	–	–	–	–	0.29**	0.293**	−0.379**	−0.286**
					(0.02)	(0.019)	(0.057)	(0.058)
Constant	−16.773**	−2.62	−3.07**	−3.078**	−3.083**	−3.14**	−12.802**	−6.001+
	(3.278)	(4.326)	(0.115)	(0.268)	(0.117)	(0.117)	(2.405)	(3.305)
Year FE	–	Yes	–	–	–	Yes	–	Yes
Court FE	Yes	Yes	–	–	–	–	Yes	Yes
Adj. R-squared	0.90	0.89	0.93	0.91	0.92	0.93	0.99	0.99
No. observaations	990	990	990	990	990	990	990	990

** = 1% s.l.
* = 5% s.l.
+ = 10% s.l.

The last four models in the table consider the specification under *rational expectations* where the quantity demanded is a function of the *current* processing time, rather than the lagged processing time. We estimated such models by using the lagged completion time as an instrument for the current completion time. The 2SLS estimator is shown as well as the same estimator using court and time dummies. As for the previous set of regression models, the models that do not include court effects (i.e., unobserved heterogeneity) produce inconsistent estimates. The last two models include court fixed effects and the associated estimated elasticity is a consistent estimator of the population quantity. The estimated elasticity for model (7) is -0.379, while the same elasticity in model (8) which includes time dummies is -0.286. These estimates possess a slightly larger elasticity of demand, although still a quite a rigid response of demand for justice to a change in the completion time. In terms of the trade-offs between parsimony and data fitting, we consider the fixed effects model with adaptive expectation to be the best of the 8 estimated models. Therefore we use an elasticity of demand of -0.159 for the subsequent analysis.

5.2 Types of equilibrium courts

Table 2 classifies courts based on the discussion arising from Figure 3, describing three types of courts: type 1 converges on the equilibrium steady state from the left, i.e., left-converging type (e.g., point c_a in Figure 3); type 2 converges on the steady state equilibrium from the right (right-converging, as in point c_b); and type 3 (point c_c) diverges from the steady state towards a fully congested outcome, despite the steady state equilibrium for these courts. Type 3 might be efficient courts, but are congested: their main problem is that they are unlikely to converge on the steady state without a policy 'shock', e.g., more judges.

The table shows the frequency of the three types of courts in the population. Type 1 represents 51.5% of the total population, has an average completion time of 10 months and in equilibrium would converge on a processing time of 11.4 months. These courts are mostly located in the North and Center of Italy. Type 2 represents 43.6% of the population, has an observed completion time of 17.7 months, and it right-converges on a steady state equilibrium completion time of 14.6 months; these courts are mostly located in the South. Lastly, type 3 courts represent only 4.9% of the population; the observed average processing time is 17.4 months and the equilibrium completion time would be 8.9 months, though, as explained below, without further intervention these 8 courts would fail to reach such a rapid processing time. As can be seen in Figure 5, these courts are evely distributed across

Table 2: Average completion times (in months) by type of courts

Type	Observed time	Equilibrium time	Frequency
1 - Left-converging	10.1	11.4	85 (51%)
2 - Right-converging	17.7	14.6	72 (44%)
3 - Diverging	17.4	8.9	8 (5%)

Italy.

5.3 Counterfactual analysis of supply policies

Four sets of policies - a) the introduction of best practices, b) the break-up of large courts, c) the reallocation of judges, and d) the increase in the number of judges - plus their combinations, can be considered tools in the supply side the effects of which can be calculated with our model. Some of these policies have now been implemented but after the period examined here.

Apart from the Pinto law of 2001 establishing damages for lenghty cases, most of the measures until 2012 were designed to reduce case inflow. They include an increase in court fees to avoid excessive recourse to courts, the reduced possibilities to appeal court decisions, the introduction of alternative dispute resolution (ADR) mechanisms, and so on (Esposito et al., 2014). Further changes have been proposed, such as backlog-reducing teams and other measures previously introduced in pilot schemes,[19] to be extended to courts throughout Italy. But overall the need for better court management to handle cases more actively, with data systems and performance accountability, is recognized (Esposito et al., 2014: 13).

One example of good case management practice, often cited, is the Strasbourg method adopted first by the Turin court and subsequently extended elsewhere (Caponi, 2016). This method is based on the active leadership role of the President of the court, making each judge responsible for reach-

[19] "Since 2004, the EU supported a roll-out of the Turin and Bozen courts' experience to the entire country (Program Title: Diffusion of best practices in the Italian Judicial Offices). This program made some progress (e.g. for the Milan Court). However, the program faced implementation constraints as well as jurisdictional issues between regional and central authorities. The central government has taken a stronger role in program management since 2010-2011, with the Ministry of Public Administration setting up an effective central monitoring system in 2011 and the Ministry of Justice putting in place professional management in 2012. This helped secure the EU structural funds" Esposito et al., 2014: 10).

ing clear and transparent objectives, to be monitored actively, and changing case management from last in - first out (LIFO) to first in - first out (FIFO) (Abravanel et al., 2015). The *Consiglio Superiore della Magistratura* (CSM), the self-governing judicial body, has recently started to introduce best practices, such as strategy and planning using a formal ‘organizational document’, working trials in sequence and not in parallel and the intensive introduction of IT technologies, etc.[20] Effective practices include case-flow management and the production of statistics, areas in which Italy has been lagging behind other OECD countries (see, e.g., Palumbo et al., 2013: 34-35), certainly in the years under consideration.

Another possible policy is the re-design of court geography (Bartolomeo, 2013). The geography of Italian courts was originally designed after Italian unification in 1865, and underwent a number of changes during the fascist period, after World War II and in the late nineties. In the early nineties, the CSM suggested the need to break up large courts such as Rome, Naples, Milan and Turin (CSM, 2010).[21] However, most of the literature on Italian court efficiency has highlighted increasing returns to scale and thus the need to merge small courts,[22] leading to similar policy suggestions offered by the Ministry of Finance,[23] eventually implemented by the Monti Government reducing the number of courts by about 20%.[24]

The increase in judges has been more difficult to achieve in recent years, because of the poor state of public finances in Italy and overstaffing compared to other OECD countries (Palumbo et al., 2013). This, together with the break-up of large courts, makes a total of 32 possible policy scenarios. For the sake of clarity, we first consider the scenarios with a fixed number of judges. For each scenario, illustrated in Table 3, we set out both the overall average equilibrium completion time and the distribution of completion times

[20]For a more detailed explanation see, e.g., www.csm.it/web/csm-internet/il-progetto-buone-prassi/il-fenomeno-buone-prassi.

[21]The CSM “suggested a split of their structures on a territorial basis, dividing their district in two or three parts with corresponding court and district attorney for each of them” (CSM, 2010: 4).

[22]A notable exception is represented by Peyrache and Zago (2016): using Italian court data for the period 2003-08, they find that the breaking up of large courts could reduce aggregate inefficiency by 22%.

[23]The Ministry of Finance estimated the elasticity of scale of Italian courts using 2006 data, finding that about 85% of courts were too small and confirming earlier results for 1996 and 2001 set out by Marchesi (2003; 2008). Therefore, the policy recommendation was “... to revise judiciary geography, by merging the smaller courts in order to realize economies of scale and specialization ...” (CTFP, 2008: 46).

[24]With Legislative Decree 155 dated 7 September 2012, the Monti Goverment merged 26 small courts (out of 165 at the national level) into larger, adjacent courts, taking effect in in 2013.

Table 3: Possible supply policy scenarios

	No Break-ups		Break-ups	
	Current efficiency	Full efficiency	Current efficiency	Full efficiency
Without reallocation of judges	1	2	3	4
With reallocation of judges	5	6	7	8

for the system.

5.3.1 With existing judges

Table 4 shows the current average processing time for the system as a whole (calculated as the total number of pending cases in the system over the total number of completed cases, i.e., the weighted average of the observed completion times of individual courts) and the average processing time associated with different policy scenarios. Note that the *current* processing time for the system is 14.1 months, while the *equilibrium* completion time - after the system adjusts to the steady state, see program of Eq. (9) - is 13 months. These figures are a weighted average for the whole system; a boxplot illustrates processing times in different courts, as shown in Figure 6a. From this figure it is clear that the equilibrium completion time for some of the slowest processing courts diminishes as it moves towards equilibrium. The problem with many of the slow processing courts is that they do not necessarily converge on the steady state, as shown above in Table 2.

Table 4 illustrates the equilibrium outcome of the different policy scenarios.[25] Overall, the single most effective policy would be the introduction of best practices (scenario 2), leading to an average completion time of 8.6 months. This is followed by the break-up of courts (scenario 3, with a reduction to an average of 11.1 months), and finally by the reallocation of judges (scenario 5, with a reduction to 12 months if two policies were adopted).

[25]For example, a policy of breaking up large courts together with an optimal reallocation of judges (policy scenario 7) would reduce the processing time of the system from 13 to 9.4 months. This would require 'only' the optimal use of scale economies and the optimal reallocation of judges.

Table 4: Average completion time (in months) for policy scenarios

Observed processing time (14.1 mos)	No Break-ups		Break-ups	
	Current efficiency	Full efficiency	Current efficiency	Full efficiency
Without reallocation of judges	13.0	8.6	11.1	7.2
With reallocation of judges	12.0	7.3	9.4	6.3

If one adopted two sets of policies, the best combination would be the use of best practices either together with break-ups (scenario 4), leading to a further reduction to 7.2 months, or the reallocation of judges (scenario 6), leading to an average completion time of 7.3 months.

The boxplot in Figure 6b is useful to investigate the distribution of processing times. By looking at single policies, for instance, with the reallocation of judges (scenario 5) the processing time of the slowest court would be dramatically reduced (to less than 18 months) as against best practices (scenario 2), where the slowest court would take more than 24 months, or the break-up of courts (scenario 3), when a certain number of courts would still take more than 24 months on average to complete cases. The effect of scenario 5 would be quite substantial, considering that it would lead to the reduction of the processing time in the worst performing courts in the system.

To summarize, from Table 4 it is clear that the proper implementation of best practices has a major effect in reducing the overall average completion times in the system (see also the even plots in Figure 6b). The combination with other policies would add to this; in particular, the implementation of the three policy tools, that is introduction of best practices, reallocation of judges, and break-ups of the large courts, i.e., scenario 8, would bring the system completion time to 6.3 months and would drastically reduce the dispersion in the system, with only one court of justice taking more than 8 months to process a case. Implementing these three policies together may be challenging, but the impact would be rather substantial, taking the system to steady state completion times comparable to other developed countries.

5.3.2 The geographical distribution of policy effects

The effects of the long processing times of the courts on the real economy are fairly well understood. Recent work highlights, among others, the effects on the size of firms and their growth (see, e.g., Giacomelli and Menon, 2017 for Italy), public procurement and corruption (see, e.g., Coviello et al., 2018), the participation of firms into global value chains and thus on the overall competitiveness of the economy (see, e.g., Accetturo et al., 2017), and so on. Since these may have a more or less important role in different parts of the country and different weights for policy-makers, looking at the geographical distribution of the impact of the suggested supply policies may be informative.

The maps below represent the reduction in processing times in courts in the different policy scenarios of Table 3, i.e., with existing judges. Figure 7 illustrates the effects of introducing best practices and the break-up of courts in terms of *reducing* the time needed to complete a case compared to the equilibrium efficiency time. First, note that from the latter it appears evident that the equilibrium (and actual) processing times are in general longer in the South and in the Islands (Sicily and Sardinia; scenario 1, panel a).[26] A comparison with panel b shows that the introduction of best practices (scenario 2, Table 3) would have a major impact in the same regions and possibly on courts where processing times are longer, i.e., in Southern Italy. In contrast, introducing a break-up of courts would have a limited effect when taken in isolation (scenario 3, panel c), or when added to the adoption of best practices (scenario 4, panel d).

Figure 8 shows the geographical impact of different policy scenarios including the reallocation of judges. Solely the latter would have a lower impact than introducing best practices (shown respectively in panels 8a and 7b), but more than merely breaking up the courts (panel 7c). However, reallocating judges may worsen the situation for some courts (highlighted in grey in the maps), located both in the North and the South, by increasing processing times.[27] On the other hand, combining reallocation with break-ups (panel 8c), and especially with best practices (panel 8b), would reduce these negative effects and shorten processing times for almost all or all courts respectively.

Moreover, while breaking-up courts added to best practices seems to have a negligible effect (compare panels 7b and 8d), the reallocation of judges appears to have an albeit limited effect (compare panels 8b and 7b). Last, panel 8d shows the reduction in processing times using the three policies

[26] Panel 7a shows the equilibrium processing time for each court, and the other panels show the *reduction* of processing times in months.

[27] The number of judges would be reduced in these courts.

together. Note that, geographically, scenarios 6 and 8 are quite similar, showing that changing the court geography after introducing best practices and optimally allocating judges would have a negligible effect. To summarize, in terms of geographical distribution, introducing best practices would have the most significant impact, especially in courts located in the South and in the Islands of Italy, which are overall more inefficient and slow in processing cases. As an alternative, or in addition, the reallocation of judges may be a second best, working more effectively when combined with best practices.

5.4 Increasing the number of judges

We now consider the effects of an increase in the number of judges, in combination with other policy scenarios. Table 5 shows the average processing times of different policy combinations, gradually increasing the number of judges.

Adding judges to the current court system First, we consider scenario I, the increase in the number of judges without break-ups and without the adoption of best practices, i.e., at current efficiency levels, represented in column I of Table 5[28] and in Figure 9a.[29] As can be seen, increasing the number of judges has a significant impact on reducing processing times: for instance, by increasing the number of judges by 10%, the average processing time in the system would fall to about 10.4 months from the equilibrium status quo of about 13 months. By increasing judges by 25%, on the other hand, processing times would fall to about 9.2 months, while increasing judges by 50% would not further reduce processing times.[30]

An interesting comparison can be made between two policy alternatives, or their combination, i.e., the increase in the number of judges vs. the optimal reallocation of all judges (existing and newly hired). As can be seen in Table 5 - column I (by comparing the even and odd plots, e.g., 2 vs. 3, 4 vs. 5, and so on, in Figure 9a), a policy of reallocating existing judges would reduce the

[28] Note that average values in tables and box-plots are slightly different. For the tables we compute the system average processing times as a *weighted* average using pending cases for each court as weights, while the box-plots illustrate processing times alone and their simple means.

[29] In the first box-plot in Figure 9a the system equilibrium time is shown without policy interventions. The *even* box-plots illustrate the distribution of processing times when optimally allocating all judges, i.e., current and new; the *odd* box-plots illustrate adding judges only. The increase in new judges is 0% (plots 2-3), +10% (plots 4-5), +25% (plots 6-7), and +50% (plots 8-9).

[30] In fact, after about 29%, a further increase in the number of judges does not have any effect on reducing the system processing time (detailed results are available on request).

Table 5: Increasing judges - Average completion time (in months)

Observed proc. time (14.1 mos)	Current efficiency		Full Efficiency	
	NO break-up I	Break-up II	NO break-up III	Break-up IV
		$\mathbf{\Delta = +0\%}$		
New judges	13.0	11.1	8.6	7.2
Opt. reallocation	12.0	9.4	7.3	6.3
		$\mathbf{\Delta = +10\%}$		
New judges	10.4	8.8	7.3	6.3
Opt. reallocation	10.3	8.4	7.2	6.3
		$\mathbf{\Delta = +25\%}$		
New judges	9.2	8.1	7.2	6.3
Opt. reallocation	9.2	8.1	7.2	6.3
		$\mathbf{\Delta = +50\%}$		
New judges	9.2	8.1	7.2	6.3
Opt. reallocation	9.2	8.1	7.2	6.3

system processing time from 13 to about 12 months (second row in Table 5), while increasing the number of judges by 10% would reduce the processing time to 10.4 months (third row). Optimally reallocating existing judges, in addition to increasing their number and, in addition, optimally allocating new judges, would have no significant effect on average processing times (apart from reducing the processing time of the worst performing courts, as can be seen in column 5 of Figure 9a). Last, from Table 5 - column I, we can infer that the effect of the optimal reallocation of existing judges on processing times (with the associated monetary and political costs) has an effect which is equivalent to adding about 3-4% of new judges to the system.

Together with a new geography of courts A relatively similar picture emerges when we consider the second policy combination, i.e., optimal break-up of large courts but with current practices (Table 5 - column II and Figure 9b). First, note that the optimal (re)allocation of existing judges, combined with an optimal court geography, i.e., break-ups, would reduce processing times to about 9.4 months. Second, an almost equivalent effect, i.e., a reduction to 8.8 months, would be obtained by increasing the number of judges by about 10%, even though reallocating existing judges as well would in this case help to further reduce processing times to 8.4 months overall. However, increasing judges from 10 to 25% would additionally reduce processing times only modestly, to 8.1 months on average, and there would be no effect beyond 25%. Last, reallocating existing judges when increasing their total number by 25% or above would have no effect on processing times, either on the overall average (column II of Table 5) or on the worst performing courts (Figure 9b).

Together with the adoption of best practices The last set of policy combinations considered is the increase in the number of judges combined with their optimal (re)allocation and/or the introduction of best practices. As stated above, introducing best practices would have a major effect: processing times would be reduced to about 8.6 months with the full implementation of best practices, and further to 7.2 months when combined with a re-design of the geography of courts (see Table 5 - columns III and IV respectively, and Figure 10, top and bottom panels). The effects of adding more judges to courts implementing best practices would be limited, i.e., a reduction of about one month by increasing the number of judges by 10% (no further effects beyond this increase). Moreover, adding new judges to courts already at optimal scale and adopting best practices would have no effect on their processing times.

Another way of looking at the results in Table 5 is to calculate the 'opportunity costs' of different policy options, in a back of the envelope sort of calculation. We have seen that the most effective policy to reduce processing times would be the full implementation of best practices: to obtain similar results - actually, slightly less effective - the number of judges would need to be increased by 25% (better still if combined with an optimum court geography), with the associated costs.[31]

Reducing processing times further by about one month in a system adopting best practices, could be achieved either by optimally breaking-up courts, or by optimally reallocating existing judges, or by increasing the number of judges by 10%. The cost of the latter can be calculated without difficulty. If it were not possible to adopt best practices, an alternative policy could be to re-design court geography combined with reallocating judges, which would reduce processing times to about 9.4 months. Alternatively, to obtain a similar result the number of judges would need to be increased by about 25%.

Adding new judges The final hypothesis concerns adding new judges throughout Italy. As can be seem in Figure 11, both in the case of an overall increase of the total number of judges of 10% and of 25%, the entire country would be involved. No preference would be made for inefficient regions in the South, nor for wealthy regions such as Lombardy, the Veneto, Emilia-Romagna or Puglia.

6 Concluding remarks

This paper reviews the relevant literature on the efficiency of the justice system. Using a rather general production model, we consider the resources used in each court allowing for possible congestion effects. This extends the standard practice of considering only processing times, and leads to an efficiency analysis investigating the sources of inefficiency (for each court) expressed in the duration of trials. A model is put forward to assess the equilibrium efficiency of the justice market and applied to Italy, a country suffering badly from inefficient courts. In such a setting, the market clearing function is performed by the time to complete a case, i.e., to deliver justice.

[31] Considering a total number of judges of about 5,650, this would correspond to about 565 new judges. With an initial cost of 70,000 euros per new judge (Senato, 2017), this would be an approximate cost of 40 million euros for the first year. With an increase of the total number of judges of 25%, the total cost would be about 100 million euros in the first year.

The shorter the time, the higher the demand for justice. The efficiency and counterfactual analyses take into consideration this feedback effect from the demand side of the market as a response to the shortening of processing times when certain supply policies are implemented.

We consider the impact on processing times of different supply policies based on the break-up of large courts, the reallocation of judges to courts, the increase in the number of judges, and the introduction of best practices. We look at the implementation of these policies in various combinations and show how the average processing times of the system (and their distribution) vary in these counterfactual analyses.

We find that the single most impactful policy would be the proper adoption of best practices by the courts, which could reduce the average overall time to complete a trial by about one third, from 13 to 8.6 months. An alternative policy would be to re-design court geography and reallocate the existing judges accordingly, leading to a reduction to about 9.4 months. Finally, with existing judges, combining the adoption of best practices with the break-up of courts and optimal reallocation of judges, the average completion time of the system would be halved, even accounting for the increased demand for justice resulting from faster courts. Increasing the number of judges would not contribute much further after implementation of the other policies.

While the costs of these three policies, taken individually or in combination, are difficult to ascertain, the cost of increasing the number of judges is easier to calculate. Thus the alternative policy of increasing the number of judges by about 25% (and their optimal allocation) would have comparable effects to the implementation of best practices or the combination of resizing the courts and reallocating judges, with a total initial cost of about 100 million euros per year. We conclude that these alternative policy scenarios would be sufficient to bring the system down to a processing time comparable to other OECD countries. The benefits of these policies would be substantial, as court inefficiencies account for a loss of about 1% in GDP (Draghi, 2011).

The paper does not consider two further issues. The first is how an alternative scenario, reducing demand, would impact on completion times. This has occurred in recent years, for instance through alternative dispute resolution (ADR) mechanisms introduced to reduce the use of the courts. Another issue not addressed is the transition towards the steady state equilibrium. For example, when the system converges from the observed completion time of 14.1 months to an equilibrium completion time of 13 months, we do not specify the timing of this adjustment. Future research could explore these transition dynamics.

References

Abravanel, R., S. Proverbio and F. Bartolomeo (2015), Misurare la performance dei tribunali, Presentation of march 26, 2015, Osservatorio per il Monitoraggio degli Effetti sull'Economia delle Riforme della Giustizia.

Accetturo, A., A. Linarello and A. Petrella (2017), Legal enforcement and Global Value Chains: Micro-evidence from Italian manufacturing firms, Questioni di economia e finanza (occasional papers) no. 397, Banca d'Italia.

Aldashev, G. (2009), 'Legal institutions, political economy, and development', *Oxford Review of Economic Policy* **25 (2)**, 257–270.

Bartolomeo, F. (2013), Linee guida sulla revisione della geografia giudiziaria per favorire le condizioni di accesso a un sistema giudiziario di qualità, Discussion paper CEPEJ-GT-QUAL(2013)2, CEPEJ.

Barzel, Y. (1974), 'A theory of rationing by waiting', *Journal of Law and Economics* **17**, 73–95.

Bianco, M. and G. Palumbo (2007), Italian Civil Justice's Inefficiencies: a Supply Side Explanation, Mimeo, Banca d'Italia.

Bray, R.L., D. Coviello, A. Ichino and N. Persico (2016), 'Multitasking, Multiarmed Bandits, and the Italian Judiciary', *Manufacturing and Service Operations Management* **18(4)**, 545–558.

Buonanno, P. and M. M. Galizzi (2014), 'Advocatus, et non Latro?: Testing the Excess of Litigation in the Italian Courts of Justice', *Review of Law and Economics* **10(3)**, 285–322.

Buscaglia, E. and M. Dakolias (1999), Comparative International Study of Court Performance Indicators, Washington, D.C., The World Bank.

Caponi, R. (2016), 'The performance of the italian civil justice system: An empirical assessment', *The Italian Law Journal* **02(1)**, 15–31.

Carmignani, A. and S. Giacomelli (2009), La giustizia civile in italia: i divari territoriali, Questioni di economia e finanza (occasional papers) no. 40, Banca d'Italia.

Carmignani, A. and S. Giacomelli (2010), Too many lawyers? Litigation in Italian civil courts, Temi di discussione (working papers) no. 745, Banca d'Italia.

CEPEJ (2016), European judicial systems efficiency and quality of justices (edition 2016), Cepej studies no. 23, European Commission for the Efficiency of Justice, Strasbourg.

Chase, Oscar G. (1988), 'Civil litigation delay in Italy and the United States', *The American Journal of Comparative Law* **36(1)**, 41–87.

Chemin, M. (2009), 'The impact of the judiciary on entrepreneurship: Evaluation of Pakistan's 'Access to Justice Programme'', *Journal of Public Economics* **93**, 114–125.

Chemin, M. (2012), 'Does court speed shape economic activity? evidence from a court reform in india', *Journal of Law, Economics, and Organisation* **28(3)**, 460–485.

Cooter, R. D. and D. L. Rubinfeld (1989), 'Economic Analysis of Legal Disputes and Their Resolution', *Journal of Economic Literature* **27(3)**, 1067–1097.

Coviello, D., A. Ichino and N. Persico (2014), 'Time Allocation and Task Juggling', *American Economic Review* **104(2)**, 609–623.

Coviello, D., A. Ichino and N. Persico (2015), 'The Inefficiency of Worker Time Use', *Journal of the European Economic Association* **13(5)**, 906–947.

Coviello, D., L. Moretti, G. Spagnolo and P. Valbonesi (2018), 'Court Efficiency and Procurement Performance', *Scandinavian Journal of Economics* **120(3)**, 826–858.

CSM (2010), Risoluzione concernente la revisione delle circoscrizioni giudiziarie, Rome, Consiglio Superiore della Magistratura.
URL: *https://www.csm.it/documents/21768/86569/Risoluzione+del+13+gennaio+2010/50f9 67b9-48af-b9b8-6bd4b67d8b18*

CTFP (2008), La revisione della spesa pubblica. Rapporto 2008, Rome, Commissione Tecnica per la Finanza Pubblica.

Dimitrova-Grajzla, V., P. Grajzl, J. Sustersic and K. Zajc (2012), 'Court output, judicial staffing, and the demand for court services: Evidence from Slovenian courts of first instance', *International Review of Law and Economics* **32**, 19–29.

Djankov, S., R. La Porta, F. Lopez-deSilanes and A. Shleifer (2003), 'Courts', *Quarterly Journal of Economics* **118(2)**, 453–517.

Draghi, M. (2011), Considerazioni finali, Technical report, Banca d'Italia.

Esposito, G., S. Lanau and S. Pompe (2014), Judicial system reform in Italy — A key to growth, Imf working paper no. wp/14/32, International Monetary Fund.

Felli, E. L., D. A. London-Bedoya, N. Solferino and G. Tria (2008), The "Demand for Justice" in Italy: Civil Litigation and the Judicial System, *in* P.F. and R.Ricciuti, eds, 'Italian Institutional Reforms: A Public Choice Perspective', Springer, New York, NY, pp. 155–177.

Giacomelli, S. and C. Menon (2017), 'Does weak contract enforcement affect firm size? evidence from the neighbour's court', *Journal of Economic Geography* **17(6)**, 1251–1282.

Ginsburg, T. and G. Hoetker (2006), 'The unreluctant litigant? an empirical analysis Japan's turn to litigation', *The Journal of Legal Studies* **35(1)**, 31–59.

Gravelle, H. S. E. (1990), 'Rationing trials by waiting: Welfare implications', *International Review of Law and Economics* **10**, 255–270.

Ichino, A., M. Polo and E. Rettore (2003), 'Are Judges Biased by Labor Market Conditions?', *European Economic Review* **47(5)**, 913–944.

Jappelli, T., M. Pagano and M. Bianco (2005), 'Courts and Banks: Effects of Judicial Enforcement on Credit Markets', *Journal of Money, Credit and Banking* **37(2)**, 223–244.

Kumar, B., R. Rajan and L. Zingales (2001), What determines firm's size, Nber working paper 7208, NBER.

Landes, W. M. and R. A. Posner (1979), 'Adjudication as a private good', *The Journal of Legal Studies* **8(2)**, 235–284.

Lewin, A., R. Morey and T. Cook (1982), 'Evaluating the administrative efficiency of courts', *OMEGA International Journal of Management Science* **4**, 401–411.

Marchesi, D. (2003), *Litiganti, avvocati e magistrati. Diritto ed economia del processo civile*, Il Mulino, Bologna.

Marchesi, D. (2008), L'enforcement delle regole. Problemi di efficienza della giustizia civile, riforme intraprese e riforme possibili, I temi dei rapporti trimestrali - Roma, ISAE.

Mora-Sanguinetti, J. S. (2012), 'Is judicial inefficacy increasing the weight of the house property market in Spain? Evidence at the local level', *SERIEs* **3**, 339–365.

Mora-Sanguinetti, J. S. and N. Garoupa (2015), 'Do lawyers induce litigation? Evidence from Spain, 2001–2010', *International Review of Law and Economics* **44**, 29–41.

Nunn, N. (2007), 'Relationship-specificity, incomplete contracts and the pattern of trade', *The Quarterly Journal of Economics* **122(2)**, 569–600.

Padrini, F., D. Guerrera and D. Malvolti (2009), La congestione della giustizia civile in Italia: Cause ed implicazioni per il sistema economico, Note tematiche no. 08, Ministero dell'Economia e delle Finanze.

Palumbo, G., G. Giupponi, L. Nunziata and J.S. Mora Sanguinetti (2013), The economics of civil justice: New cross-country data and empirics, OECD Economics Department working papers no. 1060, OECD.

Peyrache, A. and A. Zago (2016), 'Large courts, small justice! The inefficiency and the optimal structure of the Italian justice sector', *Omega* **64**, 42–56.

Ponticelli, J. and L. Alencar (2016), 'Court Enforcement, Bank Loans, and Firm Investment: Evidence from a Bankruptcy Reform in Brazil', *Quarterly Journal of Economics* **131**, 1365–1413.

Senato (2017), Bilancio di previsione dello Stato per l'anno finanziario 2018 e bilancio pluriennale per il triennio 2018-2020, A.s. 2960, Senato della Repubblica.

Svensson, L. and R. Färe (1980), 'Congestion of production factors', *Econometrica* **48(7)**, 1745–1753.

Van Wijck, P. and B. VanVelthoven (2000), 'An economic analysis of the american and the continental rule for allocating legal costs', *European Journal of Law and Economics* **9(2)**, 115–125.

Table 6: Inputs and outputs (1,320 obs.)

Variables	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2005-12**
	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
	(St.dev.)	(St.dev.)	(St.dev.)	(St.dev.)	(St.dev.)	(St.dev.)	(St.dev.)	(St.dev.)	(St.dev.)
	min.	min.	min.	min.	min.	min.	min.	min.	min.
	[Max.]	[Max.]	[Max.]	[Max.]	[Max.]	[Max.]	[Max.]	[Max.]	[Max.]
Outputs									
Defined cases	23,072	22,275	23,402	24,159	24,763	24,461	24,041	24,025	23,850
	(33,829)	(32,675)	(33,668)	(34,434)	(34,733)	(34,269)	(33,885)	(32,650)	(33,838)
	1,626	1,567	1,556	2,023	1,538	1,969	1,908	1,806	1,807
	[272,041]	[254,732]	[250,495]	[256,390]	[260,104]	[259,571]	[264,576]	[255,645]	[254,959]
Inputs									
Judges	37	34	34	34	34	34	34	34	34
	(58)	(51)	(51)	(51)	(53)	(53)	(53)	(53)	(52)
	6	6	6	6	6	6	6	6	6
	[469]	[410]	[410]	[410]	[406]	[406]	[406]	[406]	[408]
Pending cases	27,641	27,616	28,047	27,976	28,131	28,206	28,262	28,029	28,040
	(40,772)	(40,051)	(39,988)	(40,326)	(40,053)	(39,847)	(39,092)	(38,199)	(39,740)
	1,666	1,979	1,894	1,712	1,954	1,970	1,847	1,753	1,894
	[284,209[	[275,040]	[274,464]	[275,042]	[282,851]	[281,646]	[271,278]	[262,097]	[276,720]
Others									
Incoming cases	23,048	22,705	23,332	24,314	24,839	24,517	23,808	–	23,919
	(33,168)	(32,538)	(33,611)	(34,298)	(34,983)	(33,675)	(33,178)	(–)	23,603
	1,763	1,716	1,530	1,962	1,550	1,710	1,625	–	1,782
	[247,909]	[244,876]	[251,855]	[264,199]	[261,775]	[252,840]	[265,809]	[–]	[255,280]
Population	356,071	358,371	361,329	363,910	365,699	367,433	359,965	361,729	362,784
	(391,277)	(397,742)	(400,020)	(401,768)	(404,330)	(406,983)	(394,693)	(397,813)	(400,897)
	21,351	21,196	21,068	20,932	20,789	20,692	20,630	20,525	20,884
	[2,563,120]	[2,705,603]	[2,718,768]	[2,724,347]	[2,743,796]	[2,761,477]	[2,614,263]	[2,638,842]	[2,711,376]

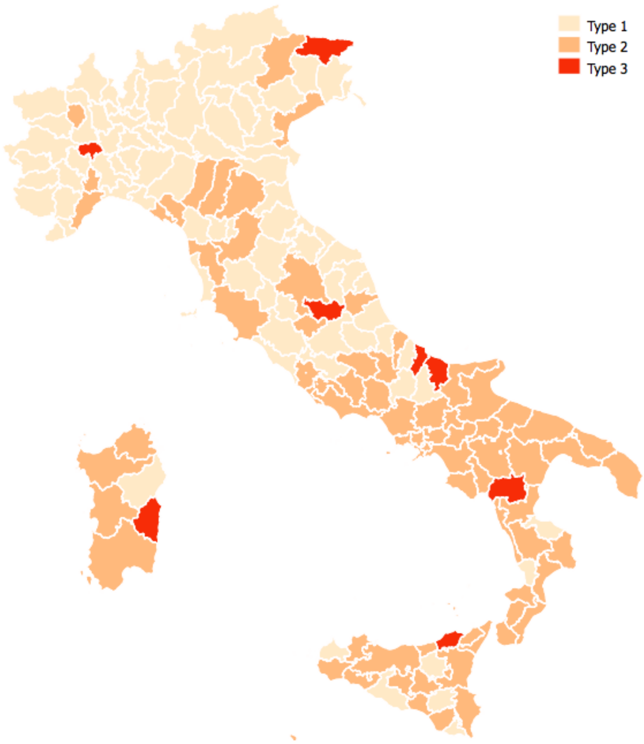
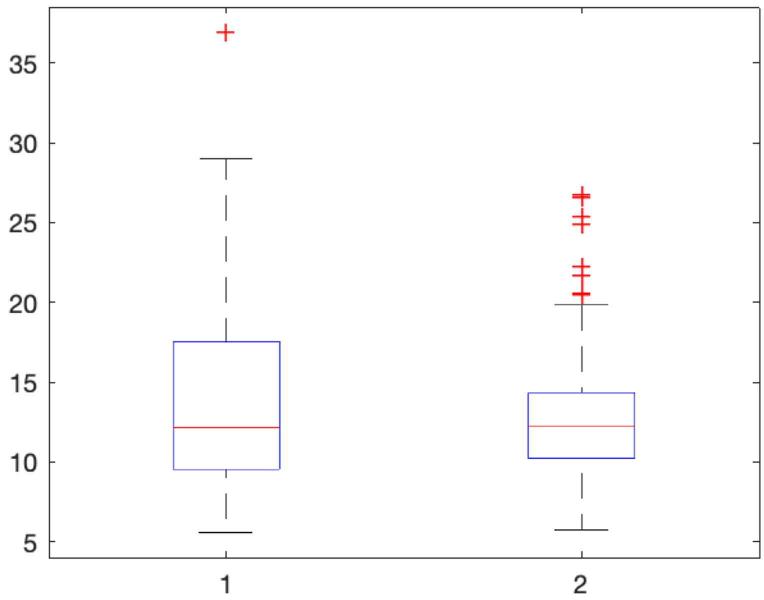
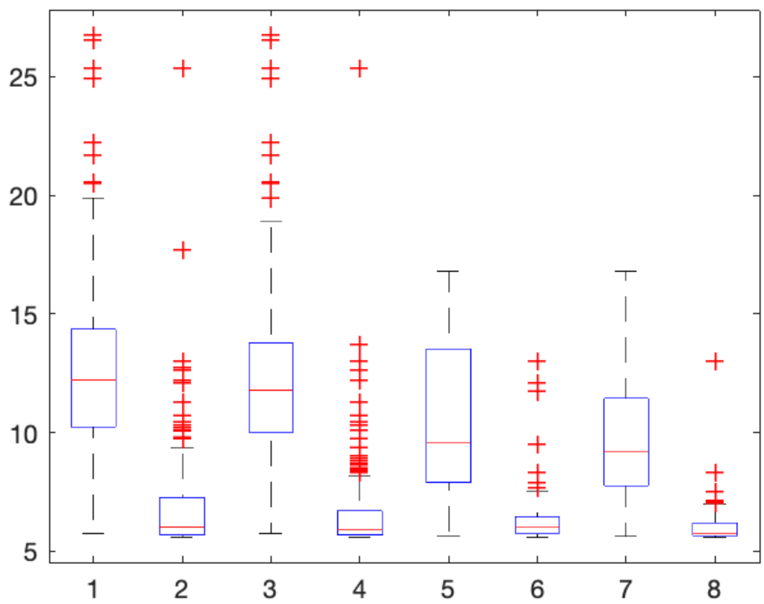


Figure 5: Court classification

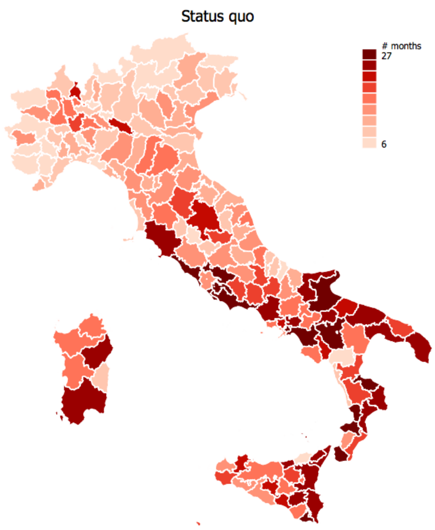


(a) Observed processing times (1) and equilibrium processing times (2)

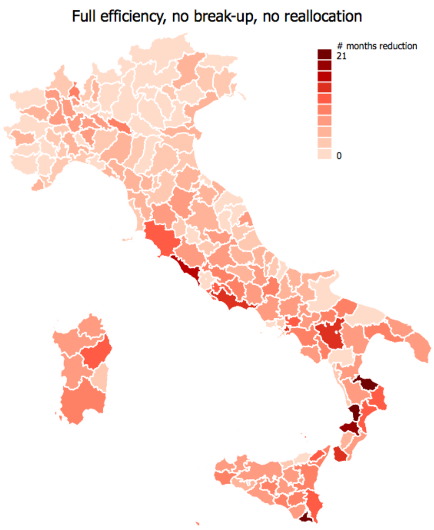


(b) Processing time distribution of courts under different policy scenarios

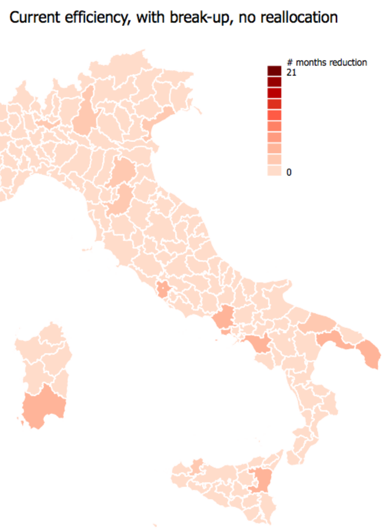
Figure 6: Boxplots of the processing time



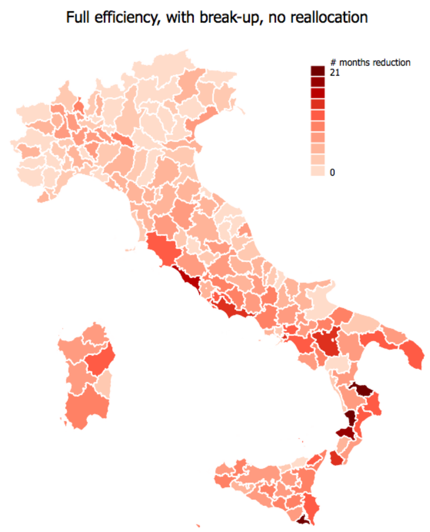
(a) Scenario 1



(b) Scenario 2

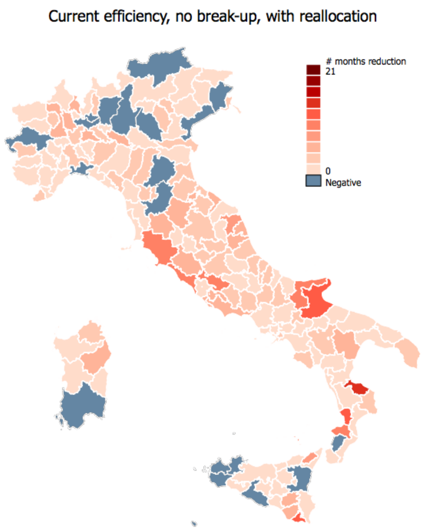


(c) Scenario 3

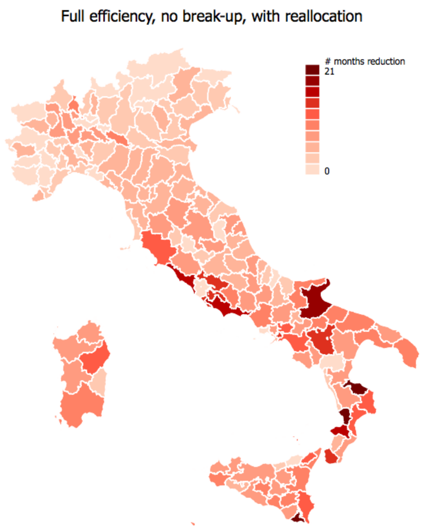


(d) Scenario 4

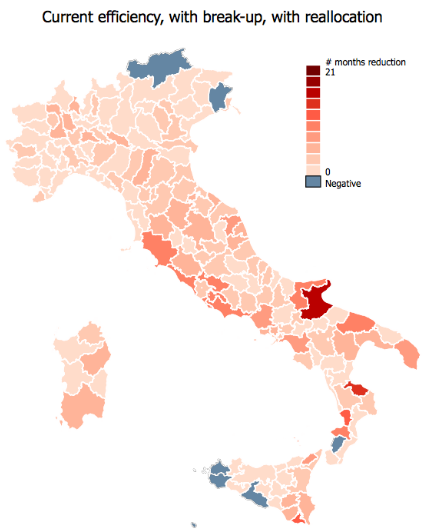
Figure 7: Policy scenarios - Without judge reallocation



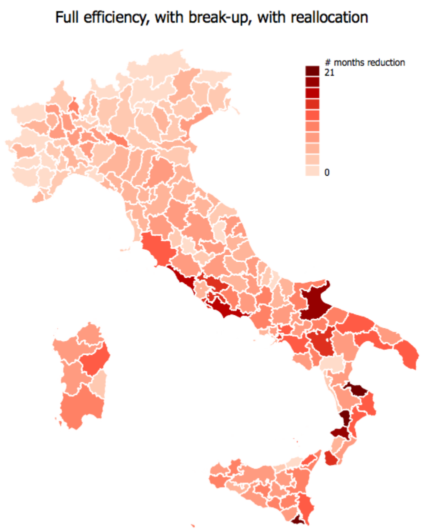
(a) Scenario 5



(b) Scenario 6

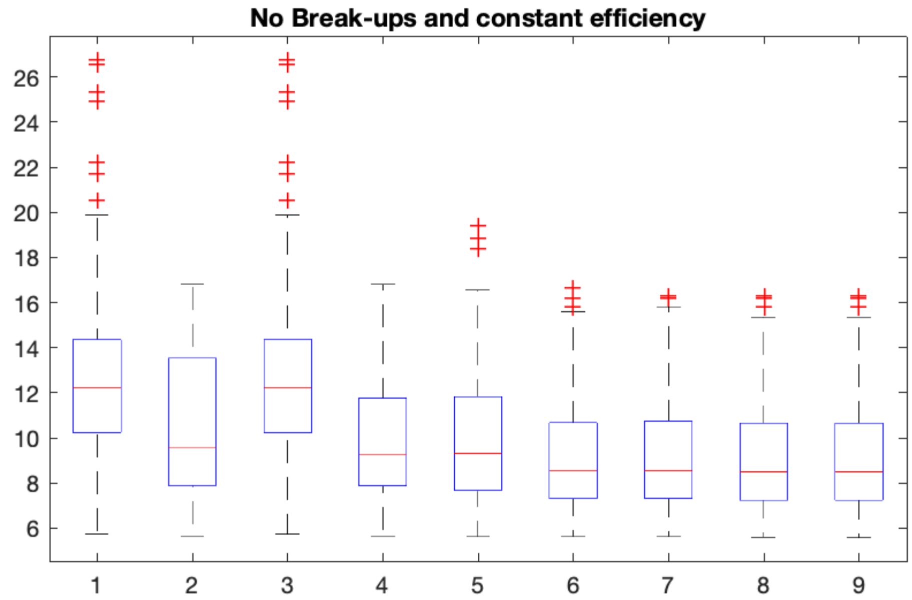


(c) Scenario 7



(d) Scenario 8

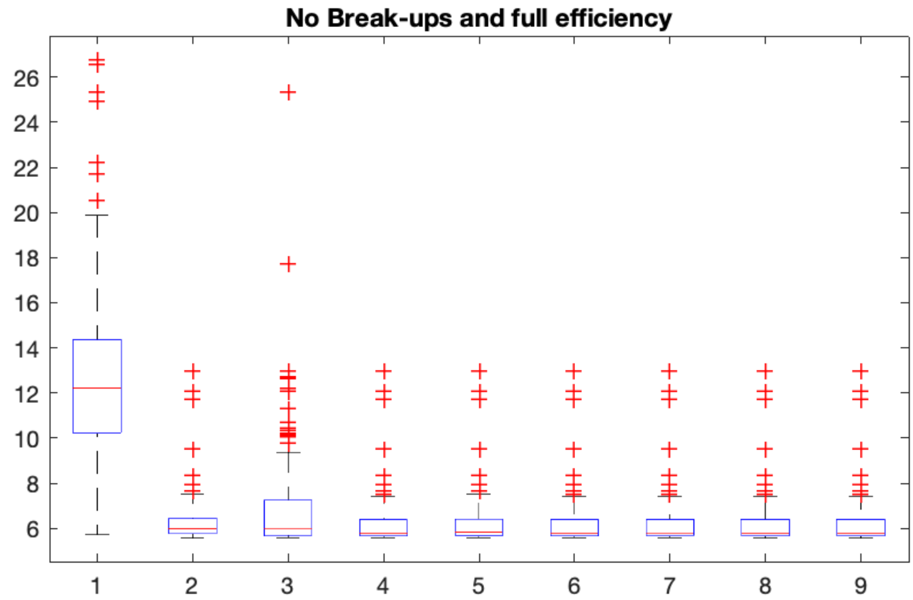
Figure 8: Policy scenarios - With judge reallocation



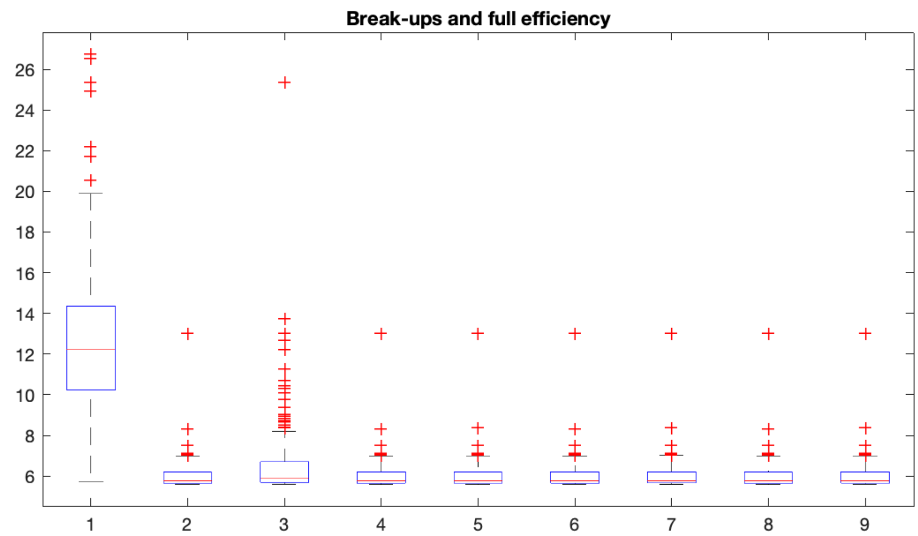
(a) Scenario I: No best practices, no break-ups

(b) Scenario II: With break-ups, no best practices

Figure 9: More judges, I

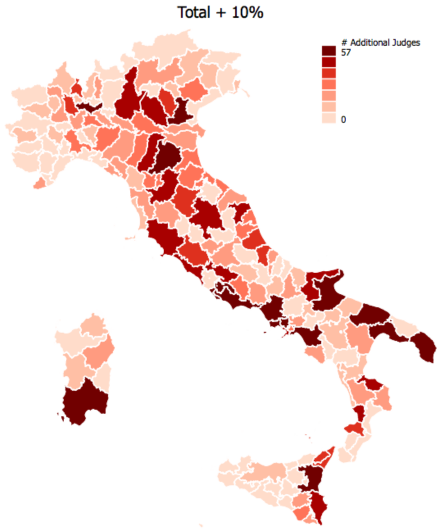


(a) Scenario III: Best practices, no break-ups

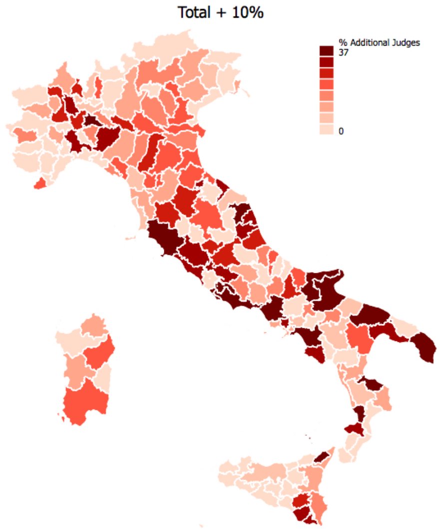


(b) Scenario IV: Best practices, with break-ups

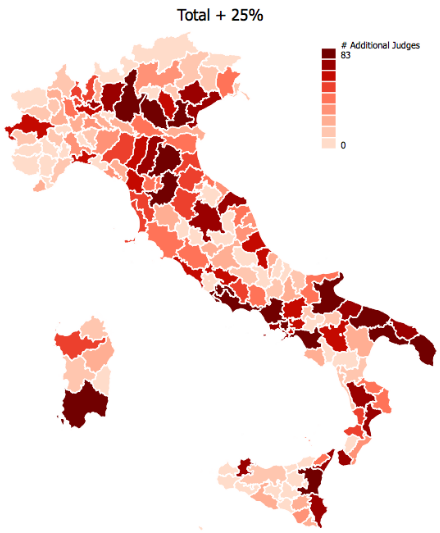
Figure 10: More judges, II



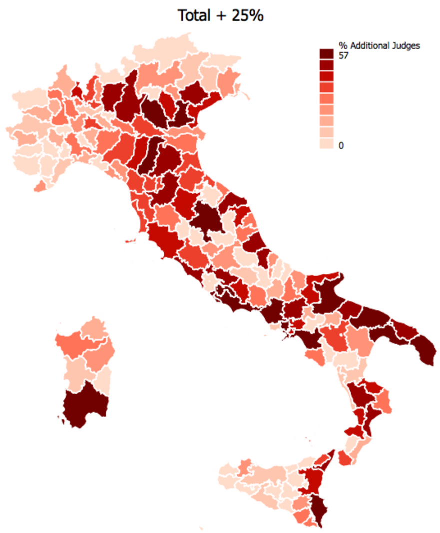
(a) #



(b) %



(c) #



(d) %

Figure 11: More judges, III