



# Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform<sup>☆</sup>

Anup Malani, Julian Reif<sup>\*</sup>

University of Chicago, United States

RFF, United States

NBER, United States

University of Illinois at Urbana–Champaign, United States

IGPA, United States



## ARTICLE INFO

### Article history:

Received 17 June 2013

Received in revised form 21 January 2015

Accepted 23 January 2015

Available online 31 January 2015

### JEL classification:

C50

K13

J20

### Keywords:

Anticipation

Medical malpractice

Endogeneity

Tort reform

## ABSTRACT

While conducting empirical work, researchers sometimes observe changes in outcomes before adoption of a new policy. The conventional diagnosis is that treatment is endogenous. This observation is also consistent, however, with anticipation effects that arise naturally out of many theoretical models. This paper illustrates that distinguishing endogeneity from anticipation matters greatly when estimating treatment effects. It provides a framework for comparing different methods for estimating anticipation effects and proposes a new set of instrumental variables to address the problem that subjects' expectations are unobservable. Finally, this paper examines a specific set of tort reforms that was not targeted at physicians but was likely anticipated by them. Interpreting pre-trends as evidence of anticipation increases the estimated effect of these reforms by a factor of two compared to a model that ignores anticipation.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

While conducting empirical work, researchers sometimes observe changes in outcomes before adoption of a new treatment program or policy. Fig. 1 provides an example from the medical malpractice liability context. It shows that equilibrium physician labor supply increased well before states adopted caps on punitive damages, which lower physician liability. The conventional diagnosis that researchers make upon observing such a pattern in the data is that the treatment was endogenous: states adopted these caps in response to the change in supply or for reasons correlated with supply (Angrist and Pischke, 2008, chapter 5).

Observing changes in outcomes prior to treatment is also consistent, however, with anticipation effects. Perhaps individuals began changing their behavior in response to an expectation that they would be treated

in the future. Anticipation is a reasonable diagnosis if individuals are forward looking, have access to information on future treatment, and there is a benefit to acting before treatment is adopted.<sup>1</sup>

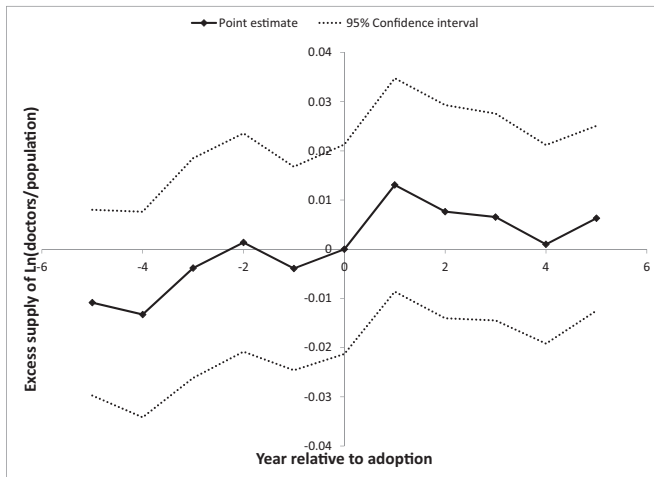
It is very difficult to rule out endogeneity as an explanation for the pre-trends such as those in Fig. 1, although previous studies have argued that the adoption of punitive damage caps was an exogenous event in the medical malpractice context (Avraham, 2007; Currie and MacLeod, 2008).<sup>2</sup> It may be equally difficult, however, to rule out anticipation as an explanation for pre-trends. For example, we present evidence that these reforms were discussed in newspapers years prior to their

<sup>☆</sup> We thank Dan Black, Amitabh Chandra, Tatyana Deryugina, Steve Levitt, Jens Ludwig, Derek Neal, Seth Seabury, Heidi Williams, participants at workshops and conferences at IGPA, Harvard University, New York University, the Searle Center at Northwestern University, and the University of Chicago, and two anonymous referees for helpful comments. Anup Malani thanks the Samuel J. Kersten Faculty Fund at the University of Chicago for funding. Julian Reif thanks the National Science Foundation for financial support.

<sup>\*</sup> Corresponding author.  
E-mail address: [jreif@illinois.edu](mailto:jreif@illinois.edu) (J. Reif).

<sup>1</sup> This paper is motivated by a case where there are differential pre-trends, i.e., different pre-trends in units that are treated and in units that are not. However, it is possible for anticipation to exist even when there are parallel pre-trends, i.e., identical pre-trends in both types of units. Suppose there are two states in which doctors have similar expectations about whether a cap will be adopted at some future date  $t$ , but at date  $t$  only one of the states actually adopts a cap. Pre-trends will be parallel and only post-treatment outcomes will diverge, yet there are anticipation effects by assumption.

<sup>2</sup> Although we selected our application because treatment is likely to be exogenous, it is possible that treatment is anticipated even when it is endogenous. Our methods may be useful even in that context, though care is required in interpreting results. For example, if it is known that endogeneity and anticipation work in opposite directions, our methods yield a bound.



**Fig. 1.** Excess physician supply before and after punitive damage caps: annual leads and lags from 5 years before to 5 years after adoption. This figure plots the normalized coefficients  $\lambda_j$  from the following regression:  $y_{ist} = \sum_{j=-5}^5 \lambda_j D_{st+j} + \gamma_{is} + \gamma_{it} + u_{ist}$ , where  $y_{ist}$  is the log of the physician count for specialty  $i$  in state  $s$  in year  $t$ ,  $D_{st+j}$  is an indicator for whether punitive damage caps is adopted in period  $t+j$ , and  $\gamma_{is}$  and  $\gamma_{it}$  are the state-specialty and specialty-year fixed effects.

adoption and that there are economic reasons for doctors to change behavior prior to reforms.

Our point is not that the pre-trends in Fig. 1 must be anticipation rather than endogeneity. Rather, we argue that there is no good reason to estimate the treatment effect of punitive damage caps on physician supply assuming that pre-trends can only be evidence of endogeneity. They could also be due to anticipation. This matters because how one interprets these pre-trends has substantial implications for how one estimates treatment effects and how large those estimates are. For example, a researcher who does not account for anticipation effects with the same sign as post-adoption effects of a policy will underestimate the full treatment effect of that policy.

With this objective in mind, we organize the paper around two contributions. Our first contribution is to provide a framework for rigorously comparing and estimating the different models that may be employed to estimate anticipation effects. We start from the premise that there exists a wide array of applied economics topics in which a researcher may be confronted with forward-looking agents whose responses anticipate future treatment. Economic theory suggests, for example, that individuals are forward looking when purchasing durable goods such as cars or houses or making human capital investments, and that firms are forward looking when investing in physical capital or entering new markets.<sup>3</sup>

Two main difficulties arise when estimating models with anticipation effects. One is that researchers may not know how many periods in advance agents anticipate treatment. A common response in the empirical microeconomics literature is to estimate a “quasi-myopic” model that includes anticipation terms for only a finite number of periods.<sup>4</sup> Within these periods, however, anticipation effects are estimated in a non-parametric manner.

An alternative approach, common in the finance and macroeconomics literature, is to posit outcomes as a function of exponentially discounted expectations about future treatment (e.g., Chow, 1989). In this formulation treatment typically has a constant contemporaneous

effect and an exponentially discounted anticipation effect. Exponential discounting has the useful feature that suitable differencing can eliminate nearly all anticipation terms.

We do not endorse any particular parameterization. The optimal approach will depend on the theory motivating the empirical analysis and on the limitations of the data. Instead, our framework advances the literature by highlighting the precise assumptions required to generate the regression models estimated in prior literature. It also provides a common benchmark for both the quasi-myopic and exponential discounting models that for the first time allows a comparison of the merits of each.

Another difficulty with estimating a model of anticipation effects is that expectations are generally unobserved. A common response is to examine shocks that alter expectations about treatment but do not actually administer a treatment. An example is a regulation that is enacted at time  $t$  but not implemented until time  $t+k$  (e.g., Alpert, 2010; Blundell et al., 2010; Gruber and Koszegi, 2001; Lueck and Michael, 2003). Unless actual expectations are observed, however, the investigator can merely demonstrate that expectations affect outcomes. She cannot identify the precise slope of the relationship and thus cannot identify treatment effects that incorporate full anticipation effects.<sup>5</sup>

An alternative approach is to assume a model of belief formation, such as rational or adaptive expectations, in order to substitute observable variables for unobservable expectations of a variable. Unless the forecast error is orthogonal to the observable variables, however, the researcher will have to instrument for them. The traditional source for these instruments is a subset of the agent's information set, for instance, lags of the observable variable (McCallum, 1976). These lags influence the agent's unobservable forecast of a variable but do not directly influence the outcome variable.

A key technical innovation in this paper is our proposal of a novel, alternative set of instruments: leads of the observable outcome or treatment variable. In general, leads can complement lags as instruments for expectations in the forward-looking regression. We show that there are situations in which lags or leads are invalid, though leads are somewhat more robust.

Our second contribution is that we explore the practical implications of the foregoing analysis in an empirical application. Specifically, we estimate the effect of punitive damage caps on equilibrium physician supply and show that accounting for anticipation could increase their estimated effect by a factor of two or more compared to a model that ignores anticipation. We first estimate a model that ignores anticipation and thus corresponds to prior analyses of tort reform, e.g., Klick and Stratmann (2007). We find that caps on punitive damages have a positive treatment effect on physician supply of 1.1% after implementation of caps. Then, we interpret the pre-period trends visible in Fig. 1 as evidence of anticipation effects and estimate the different regression models discussed in our framework. We find that damage caps have a 1.5 to 2.6% post-implementation effect after accounting for all prior anticipation effects. In addition, we estimate that damage caps had a 0.9 to 1.9% effect in each of the two years immediately preceding reform. By contrast, prior models implicitly assume zero treatment effects prior to reform. Our results are robust to different models of anticipation, which suggests that the choice of how to parameterize anticipation

<sup>3</sup> Specific examples include R&D investment decisions (Acemoglu and Linn, 2004), present value asset pricing models (Chow, 1989), pricing of durable goods (Kahn, 1986), real estate pricing (Poterba, 1984), and investment in human capital (Ryoo and Rosen, 2004).

<sup>4</sup> A less than comprehensive list includes: Acemoglu and Linn (2004); Autor et al. (2006); Ayers et al. (2005); Bhattacharya and Vogt (2003); Finkelstein (2004); Gruber and Koszegi (2001); Lueck and Michael (2003) and Mertens and Ravn (2011).

<sup>5</sup> There is also a separate literature on Ashenfelter dips, in which an observed pretrend goes in the opposite direction as the post-implementation effects of treatment (Ashenfelter, 1978). The usual interpretation of such a dip is endogenous selection. A typical solution is to net out the dip by comparing post-implementation outcomes to pre-dip outcomes, in which case the slope of the dip does not matter. However, it is also possible for anticipation to cause opposite-signed pre-trends, in which case netting them out is inappropriate. For example, Lueck and Michael (2003) discuss a case where landowners were found to have killed endangered species on their land in anticipation of a law prohibiting development in areas inhabited by these species. This anticipation effect has the opposite sign of the post-implementation effect of the law, which preserves endangered species.

effects is not critical to our estimates of those effects for our particular application.

The following is an outline of the remainder of the paper. Section 2 reviews the parameters of interest in a forward-looking regression. Section 3 elaborates on the various parametric restrictions that may be employed to reduce the number of expectation terms in the forward-looking regression. Section 4 discusses how to estimate the forward-looking regression model for a given model of belief formation. It introduces a new set of instruments, leads of the outcome variable, that can be employed to address endogeneity from forecast errors. Section 5 applies the different approaches to estimating the forward-looking model using data on tort reform and physician supply. Section 6 concludes with suggestions for future research.

## 2. Parameters of interest

Our anticipation effects framework begins with a forward-looking regression of the form

$$y_t = \lambda_0 d_t + \sum_{j=1}^{\infty} \lambda_j E_t[d_{t+j}] + e_t \quad (1)$$

where  $y_t$  is some outcome,  $\{d_{t+j}\}$  are a sequence of future values or discrete treatment states,  $E_t$  indicates the expectation taken with respect to an agent's information set at time  $t$ , and  $e_t$  is an idiosyncratic error term uncorrelated with the regressors. This specification has been used, often with further parametric restrictions, in a wide array of theoretical models (e.g., Acemoglu and Linn, 2004; Becker et al., 1994; Chow, 1989; Kahn, 1986; Poterba, 1984 and Ryoo and Rosen, 2004).

Before estimating this regression, it is useful to define the possible parameters of interest from a policy evaluation standpoint. For simplicity, consider the case where a binary treatment is implemented permanently in time  $t^*$ , and assume that individuals have perfect foresight. The baseline is outcomes at time  $t^* - \infty$ , before agents anticipated adoption of treatment.

The first parameter of interest is the average per-period effect of treatment on outcomes after implementation:  $\lambda_0 + \sum_{j=1}^{\infty} \lambda_j$ .<sup>6</sup> We call this the *ex post effect* of treatment. This parameter can itself be broken down into two components. One is the *ex post non-anticipation effect* of treatment, or  $\lambda_0$ . This is the effect on outcome  $y_t$  of a treatment at time  $t$  that is expected to last only one period. The other component is the *ex post anticipation effect*, or  $\sum_{j=1}^{\infty} \lambda_j$ . One can interpret this as the effect on  $y_t$  of the agent's expectations about whether treatment will continue to be in effect in future periods. A treatment that is implemented permanently will have the same *ex post non-anticipation effect* as a treatment that is only implemented for one period, but the latter will have a smaller *ex post anticipation effect* if agents correctly anticipate that it is temporary.

A second parameter of interest is the *ex ante effect* in period  $t^* - k$  for  $k > 0$ :  $\sum_{j=k}^{\infty} \lambda_j$ . This captures the average effect on pre-implementation outcome  $y_{t-k}$  of a permanent treatment implemented at time  $t^*$  and is driven entirely by anticipation effects. This parameter changes over time because of discounting and because agents might update their expectations when they receive new information.

Fig. 2 illustrates the effect of a permanent treatment implemented in period  $t^*$ , but anticipated in all prior periods. To simplify the graphical expression, we assume that time is continuous rather than discrete. The height of the curve prior to time  $t^*$  corresponds to the *ex ante effect*, and the height of the curve after time  $t^*$  corresponds to the *ex post effect*. The *full effect* of treatment on outcomes is the area under the solid line. It depends on both the *ex ante* and *ex post* effects.

We emphasize that observing data as in Fig. 2 is not proof of anticipation. It could be evidence of endogeneity, or both anticipation and endogeneity. We do not provide a test that sharply discriminates between the two explanations. However, if agents do anticipate treatment and the investigator does not account for this during estimation, there are two reasons to think that her estimate of the full effect of treatment will be biased.

First, the investigator is likely to ignore the *ex ante* effect, which occurs prior to adoption of treatment. Standard econometric models implicitly assume that this effect is equal to zero. But if outcomes change prior to treatment and this change is caused by anticipation of treatment, it should be counted towards the full treatment effect. Second, ignoring the *ex ante* effect will cause the investigator to estimate with bias the *ex post* treatment effect. We elaborate on this second point in the next section.

The sign of the bias from ignoring anticipation effects in econometric modeling depends on whether the sign of anticipation effects is the same as that of non-anticipation effects and whether there is endogeneity. If the signs are the same and there is no endogeneity, the investigator will likely underestimate the absolute value of the full treatment effect. First, she will ignore the *ex ante* effect, which has the same sign — a conceptual error. Second, her empirical model, likely the myopic model we describe in the next section, will not fully capture *ex post* anticipation effects — an econometric error. If the signs of anticipation effects and non-anticipation effects are different or if there is endogeneity, one cannot in general sign the bias from ignoring anticipation effects.

If there are no anticipation effects but the investigator's econometric models mistakenly assume that anticipation is present, the nature of bias in the resulting estimates will depend on the presence of endogeneity. If there is no endogeneity, the anticipation models we describe in the next sections will inefficiently include extraneous regressors, but estimates will remain unbiased. If there is endogeneity, bias may ensue, and it cannot in general be signed without making further assumptions. This underscores the need for the investigator to make cogent arguments in favor of exogeneity.

## 3. Simplifying the forward-looking model

The primary challenges with estimating the forward-looking model (1) are the potentially infinite number of expectation terms (“dimensionality” problem) and the lack of data on expectations (“unobservable expectations” problem). A researcher with a finite data set must choose some parametric restriction on the model because (1) cannot be estimated due to the dimensionality problem.

Here we discuss three different ways to tackle these two problems. First, a researcher might completely ignore anticipation effects. Second, a researcher might estimate a quasi-myopic model that includes only a finite number of anticipation terms. Third, a researcher could adopt an exponential discounting model that assumes outcomes are a function of exponentially discounted expectations about treatment.

We do not endorse any particular approach, though the first approach is surely the least satisfactory due to omitted variable bias. Towards the end of this section we discuss the relative merits of the quasi-myopic and exponential discounting models and emphasize that these two approaches are complementary.

### 3.1. Addressing the dimensionality problem

#### 3.1.1. Myopic model

The simplest approach to dealing with anticipation effects is for the econometrician to ignore them and estimate a myopic model such as

$$y_t = \beta_0^{myopic} d_t + u_t \quad (2)$$

where  $y_t$  is the outcome of interest,  $d_t$  is the treatment, and  $u_t$  includes the model error,  $e_t$ , from the forward-looking model (1) and also

<sup>6</sup> If individuals lack perfect foresight, the estimated effect will differ as each  $\lambda_j, j > 0$  will have to be weighted by  $E[d_{t+j}] < d_{t+j}$ . A similar issue will arise for the *ex ante* effect.

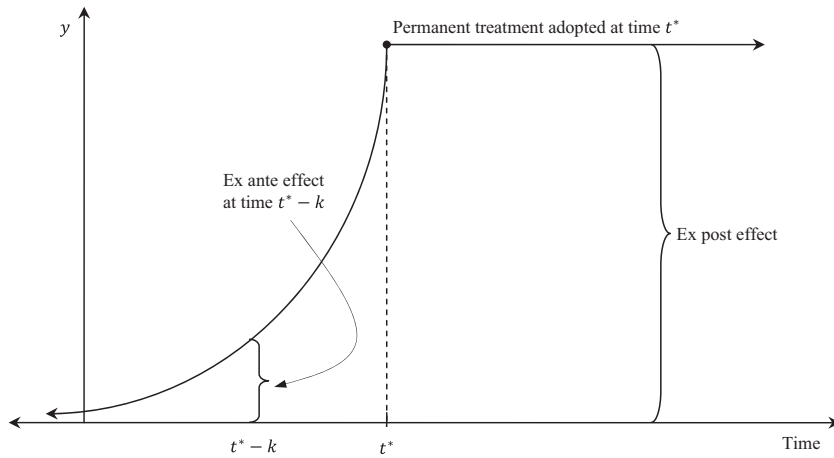


Fig. 2. Parameters of interest.

additional error resulting from model misspecification. This is the model researchers typically use when they assume no anticipation effects.

The omission of anticipation effects generates omitted variable bias. The specific nature of the bias depends on which parameter of interest one seeks to estimate. We will focus on the ex post effect of treatment because it corresponds closely to what most applied researchers are interested in estimating when working in an environment with no anticipation. If the true specification is the forward-looking model (1), then the treatment effect estimated by the myopic model is equal to

$$\text{plim}|\hat{\beta}_0^{\text{myopic}}| = |\lambda_0| + \sum_{j=1}^{\infty} |\lambda_j| \alpha_j$$

where  $\alpha_j$  is the coefficient from a regression of expected future treatment,  $E_t[d_{t+j}]$ , on current treatment,  $d_t$ .<sup>7</sup> Intuitively, the coefficient on current treatment in the myopic model captures some of the effect of future treatment. It differs from the ex post effect,  $\lambda_0 + \sum_{j=1}^{\infty} \lambda_j$ , by the coefficients  $\alpha_j$ .

The coefficient  $\alpha_j$  will usually be positive because current treatment and expectations of future treatment are likely to be positively correlated.<sup>8</sup> Moreover, the magnitude of  $\alpha_j$  will be less than one unless the current state of treatment perfectly predicts all future expected states of treatment. Thus, assuming  $\lambda_0$  and  $\{\lambda_j\}$  have the same sign, the myopic model is likely to underestimate the magnitude of the ex post effect.

This point is illustrated in Fig. 3, which plots the outcome of a forward-looking process after adoption of a permanent treatment at time  $t^*$ . The true ex post effect of the intervention is to increase outcomes by  $\sum_{j=0}^{\infty} \lambda_j = y_{\text{post}} - y_{\text{pre}}$  per period. Estimation of a myopic model, however, yields a treatment effect,  $\hat{\beta}_0^{\text{myopic}}$ , that is the difference between the average outcome,  $y_{\text{pre}}^{\text{sample}}$ , before the law is passed and the average outcome,  $y_{\text{post}}$ , after the law is passed. The myopic estimate is less than the true ex post effect because the researcher observes a finite number of pre-treatment periods, say  $[t^* - k, t^*]$ , but expectations may have begun shifting outcomes well before  $t^*$ . Therefore the average pre-treatment outcome in the sample,  $y_{\text{pre}}^{\text{sample}}$ , is greater than the true pre-treatment outcome,  $y_{\text{pre}}$ .

<sup>7</sup> The econometrician has the same problem estimating the  $\alpha_j$  coefficients as she does estimating the coefficients in the quasi-myopic model we describe in the next section: she does not observe expected future treatment terms.

<sup>8</sup> Negative correlation between current treatment and expected future treatment implies that subjects frequently alternate between treated and untreated states. It is difficult to come up with examples of such treatments. Zero correlation is possible, but rules out infrequent treatment or treatment that lasts multiple periods.

### 3.1.2. Quasi-myopic model

To address the shortcomings of the myopic model, a researcher might estimate a quasi-myopic model that assumes agents have anticipation effects, but only for a finite number of periods  $S$ :

$$y_t = \beta_0^{\text{quasi}} d_t + \sum_{j=1}^S \beta_j^{\text{quasi}} E_t[d_{t+j}] + u_t. \quad (3)$$

This addresses the dimensionality problem by ignoring anticipation terms after  $S$  periods, perhaps on the theory that agents do not forecast further than  $S$  periods ahead or that anticipation effects past  $S$  years have negligible effects.

### 3.1.3. Exponential discounting model

The third approach to reducing the dimensionality of the forward-looking model is to assume that treatment has a constant contemporaneous effect of  $\lambda_0 = \beta$  and an anticipation effect  $j$  periods prior to treatment of  $\lambda_j = \beta \theta^j$ :

$$y_t = \beta d_t + \beta \sum_{j=1}^{\infty} \theta^j E_t[d_{t+j}] + u_t. \quad (4)$$

The ex post effect of treatment is then estimated as  $\hat{\beta}/(1-\theta)$ . The central benefit of the assumption that outcomes are a function of exponentially discounted expectations about treatment is that subtracting  $\theta y_{t+1}$  from Eq. (4) will enable the researcher to generate an equation with only two regressors.

## 3.2. Addressing the unobservable expectations problem

Both the quasi-myopic and exponential discounting models contain unobserved expectations as regressors. A common solution to addressing the unobservable expectations problem is to assume a model of belief formation, such as rational or adaptive expectations, in order to substitute observable variables for unobservable expectations of a variable. Below we derive a solution under the assumption that agents have rational expectations. We model adaptive expectations in Appendix A. We confine our discussion to the exponential discounting model because estimates of the quasi-myopic model typically substitute realizations,  $d_{t+j}$ , for expectations,  $E_t[d_{t+j}]$ , rather than modeling expectations directly. However, it is possible to generalize our modeling of beliefs for the exponential discounting model to the quasi-myopic model.

In general, expectations may depend on realizations ( $E[z] = z + v$ ) or vice versa ( $z = E[z] + v$ ). Economic theory should dictate which path to take. We focus here on the case where expectations depend on realizations because this corresponds to the application we consider



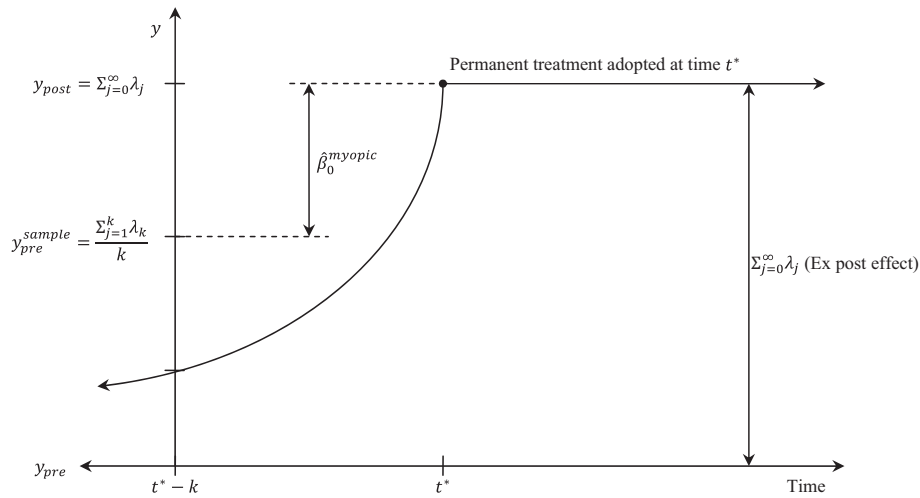


Fig. 3. Estimate from a myopic model.

in Section 5. We consider alternative formulations of the rational expectations assumption in Appendix B.

Let expectations be a function of treatment so that  $E_t[d_{t+j}] = d_{t+j} + v_{t,t+j}$ , where  $E_t[d_{t+j}v_{t,t+j}] = 0$  and  $v_{t,t+j}$  is the forecast error resulting from an agent's time- $t$  forecast of the treatment  $d_{t+j}$ . In this case we can substitute the rational expectations assumption directly into the exponential discounting model to obtain

$$y_t = \beta \sum_{j=0}^{\infty} \theta^j d_{t+j} + e_t + \beta \sum_{j=1}^{\infty} \theta^j v_{t,t+j}.$$

Subtracting  $\theta y_{t+1}$  from  $y_t$  then yields

$$y_t = \theta y_{t+1} + \beta d_t + w_t \quad (5)$$

where

$$\begin{aligned} w_t &= e_t - \theta e_{t+1} + \beta \sum_{j=1}^{\infty} \theta^j v_{t,t+j} - \beta \sum_{j=2}^{\infty} \theta^j v_{t+1,t+j} \\ &= [e_t - \theta e_{t+1}] + \beta \theta v_{t,t+1} + \beta \sum_{j=2}^{\infty} \theta^j [v_{t,t+j} - v_{t+1,t+j}]. \end{aligned}$$

The error term  $w_t$  has three components. One is the change in model error,  $e_t - \theta e_{t+1}$ . A second is the error in forecasting time  $t+1$  treatment at time  $t$ . The third component is the change in forecasts about time  $t+j$  treatment ( $j > 1$ ) from time  $t$  to time  $t+1$ . There is, however, only one definite source of endogeneity between outcome  $y_{t+1}$  and the error term: the model error  $e_{t+1}$ .<sup>9</sup> Section 4 of our paper discusses how to employ instruments to overcome this problem.

### 3.3. Comparing models of anticipation

We do not endorse the quasi-myopic model over the exponential discounting model or vice versa. Each model has imperfections and the choice between the two should be dictated by the theory motivating the empirical analysis and the available data.

The quasi-myopic model has two shortcomings. First, it assumes that there are no more than  $S$  periods of anticipation effects. If this

assumption is incorrect then, as in the myopic model, bias ensues. Moreover, if the length  $T$  of the panel of data available to the researcher is less than  $S$ , the researcher cannot estimate all anticipation effects.

Second, the way that the quasi-myopic model is actually implemented in the literature implicitly substitutes realizations,  $d_{t+j}$ , for expectations,  $E_t[d_{t+j}]$ , of treatment in (3). This introduces measurement error bias unless agents have perfect foresight. One solution to the measurement error problem is to instrument for the affected regressors, using analogous methods to those we employ for the exponential discounting model. However, this requires at least  $S$  instruments for the  $S$  periods of anticipation effects the researcher seeks to estimate. The exponential discounting model, however, only requires one instrument.

A problem with the exponential discounting model is that it assumes expectations decay at an exponential rate, which is an admittedly strong assumption, perhaps motivated by convenience of estimation rather than theory. That said, there are numerous theoretical models with exponential discounting of future income or utility that yield regression models with exponential discounting of expected treatment. Even as an ad hoc assumption on a regression model, exponential discounting does not appear any more troubling than the possibly incorrect but frequent assumption in theoretical models that future profits or utility are exponentially discounted.

A second issue with the exponential discounting model is that modeling expectations generally requires imposing structure on the error term. We discuss this in more detail in Section 4. Of course, if one is to take unobserved expectations seriously in the quasi-myopic model, one may have to impose structure on the error term in that model as well.

We emphasize that one cannot, a priori, determine whether the parametric restrictions in the quasi-myopic model or those embodied in the exponentially discounted model yield lower bias. If there are more than  $S$  periods of anticipation effects, then the quasi-myopic model suffers omitted variable bias. But exponential discounting may also be a poor approximation to the time path of anticipation effects and suffer misspecification bias.

These competing arguments aside, we note that these regression models are not mutually exclusive. For example, one could specify an estimating equation with a few quasi-myopic terms in addition to an exponentially discounted expectations process. Although this layering increases data and power requirements, it does permit one to check if the quasi-myopic model is empirically more appropriate by testing whether the included quasi-myopic terms are statistically significant.

<sup>9</sup> There is no endogeneity from  $v_{t,t+1}$  because, although  $y_{t+1}$  is a function of  $d_{t+1}$ , we have assumed that  $d_{t+1}$  is orthogonal to  $v_{t,t+1}$ . Nor is there endogeneity from the change in forecasts ( $v_{t,t+j} - v_{t+1,t+j}$ ,  $j > 1$ ) because under rational expectations these forecast updates are orthogonal to prior forecast errors ( $v_{t,t+j}$ ) and thus orthogonal to  $E_t[d_{t+j}]$  too. Indeed, there is no additional endogeneity even if  $\{d_t\}$  are serially correlated because  $E_t[d_{t+j}v_{t,t+j}] = 0$  by assumption.

#### 4. Estimation

This section takes up estimation of anticipation effects models. The focus will be on estimating the exponential discounting model, though the section will contain lessons for the quasi-myopic model as well. The Euler Eq. (5) derived in the previous section greatly reduced the dimensionality of the forward-looking model (1), but consistent estimation still requires, at a minimum, that the researcher addresses the correlation between  $y_{t+1}$  and the model error  $e_{t+1}$  contained in the error term  $w_t$ . One solution is to find an instrument.

The usual source for these instruments is a subset of the agent's information set, for instance, lags of the endogenous variable (McCallum, 1976). This is typically motivated by modeling expectations as a linear projection of the variables in the agents' data sets, which include lagged values of the endogenous variable. An alternative motivation is to note that since  $y_{t+1}$  and its lags all depend, according to the forward-looking model (1), on expectations about future treatment, changes in those future expectations will move both lags of  $y_{t+1}$  and  $y_{t+1}$ . The exclusion restriction is completed by noting that  $y_t$  in Eq. (1) does not depend on lagged values.

This alternative motivation suggests a new set of instruments we propose here: leads of the endogenous variable. Like  $y_{t+1}$ ,  $\{y_{t+k}\}$  for  $k > 1$  depend on expectations about future treatment and thus are correlated with  $E[y_{t+1}]$ . Leads meet the exclusion restriction because, as shown by Eq. (5), the current outcome  $y_t$  is related to future treatments, and thus future outcome variables, only through  $y_{t+1}$ .

One could alternatively instrument for  $y_{t+1}$  using leads of  $d_{t+1}$  rather than leads of  $y_{t+1}$  since shocks to future expectations of  $d_{t+1}$  are what ultimately drive identification. Whether this is more efficient than using leads of  $y_{t+1}$  depends on the variance–covariance structure of model error  $e_t$  and forecast error  $v_t$ , which is unknown a priori. However, there is a strong practical reason to prefer leads of  $y_{t+1}$ : they embed more information than leads of  $d_{t+1}$  for any finite data set. For example, consider a panel with 10 time periods. Instrumenting for  $y_9$  with  $y_{10}$  necessarily includes information about  $\{d_{11}, d_{12}, \dots\}$  because  $y_{10}$  is a function of future treatments. That information is unavailable when using leads of  $d_9$ , however, because the data set only contains 10 time periods. This advantage is reduced if data on the future treatments  $\{d_{11}, d_{12}, \dots\}$ , but not the future outcomes  $\{y_{11}, y_{12}, \dots\}$ , are available. There are some scenarios, however, where using leads of  $d_t$  is preferable to using leads of  $y_{t+1}$ . We discuss them at the end of this section.

Below, we derive the technical conditions necessary for estimation to be consistent. We work with a model that generalizes Eq. (5) to a panel setting and allows us to draw on results from the literature on dynamic panel estimation. The main result is that estimation requires imposing some degree of limited serial correlation. One might think that this is a poor assumption in a panel data setting (Bertrand et al., 2004). However, if errors are uncorrelated across panels then it is possible to test for serial correlation (Arellano and Bond, 1991). Moreover, we emphasize that this issue is not unique to the exponential discounting model. Estimating the quasi-myopic model requires the researcher either to instrument for that model's future expectations, which bring up analogous problems, or to assume perfect foresight, which is difficult to justify (Gruber and Koszegi, 2001).

##### 4.1. Instrumenting with lags and leads

We are interested in estimating a model of the form

$$y_{it} = \theta y_{i,t+1} + \alpha_1 x_{it} + \alpha_2 x_{i,t+1} + \beta d_{it} + \eta_i + w_{it} \quad (6)$$

where  $i = 1 \dots N$ ,  $t = 1 \dots T$ , and  $0 < \theta < 1$ . We assume that  $\eta_i$  and  $w_{it}$  are independently distributed across  $i$  with  $E[\eta_i] = E[w_{it}] = E[\eta_i w_{it}] = 0$ . The number of time periods  $T$  is fixed and the number of individuals  $N$  is large. Without loss of generality, we assume that  $x_{it}$  and  $x_{i,t+1}$  are

strictly exogenous and known in advance. Although we shall focus here on the validity of instrumenting with leads, it is easy to adapt our argument to show that lags are also valid.

Direct OLS estimation of Eq. (6) is inconsistent because  $E[y_{i,t+1} \eta_i] \neq 0$ . Estimating first differences (defined here as  $\Delta y_{it} = y_{it} - y_{i,t+1}$ ) fails because  $E[\Delta y_{i,t+1} \Delta w_{it}] \neq 0$ . A within estimator suffers from this same problem, although the bias may disappear as  $T \rightarrow \infty$ . The logic in Arellano and Bond (1991); Arellano and Bover (1995) and Blundell and Bond (1998) suggests that one solution is to instrument for  $\Delta y_{i,t+1}$  using leads of  $y_{i,t+1}$ . Leads are valid instruments if the following standard assumptions are met:

##### A1.

$$E[y_{i,T} w_{it}] = 0 \forall i, \forall t \leq T-1$$

##### A2.

$$E[w_{is} w_{it}] = 0 \forall t \neq s$$

##### A3.

$$E[n_i \Delta w_{i2}] = 0 \forall i.$$

Assumption A1 requires  $y_{iT}$  to be uncorrelated with past disturbances. Assumption A2 requires these disturbances to be uncorrelated. These two assumptions together imply the following moment conditions:

$$E[y_{i,t+j} \Delta w_{it}] = 0 \quad \forall j \geq 2, \quad \forall t. \quad (7)$$

Assumption A3 requires the terminal conditions to be mean stationary. In other words, conditional on the covariates  $x_{it}$ , individuals with large random effects  $\eta_i$  must not be systematically closer or farther away from their steady states than individuals with small random effects, so that the terminal conditions are representative of the steady state behavior of the model. If it holds, A3 implies the following additional moment conditions:

$$E[\Delta y_{i,t+1} w_{it}] = 0 \quad \forall t. \quad (8)$$

The model is overidentified if  $T > 3$  but can be estimated using the Generalized Method of Moments (GMM) framework developed by Hansen (1982). “Difference GMM” estimation exploits the moment conditions (7) while “system GMM” estimation exploits both conditions (7) and (8).

Assumption A2 is central to the validity of these estimation procedures. As currently stated, however, it is actually stronger than necessary. Limited serial correlation of order  $H > 0$  is acceptable so long as the researcher takes care to omit the affected instruments and enough instruments remain for identification. We therefore loosen A2:

##### A2'.

$$E[w_{it} w_{i,t+j}] = 0 \forall j > H, H \geq 0.$$

This changes our moment conditions (7) and (8) to

$$\begin{aligned} E[y_{i,t+j} \Delta w_{it}] &= 0 \forall j \geq H+2, \quad \forall t \\ E[\Delta y_{i,t+H+1} w_{it}] &= 0 \forall t. \end{aligned}$$

Whether or not these assumptions are satisfied depends on the content of the error term in the Euler Eq. (5), which in turn depends on how expectations are specified.

For notational convenience, we now drop the  $i$  subscript for the remainder of this section. Following Section 3.2, we consider the case where expectations are a function of treatment:  $E_t[d_{t+j}] = d_{t+j} + v_{t,j}$ . The error term, derived in Eq. (5), is

$$w_t = e_t - \theta e_{t+1} + \beta \theta v_{t,t+1} + \beta \left( \sum_{j=2}^{\infty} \theta^j [v_{t,t+j} - v_{t+1,t+j}] \right)$$

In order for the error term  $w_t$  to satisfy Assumption A2', the following four conditions must be satisfied for all periods and all agents, for some  $H \geq 1$ :

1.  $E[e_t e_{t+j}] = 0 \forall j > H$
2.  $E[e_t v_{t,j,t+k}] = 0, \forall k > j, \forall j > H$
3.  $E[(v_{t,t+k} - v_{t+1,t+k}) v_{t,j,t+j+1}] = 0 \forall k > 1, \forall j > H$
4.  $E[(v_{t,t+k} - v_{t+1,t+k})(v_{t,j,t+m} - v_{t+j+1,t+m})] = 0 \forall k > 1, m > j + 1, j > H$ .

In words, condition 1 means that autocorrelation in model error  $e_t$  cannot be higher than order  $H$ . Condition 2 means that the model error is orthogonal to the  $H$ -step-ahead-and-beyond forecast error. Condition 3 means that the change in a forecast from period  $t$  to period  $t + 1$  is uncorrelated with the level of a forecast in period  $t + j$ . Condition 4 means that independent information is used to update forecasts.

Conditions 1, 2 and 4 are plausible in many scenarios, but condition 3 may be an unrealistic assumption. It holds in the cases of perfect serial correlation (so the change in forecast ( $v_{t,t+k} - v_{t+1,t+k}$ ) is equal to 0) or no serial correlation (so  $E[v_{t,j,t+j+1} v_{t+1,t+k}] = 0 \forall j, k, l$ ). These two extremes are not satisfied in most applications. Rational expectations, however, offers some hope. It implies that the (perhaps nonzero) expectation in condition 33 is not a function of  $t$ . In other words, an agent's forecast error might depend on whether she is predicting an event three time periods in the future versus four time periods in the future, but it does not depend on the particular time period she is forecasting from.

This means that we can rewrite our moment conditions (7) and (8) as

$$\begin{aligned} E[y_{t+j} \Delta e_t - k_1(j; \beta, \theta)] &= 0 \\ E[\Delta y_{t+j} e_t - k_2(j; \beta, \theta)] &= 0 \end{aligned}$$

where  $k_1(\cdot)$  and  $k_2(\cdot)$  are constants that do not depend on  $t$  or the data. They will be absorbed into the constant term, which means that the researcher can still identify the parameters of interest,  $\beta$  and  $\theta$ .<sup>10</sup> Unfortunately, the non-zero moment condition differs for each instrument, which means that the researcher may not be able to simultaneously include multiple instruments from different time periods. The optimal solution is for the researcher to specify each instrument as a separate GMM equation and then estimate the entire system simultaneously.

Finally, we note that estimation is further complicated when the treatment variable is discrete. We discuss how to handle the binary case in Appendix C.

<sup>10</sup> It may appear surprising that we can effectively ignore  $k_1(\cdot)$  and  $k_2(\cdot)$  even though they are functions of our parameters. We are able to do this because  $k_1(\cdot)$  and  $k_2(\cdot)$  are constants and thus merely represent level shifts of the GMM minimization problem. Consider the analogous problem for OLS:  $\min_{\beta_0, \beta_1} (y - \beta_1 x_1 - \beta_0 - k(\beta_1))^2$  where  $k(\beta_1)$  is some constant that is a function of  $\beta_1$  and independent of  $x_1$ . Identification of  $\beta_0$  is impossible but OLS can still identify  $\beta_1$  even without knowing the functional form of  $k(\beta_1)$ .

## 4.2. Comparing lag and lead instruments

The discussion above showed that lags and leads of the outcome variable are both valid instruments under our standard forward-looking specification when agents have rational expectations. We also noted that the researcher could alternatively use leads (but not lags) of the treatment variable as instruments. Here we show that the possibility of instrumenting with leads of the treatment variable provides a flexibility that makes leads generally better instruments than lags. We demonstrate this by considering two common situations that depart from the standard specification.

Suppose, for example, that agents do not continuously update their forecasts of future variables with new, orthogonal information. In this case lags of the outcome variable are no longer valid instruments. To illustrate why, we examine the case where agents have rational (i.e., unbiased) forecasts of treatment but never update these forecasts (e.g., Anderson et al., 2013; Carroll, 2003). This implies that the forecast error no longer depends on the date the forecast was made, so that  $E_t[d_{t+j}] = d_{t+j} + v_{t,t+j} = d_{t+j} + v_{t+j}$ . The exponential discounting model may now be written as

$$y_t = \beta \sum_{j=0}^{\infty} \theta^j (d_{t+j} + v_{t+j}) + e_t.$$

Subtracting  $y_{t+1}$  yields the Euler equation

$$y_t = \theta y_{t+1} + \beta d_t + \beta \theta v_{t,t+1} + e_t - \theta e_{t+1}.$$

Lags are necessarily invalid instruments in this specification because  $y_{t-j}$  for any  $j > 1$  is correlated with  $v_{t+1}$ . Leads, however, remain valid instruments.

Alternatively, suppose the researcher derives an Euler condition that includes lagged dependent variables, e.g.,

$$y_t = \theta y_{t+1} + \gamma y_{t-1} + \beta d_t + w_t.$$

In this scenario, leads of  $y_{t+1}$  are invalid instruments for  $y_{t+1}$ . The reason is that  $y_{t+j}$  for any  $j > 1$  is correlated with  $y_t$  and thus  $w_t$  due to the autoregressive specification for  $y_t$ . Leads of the outcome variable are thus no longer orthogonal to the error term. Lags, however, remain valid instruments. Interestingly, leads of the treatment variable,  $d_t$ , are also valid instruments, so long as treatment remains exogenous to the error term.

## 5. Application

In this section we estimate the effect of punitive damage caps on physician supply using the different methods of estimating anticipation effects that we have described. Section 5.1 describes the data we employ. Sections 5.2 and 5.3 provide background on medical malpractice liability and present evidence that changes in the supply of physician prior to adoption of tort reform are plausibly due to anticipation. Finally, Section 5.4 explains our empirical strategy and Section 5.5 reports our estimates.

### 5.1. Data

Our analysis employs annual physician count data from the American Medical Association's Physician Masterfile. These data include private practitioners, hospital staff, residents, and locum tenens (temporary employees). They exclude military doctors. Physicians are categorized into one of 20 possible specialties and have state identifiers. The data span the period 1980–2001, with gaps in 1984 and 1990.

Klick and Stratmann (2007) note that some physician specialties are sued more often than others and correspondingly group them into four equally-sized risk tiers, displayed in Table 1. We use their definitions to limit our data and analysis to the two riskiest tiers (tiers 1

**Table 1**

Physician specialties by risk tier.

Source: Klick and Stratmann (2007).

| Tier 1                  | Tier 2             | Tier 3                             | Tier 4                                      |
|-------------------------|--------------------|------------------------------------|---|
| Emergency medicine      | Anesthesiology     | Allergy & immunology               | Diabetes                                    |
| General practice        | General surgery    | Dermatology                        | Medical oncology                            |
| Neurological surgery    | Orthopedic surgery | Nephrology                         | Neoplastic diseases                         |
| Obstetrics & gynecology | Plastic surgery    | Physical medicine & rehabilitation | Psychiatry                                  |
| Thoracic surgery        | Radiology          | Rheumatology                       | Public health & general preventive medicine |

Specialties in tier 1 exhibit the highest average medical malpractice award per doctor and specialties in tier 4 exhibit the lowest average.

and 2) because we expect these to be more affected by tort reform than the other two tiers.

Fig. 4 graphs the total counts over time of the five most populated specialties in the data. The supply of general practitioners is declining over time, the supply of general surgeons is stagnant, and the rest are rising.

Our tort reform data come from Avraham (2010). These data indicate, for the same time period as our physician supply data, whether ten different tort reforms are in effect at the state-year level. These reforms are defined in Table 2 and are coded as 0–1 indicator variables.

The supporting analysis described in Section 5.3 employs state-level data on per capita income and medical transfer payments from the Regional Economic Information System (REIS). Medical transfer payments include Medicare and Medicaid payments received by state governments.

## 5.2. Background on tort liability and tort reform

Tort liability is akin to a legally-mandated alteration of the implicit labor contract between a patient and her physician. In most cases, it requires the physician to provide the quality of care that a “reasonable physician” would provide and compensates patients who suffer injuries due to inadequate care. Compensation may include economic damages for lost wages and the cost of additional medical care; non-economic damages for pain and suffering from the injury; and punitive damages intended to punish the doctor for outrageous misconduct.

Policymakers are concerned that tort liability is driving away doctors and thus reducing patients’ access to care. This claim has received substantial attention from scholars and the media.<sup>11</sup> However, the effect of tort liability on equilibrium physician supply is theoretically ambiguous. On the one hand, an increase in liability can increase the cost of providing service and drive physicians away. On the other hand, if transaction costs prevent a patient and physician from writing a complete contract that covers all contingencies, including instances of malpractice, then the mandatory terms imposed by tort liability that improve the contract will increase the value of (and demand for) physician care.

Tort reform refers to various changes to these mandatory contract terms. Table 2 provides a description of the most common reforms. Most of them, such as caps on punitive damages, lower damages paid by doctors found liable. Some, like split recovery, reduce the damages received by patients, and thus their incentive to sue. Others, such as reform of joint and several liability, reduce the extent to which hospitals share liability and thus effectively increase the overall liability of doctors (Currie and MacLeod, 2008).

Several recent studies analyze the impact of tort liability on physician supply. Kessler et al. (2005) perform a state-level difference-in-differences analysis and find evidence that reforms directly affecting how much a liable defendant has to pay increase physician supply by

3%. Matsa (2007) examines the effect of damage caps on physician supply and finds it increases the supply of physicians by about 10%, but only in rural counties. Klick and Stratmann (2007) employ a triple-differences model and estimate that caps on non-economic damages are associated with a 6% increase in physician supply for high-risk specialties relative to low-risk specialties.

In this paper, we focus on the effect of punitive damage caps on physician supply. We chose this reform for two reasons. First, Fig. 1 shows that this reform exhibits pretrends that are strikingly similar to Fig. 2, a canonical illustration of anticipation effects. This suggests that it is a good candidate to illustrate how to estimate our forward-looking model. Second, prior empirical studies have argued that this reform was adopted for reasons unrelated to physician supply (Avraham, 2007; Currie and MacLeod, 2008).<sup>12</sup> We elaborate on this second point in the next section.

## 5.3. Evidence on anticipation of punitive damage reforms

The existence of pre-trends does not prove that anticipation exists. Indeed, the conventional diagnosis is that pre-trends imply endogeneity. In this section we cast doubt on that diagnosis and argue that anticipation is a plausible cause. Historical accounts of tort reform suggest that punitive damage caps were not driven by events in the health care industry. Moreover, the data on physician supply contradict the usual story given for why adoption of punitive damage caps might be endogenous.

Historically, physician lobbies have not focused on enacting punitive damage caps. Hubbard (2006) notes that physician lobbies focused instead on noneconomic damage caps, period payment, contingency fees, and collateral source. The effects of these differential lobbying efforts are visible in Table 3, which shows when different tort reforms were adopted and whether they apply only to medical malpractice. As expected, the reforms targeted by the physician lobby (noneconomic damage caps, period payment, contingency fees, and collateral source) are generally aimed at medical malpractice. By contrast, punitive damage caps, along with joint and several reform and split recovery, stand out as applying to all industries.<sup>13</sup>

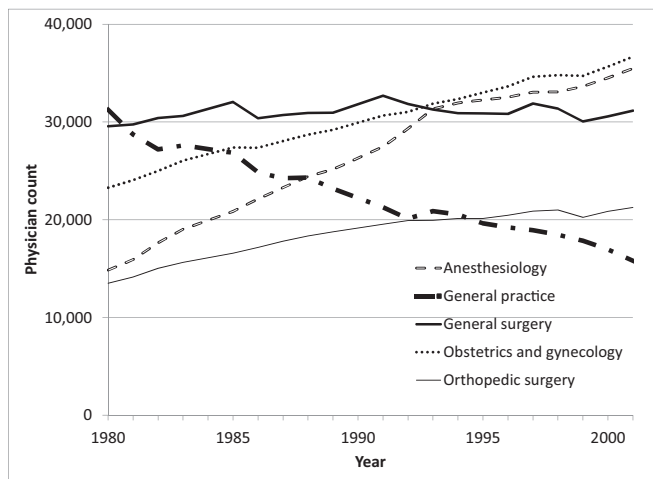
Indeed, punitive damage caps reforms were spurred by events unrelated to the health care industry (Avraham, 2007; Currie and MacLeod, 2008). These reforms were adopted in two waves, one in the mid-1980s and the second in the mid-1990s, and were driven by well-publicized anecdotes unrelated to physician suits. For example, the 1980s reforms were triggered by incidents where an insolvent drunk driver or hazardous waste operator’s liabilities were reallocated to a solvent municipal entity under joint and several liability (Rabin, 1988). The 1990s reforms to punitive damages were driven by the well-publicized case of Liebeck v. McDonald’s Restaurants (1994), which resulted in an initial award of

<sup>11</sup> See Born et al. (2006); Currie and MacLeod (2008); Helland et al. (2005); Holtz-Eakin (2004); Kessler et al. (2005); Klick and Stratmann (2007); Matsa (2007) and The Economist (2005).

<sup>12</sup> Split recovery and joint and several liability also fit these two criteria. An earlier version of this paper also examined these reforms and found similar results to the ones presented for punitive damage caps.

<sup>13</sup> Table 3 indicates that there are a few states that did actually target their reforms to medical malpractice. We exclude them from our main empirical specification.





Note: data for 1984 and 1990 are interpolated.

Fig. 4. Physician supply from 1980 to 2001. Note: Data for 1984 and 1990 are interpolated.

\$2.7 million in punitive damages to a lady who had spilled hot coffee on herself (Nockleby and Curreri, 2005).

Looking beyond the history of tort reform and focusing on changes in physician supply around adoption of reform, we find that the data are not consistent with the usual theory for why supply might drive punitive damage caps. The usual theory is that states adopt liability-reducing reforms in order to stem a decline in physician supply. Fig. 5 illustrates this argument in the context of total damage caps, a reform that is always targeted at physicians. Supply was falling before adoption of total damage caps; after adoption supply recovered. By contrast, Fig. 1 shows that supply was rising before adoption of punitive damage caps, a reform that reduced liability.

A related concern is that adoption of punitive damage caps might be correlated with unobserved variables that affect physician supply. For example, business-friendly states may attract physicians because they make it easy to set up a new practice, and coincidentally these states may also be more likely to pass tort reform. If this were the case, we would expect to observe a trend in physician supply both before and after the enactment of punitive damage caps. But, Fig. 1 shows a pre-trend only.

Finally, Fig. 6 provides evidence consistent with the plausible exogeneity of punitive damage caps by showing that per capita income and medical transfer payments, variables that may be important to physician labor supply, do not appear to be correlated with the passage of reform.

While these arguments suggest that pre-trends may not be due to endogeneity, they do not directly address anticipation. To justify estimation of anticipation effects, one must show that physicians could directly or indirectly anticipate the reform years prior to adoption.

Physicians clearly have a large incentive to care about tort reform because it affects their insurance premiums.<sup>14</sup> For example, neurosurgeons in St. Clair County, Illinois, paid an average premium of \$228,396 in 2004, but their colleagues in neighboring Wisconsin, which is not as friendly to plaintiffs, paid less than one-fifth of that (The Economist, 2005). Correctly anticipating future reforms is important for at least three reasons. First, relocation costs are large so it is advantageous to locate in a state where tort liability laws are not only currently physician-friendly but also are expected to remain that way. Chou and Lo Sasso (2009) present evidence from a survey of graduating medical residents that tort laws significantly affect a physician's practice

Table 2  
Tort reform descriptions.  
Source: Avraham (2010).

| Tort reform             | Description  |
|-------------------------|--|
| Collateral source       | Allows damages to be reduced by the value of compensatory payments already made to the plaintiff |
| Contingency fees        | Places limits on attorney contingency fees   |
| Joint and several       | Limits damages recoverable from parties only partially responsible for the plaintiff's harm      |
| Noneconomic damage caps | Limits awards for noneconomic damages in malpractice cases                                       |
| Periodic payment        | Requires part or all of damages to be paid in the form of an annuity                             |
| Punitive damage caps    | Prohibits or limits recovery of punitive damages from physicians                                 |
| Punitive evidence       | Requires plaintiff to show by clear and convincing evidence that a defendant acted recklessly    |
| Split recovery          | Requires some of the punitive damages to go to a state fund for uncompensated tort victims       |
| Total damage caps       | Limits awards for total damages  |
| Victims' fund           | Establishes a no-fault compensation fund for medical malpractice victims                         |

location choice.<sup>15</sup> Second, some changes to tort reforms are implemented retroactively (Avraham, 2007). Third, assets accumulated prior to adoption of reforms are at risk (or removed from risk) after reforms. Therefore, the economic return to current supply depends on future liability rules.

Newspaper articles are one obvious channel through which physicians could be alerted to forthcoming reform. We investigated this possibility by searching the newspaper archives of Pennsylvania, the largest state that adopted punitive damage caps and that also has an online database of newspaper articles available prior to adoption. Pennsylvania adopted punitive damage caps on January 25, 1997, and, conveniently, adopted no other reforms during that decade. We found eighty-four articles published in its two largest local newspapers mentioning "tort reform" prior to 1997. Fig. 7 displays the frequency of these articles in each quarter preceding the reform and shows the policy discussion that took place over an extended period of time. One article published about two years prior to enactment wrote that "the key goals of the [state] administration... have been to place a cap on punitive-damage awards" (Siegel, 1995).

Finally, we briefly discuss why physician supply should respond at all to punitive damages. It has been reported that punitive damages are awarded in only 1–4% of all medical malpractice trials (Cohen, 2004; Cohen and Harbacek, 2011). However, this figure underestimates the impact of punitive damages. First, according to Table 3, nineteen states have adopted caps on punitive damages. Since the 1–4% figure is a national figure, it reflects both states that allow punitive awards and states that cap or prohibit those awards. Punitive damages will play a larger role in states that do not cap those damages. Second, even if punitive damages are a small percentage of awards, they may be the one aspect of tort damages that cannot be insured by physicians. Nearly half the states do not allow malpractice liability insurance to cover punitive damages (McCullough et al., 2004; Wilson et al., 2008).<sup>16</sup> Moreover,

<sup>15</sup> The change in physician supply resulting from tort reform is most likely due to the large inflow of new residents and outflow of retirees because it is costly for currently practicing physicians to move states (Kessler et al., 2005). We are unfortunately unable to verify this with our data. However, we did perform a simple calculation using the published fact that in 1996 approximately five percent of the physicians in our sample were new residents (AMA, 1997). Extrapolating this trend implies that more than one half of all practicing physicians entered the profession within the past 14 years.

<sup>16</sup> Viscusi and Born (2005) estimate that medical malpractice insurers incur 6–7% lower losses in states that prohibit insurance coverage of punitive awards. Given that 98.8% of total malpractice awards are covered by liability insurance (Zeiler et al., 2007), this implies that punitive damages are a disproportionate source of risk to doctors, perhaps more than 83% (= 6%/(6% + 1.2%)) of all financial risk they bear from medical malpractice liability.

<sup>14</sup> Even if doctors can pass these costs through to patients, many uninsurable risks such as reputational effects and lost time remain. See Currie and MacLeod (2008) for additional discussion.

**Table 3**

Summary of tort reform laws enacted during 1980–2001.

Source: Avraham (2010).

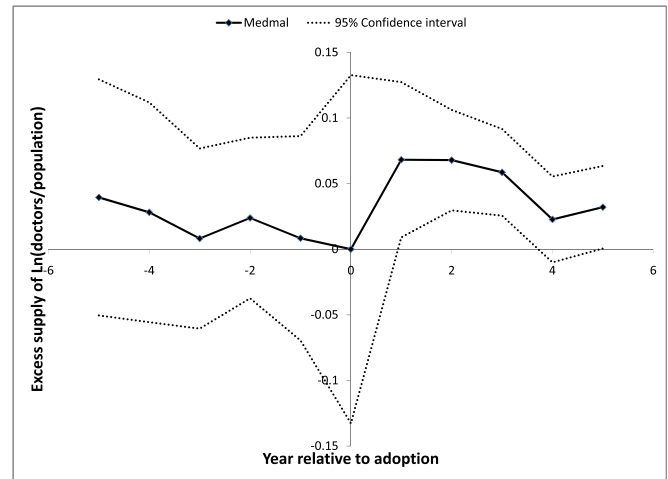
| Tort reform             | States enacting tort reform   | Proportion medical |
|-------------------------|---|--------------------|
| Collateral source       | <b>AL (87)</b> , <b>CO (87)</b> , CT (85), <b>HI (87)</b> , ID (90), IN (87), KY (89), <b>MA (87)</b> , <b>ME (90)</b> , MI (86), MN (85), MT (88), ND (88), NJ (88), <b>NY (85)</b> , OR (88), <b>UT (87)</b> , <b>WI (95)</b>   | 0.44               |
| Contingency fees        | CT (87), <b>FL (86)</b> , HI (87), IL (85), <b>MA (87)</b> , ME (89), <b>MI (85)</b> , NH (87), <b>UT (86)</b>  | 0.44               |
| Joint and several       | AK (86), AZ (87), CA (86), CO (87), CT (87), <b>FL (86)</b> , GA (88), HI (87), IA (84), ID (88), LA (81), <b>MI (87)</b> , MN (89), <b>MO (86)</b> , MS (90), MT (88), ND (88), NE (92), NH (90), NJ (88), NM (82), NY (87), TN (92), TX (86), UT (86), WA (86), WI (94), <b>WV (86)</b> , WY (86) | 0.14               |
| Noneconomic damage caps | AK (86), AL (87), <b>CO (87)</b> , <b>FL (89)</b> , HI (87), ID (87), IL (95), <b>KS (87)</b> , <b>MA (87)</b> , MD (87), <b>MI (87)</b> , MN (86), <b>MO (86)</b> , <b>MT (96)</b> , <b>ND (96)</b> , OR (88), <b>UT (88)</b> , WA (86), <b>WI (86)</b> , <b>WV (86)</b>                           | 0.55               |
| Periodic payment        | <b>AZ (89)</b> , <b>CO (89)</b> , CT (88), FL (87), IA (88), ID (88), <b>IL (86)</b> , <b>IN (85)</b> , <b>LA (85)</b> , MD (87), ME (87), MI (86), <b>MN (89)</b> , <b>MO (86)</b> , <b>MT (87)</b> , <b>NY (86)</b> , <b>OH (88)</b> , <b>RI (88)</b> , SD (88), <b>UT (86)</b> , <b>WA (86)</b>  | 0.62               |
| Punitive damage caps    | AK (98), AL (87), CO (87), FL (87), GA (88), <b>IL (85)</b> , IN (95), KS (88), NC (96), ND (93), NH (87), NJ (96), NV (89), OK (96), <b>OR (88)</b> , <b>PA (97)</b> , TX (88), <b>VA (89)</b> , <b>WI (85)</b>  | 0.26               |
| Punitive evidence       | AK (86), AL (87), AZ (87), CA (88), DC (96), <b>FL (00)</b> , GA (88), IA (87), ID (88), IN (84), KS (88), KY (89), MD (92), ME (85), MO (86), MS (94), MT (85), NC (96), ND (87), NJ (96), NV (89), OH (88), OK (87), OR (88), SC (88), TN (92), TX (88), UT (90), WI (95)                         | 0.03               |
| Split recovery          | AK (98), CO (87), <b>FL (87)</b> , IA (87), IN (96), OR (88), UT (90)   | 0.14               |
| Total damage caps       | <b>CO (89)</b> , <b>SD (86)</b>   | 1                  |
| Victims' fund           | <b>ND (83)</b>  | 1                  |

Year of enactment given in parentheses. Boldface indicates that the reform applies to medical malpractice torts only. The third column presents the proportion of all adopted reforms that applied to medical malpractice torts only.

even in states that allow liability insurance to cover punitive damages, many insurers refuse to do so. It is not surprising, therefore, that a number of papers that examine medical malpractice tort reform have found a significant effect of punitive damages on physician behavior (Avraham, 2007; Avraham et al., 2010; and Currie and MacLeod, 2008).

#### 5.4. Empirical strategy

We estimate the effect of caps on punitive damages on the log of physician supply using a difference-in-differences strategy. Treatment effects are identified by comparing within-state changes in physician supply among states that adopt reform to within-state changes in supply among states that do not adopt reform. It would be sufficient to include state and year fixed effects to implement our difference-in-differences estimator. However, we go further and employ state-specialty and specialty-year fixed effects. The former control for specialty-level unobservables within each state. The latter allow time paths for physician supply to vary across specialty, as Fig. 4 suggests that it may be appropriate.



**Fig. 5.** Excess physician supply before and after total damage caps: annual leads and lags from 5 years before to 5 years after adoption. This figure plots the normalized coefficients  $\lambda_j$  from the following regression:  $y_{ist} = \sum_{j=-5}^5 \lambda_j D_{st+j} + \gamma_{is} + \gamma_{st} + u_{ist}$ , where  $y_{ist}$  is the log of the physician count for specialty  $i$  in state  $s$  in year  $t$ ,  $D_{st+j}$  is an indicator for whether total damage caps was first adopted in period  $t+j$ , and  $\gamma_{is}$  and  $\gamma_{st}$  represent the state-specialty and specialty-year fixed effects.

We must select a pre- and a post-treatment period in order to implement our difference-in-differences design. We could use the entire 1980–2001 panel to calculate these contrasts but this is unappealing: observations from states that adopted reform early (late) would receive less weight in the pre (post) period than states that adopted reform later (earlier). Fig. 8 shows that all reforms adopted during our sample period occurred between 1984 and 1998 (the circled points). Given the 1980 beginning and 2001 end of our sample, we implement the widest window that permits full pre and post coverage for each treated state: a 9-year pre–post moving window that includes the 5 years preceding adoption of punitive damage caps and the 4 years following adoption.<sup>17</sup>

Our main identifying assumption is that the adoption of punitive damage caps is exogenous. We estimate several different specifications in order to increase the plausibility of this assumption. First, punitive damage caps are most likely to be exogenous in states where the reform is not targeted solely at medical malpractice cases (see Table 3). We therefore estimate specifications with and without these potentially endogenous states.<sup>18</sup> Second, even if punitive damage caps are exogenous, other tort reforms may not be, and their endogeneity could contaminate our estimates. We thus also estimate specifications that exclude these other tort reforms as controls, though this may introduce omitted variable bias if tort reforms are correlated. Third, although we have argued that punitive damage caps were adopted for reasons unrelated to physician supply, skeptics may still worry that states endogenously adopt reform in response to trends in physician supply. We rule out this possibility by also estimating specifications that include state trends.

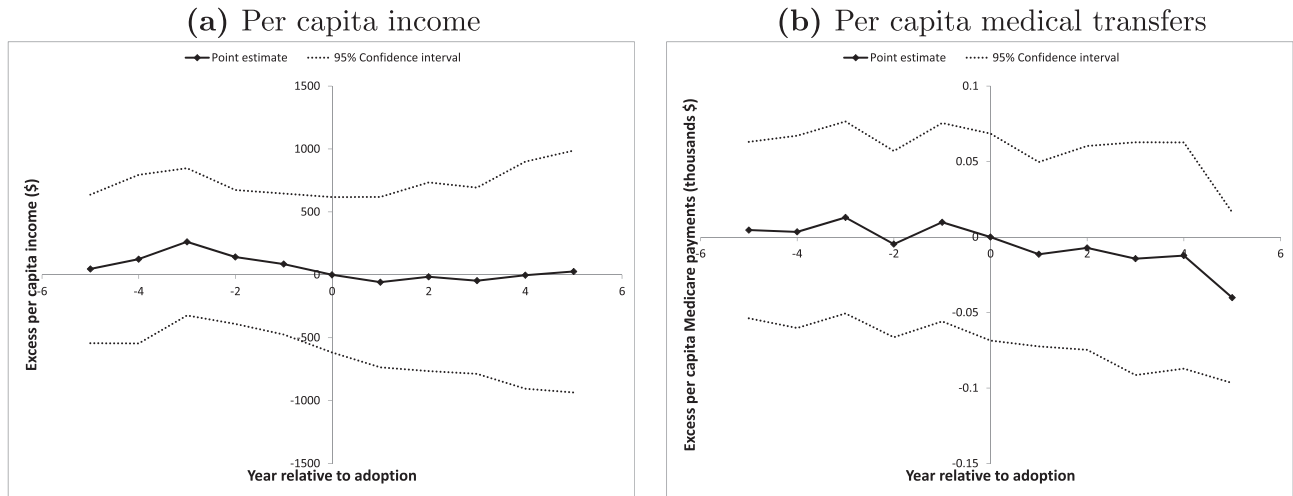
We begin by estimating a baseline myopic model:

$$y_{ist} = \beta_0^{myopic} d_{st} + \gamma_{is} + \gamma_{it} + \gamma_s \times t + u_{ist}. \quad (9)$$

The outcome  $y_{ist}$  is the log of the number of physicians per capita practicing specialty  $i$  in state  $s$  in period  $t$ ,  $d_{st}$  is an indicator for the presence of punitive damage caps in state  $s$  in period  $t$ , and  $\gamma_{is}$  and  $\gamma_{it}$  are the

<sup>17</sup> Appendix D presents results for alternative windows. Including all available years generates similar results. Including less than four years in the pre- and post-periods causes estimates from the quasi-myopic model to become insignificant. Including less than three years causes all estimates to become insignificant due to lack of statistical power.

<sup>18</sup> The potentially endogenous states are CO, IL, OR, PA, VA, and WI.



**Fig. 6.** Per capita income and medical transfer payments are uncorrelated with enactment of punitive damage caps. This figure plots the normalized coefficients  $\lambda_j$  from the following regression:  $y_{st} = \sum_{j=-5}^5 -\lambda_j D_{st+j} + \gamma_s + \gamma_t + u_{st}$ , where  $y_{st}$  is per capita income or medical transfers in state  $s$  in year  $t$ ,  $D_{st+j}$  is an indicator for whether punitive damage caps is adopted in period  $t+j$ , and  $\gamma_s$  and  $\gamma_t$  are the state and year fixed effects.

Source: REIS and Avraham (2010).

state-specialty and specialty-year fixed effects, respectively. Some specifications that we estimate also include state trends,  $\gamma_s \times t$ . The ex post effect of reform is estimated by  $\hat{\beta}_0^{myopic}$ . Because this model ignores anticipation, it implicitly assumes that the ex ante effect of reform is equal to zero.

Next, we turn to models that account for anticipation. The quasi-myopic model addresses anticipation by including leading indicators for whether reform is adopted:

$$y_{ist} = \beta_0^{quasi} d_{st} + \sum_{j=1}^S \beta_j^{quasi} D_{s,t+j} + \gamma_{is} + \gamma_{it} + \gamma_s \times t + u_{ist}. \quad (10)$$

We estimate four versions of Eq. (10) so that we can allow the number of leading indicators,  $S$ , to range from one to four.  $D_{s,t+j}$  is a dummy variable equal to 1 if a reform is adopted in time period  $t+j$ . For example, if a reform is adopted in period 5, then  $D_{s,t+1} = 1$  in period 4 and 0 otherwise. The coefficient  $\hat{\beta}_0^{quasi}$  identifies the ex post effect of treatment. The ex ante effect in period  $t-j$  is estimated by  $\hat{\beta}_j^{quasi}$ .

The quasi-myopic model replaces expectations of reform,  $E_t[D_{s,t+j}]$ , with realizations of reform,  $D_{s,t+j}$ , and thus implicitly assumes perfect foresight. It also assumes that expectations over a time horizon greater than length  $S$  do not matter. Estimation is inconsistent if either of these assumptions fails to hold.

An alternative solution is to assume that individuals discount the future exponentially and have rational expectations. This imposes no restrictions on the length of the time horizon for expectations and relaxes the assumption of perfect foresight:

$$y_{ist} = \theta y_{is,t+1} + \beta d_{st} + \delta d_{s,t+1} + \gamma_{is} + \gamma_{it} + \gamma_s \times t + w_{ist}. \quad (11)$$

The regressor  $d_{s,t+1}$  accounts for the binary nature of treatment, as described in Appendix C. The forecast errors contained in the error term,  $w_{ist}$ , cause the lead dependent variable,  $y_{is,t+1}$ , to be endogenous. We estimate Eq. (11) first using OLS, then using leads of the outcome variable ( $y_{t+j}$ ,  $j \geq 2$ ) as instruments for  $y_{is,t+1}$ , and finally using lags of the outcome variable ( $y_{t-j}$ ,  $j \geq 2$ ) as instruments. The key identifying assumption is that the lags and leads of physician supply are uncorrelated with unobserved determinants of current supply. This imposes a restriction on the degree of serial correlation in the error term. Arellano and Bond (1991) derive a test commonly used to check this assumption. We report those tests in our tables.

In the exponential discounting model, the ex post effect is equal to  $\hat{\beta}/(1-\hat{\theta})$  and the ex ante effect in period  $t-j$  is equal to  $\hat{\beta} \sum_{i=j}^{\infty} \hat{\theta}^i$ . Note that one can evaluate the statistical significance of the ex ante effect by testing the nonlinear null hypothesis  $\beta\theta = 0$ .<sup>19</sup> We report those test results in our tables.

A weak instrument test for panel GMM does not exist.<sup>20</sup> This is problematic because employing a large number of weak instruments can bias estimation. As a robustness check, we estimate additional IV specifications that employ only one instrument. This is less efficient than employing all available instruments, but also less prone to bias.

We weight all estimations by state population because  $y_{ist}$  is a per capita measure. Standard errors are clustered at the state level. We employ one-step GMM estimation when estimating the exponential discounting model to alleviate concerns about finite sample problems associated with two-step GMM estimation (Doran and Schmidt, 2006; Judson and Owen, 1999). Since we do not have data on physician counts for 1984 or 1990, we transform our data using forward orthogonal deviations instead of the usual first differences when estimating the exponential discounting models because this preserves sample size in panels with gaps (Arellano and Bover, 1995; Roodman, 2009).<sup>21</sup>

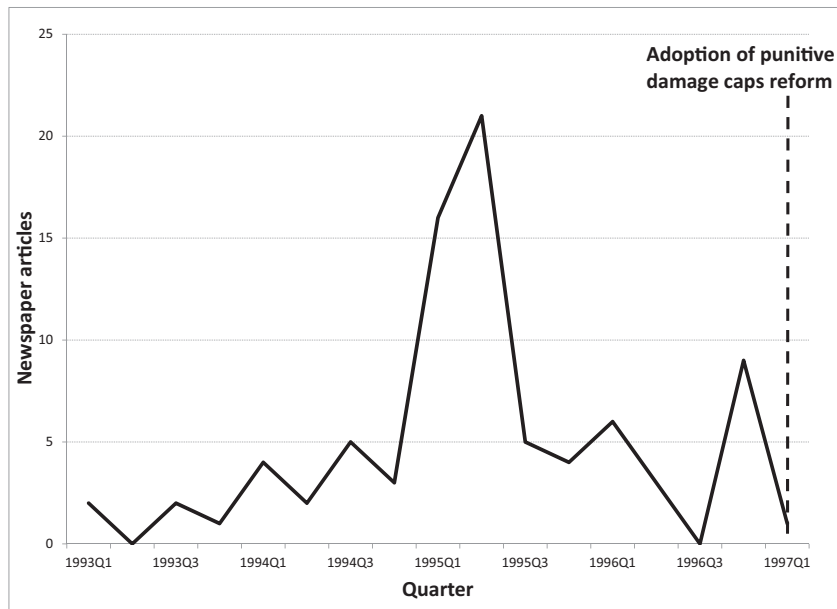
## 5.5. Results and discussion

We first report estimates from a myopic model – our baseline model (9) – that ignores expectations. Column 1 of Table 4 estimates that caps on punitive damages increases a state's equilibrium supply of physicians by 4.5% relative to states that did not adopt the reform. This is the myopic model's estimate of what we call the ex post effect. This estimate increases if we exclude potentially endogenous states (Column 2) and decreases slightly if we instead exclude all other tort reforms from the regression (Column 3). Columns 4 and 5 show that the estimated effect

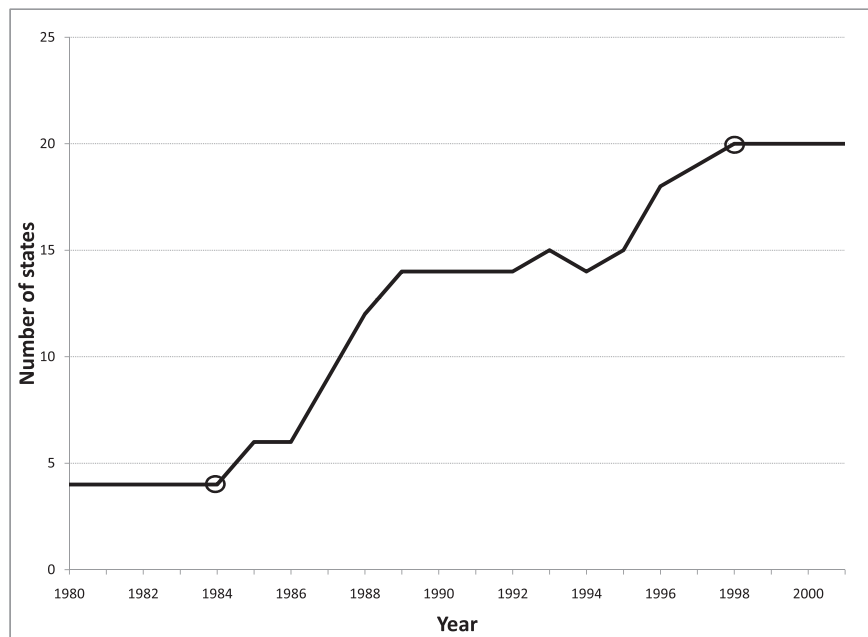
<sup>19</sup> Similarly, the statistical significance of the ex post effect can be evaluated by testing the null hypothesis  $\beta = 0$ . It is well known that Wald tests of nonlinear restrictions are not invariant to algebraically equivalent restrictions. Existing evidence suggests that researchers should use simple multiplicative forms when possible (Phillips and Park, 1988).

<sup>20</sup> Clemens and Bazzi (2009) suggest that researchers estimate a two-stage least squares model in these cases to ensure that their instruments are not radically weak. When we do this for Eq. (11) we obtain first-stage partial F statistics substantially larger than 10.

<sup>21</sup> Let  $y_{it}$  be a variable. Then the transformed version under forward orthogonal deviations is equal to  $y_{it}^* = c_{it} \left( y_{it} - \frac{1}{T_{it}} \sum_{j=t+1}^{T_{it}} y_{ij} \right)$ , where  $T_{it}$  is the number of available future observations and  $c_{it}$  is a scale factor equal to  $\sqrt{T_{it}/(T_{it}+1)}$ .



**Fig. 7.** Frequency of articles mentioning tort reform in Pennsylvania, by quarter. Pennsylvania adopted punitive damage caps reform on January 25, 1997. This figure shows the distribution over time of the 84 articles in *The Philadelphia Inquirer* and *The Pittsburgh Post-Gazette* mentioning the term “tort reform” during the four years prior to reform. Online archives are unavailable prior to 1993.



**Fig. 8.** Cumulative number of states implementing punitive damage caps.

decreases substantially if we include state-specific trends. The specification in Column 5 is the most robust to an endogeneity critique because it excludes endogenous states, omits other reforms from the regression, and includes state trends. We thus adopt it as our preferred specification for accounting for anticipation effects. For compactness, we confine subsequent estimates to the specifications presented in Columns 2 and 5.<sup>22</sup>

<sup>22</sup> Estimates for the specification presented in Column 1 are available in [Appendix D](#). Estimates for the other two specifications (Columns 3 and 4) are similar to those presented in the main text.

Estimates of the ex post effect from the quasi-myopic model are slightly larger, but generally yield insignificant estimates of the ex ante effect. We report those results in [Tables 5 and 6](#), which correspond to the specifications reported in Columns 2 and 5 from [Table 4](#), respectively. The first column in each table replicates the corresponding baseline estimate from [Table 4](#). Columns 2–5 in each table report estimates for versions of the quasi-myopic model, with each column sequentially adding one lead to the model. The coefficient estimates on the time- $t$  treatment variable identify the ex post effect of reform, including anticipation effects, and can be interpreted as relative changes in physician supply. Moving across the first row of [Table 5](#) shows that the estimated ex post effect increases monotonically from 7.1% to 8.9% as we add leads.



**Table 4**  
OLS estimates of baseline myopic model.

|   | (1)                | (2)               | (3)               | (4)              | (5)               |
|---|--------------------|-------------------|-------------------|------------------|-------------------|
| Ex post effect ( $\hat{\beta}_0^{myopic}$ ) | 0.045**<br>(0.022) | 0.071*<br>(0.035) | 0.034*<br>(0.020) | 0.006<br>(0.008) | 0.011*<br>(0.006) |
| Exclude IL, OR, PA, VA, WI                  | No                 | Yes               | No                | No               | Yes               |
| Exclude tort controls                       | No                 | No                | Yes               | No               | Yes               |
| State trends                                | No                 | No                | No                | Yes              | Yes               |
| Observations                                | 6493               | 6103              | 6493              | 6493             | 6103              |
| R-squared                                   | 0.990              | 0.990             | 0.990             | 0.993            | 0.992             |

This table reports estimates of Eq. (9). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the coefficient indicates significance at the 10/5% level.

**Table 6**, our preferred specification, reveals a similar pattern. It estimates an ex post treatment effect of 1.5%, which is slightly larger than the corresponding baseline estimate of 1.1%. As with the myopic model, introduction of state trends dramatically reduces treatment effects. The estimated ex ante effects in Column 5 of **Table 6** range from 0.002 to 0.005 and are insignificant.

The exponential discounting model (11) yields even higher estimates of the ex post effect and typically yields significant ex ante effects. Column 1 of **Table 7** reports OLS estimates of the exponential discounting model. Columns 2 and 3 instrument for  $y_{is,t+1}$  using leads and lags, respectively, of the outcome variable. The estimated coefficient on the lead dependent variable,  $\hat{\theta}$ , is significant across all three models. The estimated coefficient on the treatment variable,  $\hat{\beta}$ , is significant in the first two columns of **Table 7** but insignificant in Column 3, when we instrument for  $y_{is,t+1}$  using lags. The estimated ex post effect of treatment ranges from 6.1% to 7.7% is statistically significant except when we instrument using lags. **Table 8**, our preferred specification, reveals broadly similar results. It shows an estimated ex post effect of 1.8% when using leads as instruments. The estimated effect is again insignificant, however, when we instrument with lags. Finally, **Table 8** reports ex ante effects that range from 0.1% to 1.8% when leads are used as instruments. The Wald test of the hypothesis that  $\hat{\beta}\hat{\theta} = 0$  shows that these effects are statistically significant. When we employ lags as instruments, however, our point estimates fall close to zero and become insignificant.

To address the concern that the instruments we employ are over-fitting the model, we re-estimate the exponential discounting model (11) but use only one lead of the outcome variable as the instrument.

**Table 5**  
OLS estimates of myopic and quasi-myopic (QM) models: Exclude endogenous states, all controls, no trends.

|  | (1)               | (2)               | (3)               | (4)                | (5)                |
|--|-------------------|-------------------|-------------------|--------------------|--------------------|
| Ex post effect ( $\hat{\beta}_0^{quasi}$ )         | 0.071*<br>(0.035) | 0.074*<br>(0.037) | 0.080*<br>(0.040) | 0.083**<br>(0.041) | 0.089**<br>(0.044) |
| Ex ante effect (t – 1) ( $\hat{\beta}_1^{quasi}$ ) |                   | 0.016<br>(0.014)  | 0.022<br>(0.019)  | 0.026<br>(0.021)   | 0.031<br>(0.024)   |
| Ex ante effect (t – 2) ( $\hat{\beta}_2^{quasi}$ ) |                   |                   | 0.019<br>(0.015)  | 0.023<br>(0.017)   | 0.028<br>(0.021)   |
| Ex ante effect (t – 3) ( $\hat{\beta}_3^{quasi}$ ) |                   |                   |                   | 0.011<br>(0.009)   | 0.016<br>(0.013)   |
| Ex ante effect (t – 4) ( $\hat{\beta}_4^{quasi}$ ) |                   |                   |                   |                    | 0.015<br>(0.015)   |
| Model  | Myopic            | QM                | QM                | QM                 | QM                 |
| Exclude IL, PA, OR, VA, WI                         | Yes               | Yes               | Yes               | Yes                | Yes                |
| Exclude tort controls                              | No                | No                | No                | No                 | No                 |
| State trends                                       | No                | No                | No                | No                 | No                 |
| Observations                                       | 6103              | 6103              | 6103              | 6103               | 6103               |
| R-squared  | 0.990             | 0.990             | 0.990             | 0.990              | 0.990              |

This table reports estimates of Eq. (10). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the coefficient indicates significance at the 10/5% level.

**Table 6**  
OLS estimates of myopic and quasi-myopic (QM) models: Exclude endogenous states, no controls, trends.

|  | (1)               | (2)               | (3)               | (4)              | (5)               |
|--|-------------------|-------------------|-------------------|------------------|-------------------|
| Ex post effect ( $\hat{\beta}_0^{quasi}$ )         | 0.011*<br>(0.006) | 0.013*<br>(0.007) | 0.012<br>(0.009)  | 0.014<br>(0.009) | 0.015*<br>(0.009) |
| Ex ante effect (t – 1) ( $\hat{\beta}_1^{quasi}$ ) |                   | 0.002<br>(0.005)  | 0.002<br>(0.007)  | 0.003<br>(0.007) | 0.004<br>(0.006)  |
| Ex ante effect (t – 2) ( $\hat{\beta}_2^{quasi}$ ) |                   |                   | –0.000<br>(0.007) | 0.001<br>(0.007) | 0.002<br>(0.007)  |
| Ex ante effect (t – 3) ( $\hat{\beta}_3^{quasi}$ ) |                   |                   |                   | 0.004<br>(0.004) | 0.005<br>(0.005)  |
| Ex ante effect (t – 4) ( $\hat{\beta}_4^{quasi}$ ) |                   |                   |                   |                  | 0.002<br>(0.004)  |
| Model  | Myopic            | QM                | QM                | QM               | QM                |
| Exclude IL, PA, OR, VA, WI                         | Yes               | Yes               | Yes               | Yes              | Yes               |
| Exclude tort controls                              | Yes               | Yes               | Yes               | Yes              | Yes               |
| State trends                                       | Yes               | Yes               | Yes               | Yes              | Yes               |
| Observations                                       | 6103              | 6103              | 6103              | 6103             | 6103              |
| R-squared  | 0.992             | 0.992             | 0.992             | 0.992            | 0.992             |

This table reports estimates of Eq. (10). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the coefficient indicates significance at the 10/5% level.

We report those results in **Table 9**. A comparison with **Tables 7** and **8** reveals that employing only one instrument increases the estimated effect, which suggests that our previous estimates may have been biased downwards in magnitude. Our preferred specification, presented in Columns 5 and 6 of **Table 9**, estimates an ex post treatment effect of 2.2–2.6%. This is larger than the corresponding quasi-myopic estimate of 1.5% and more than double the estimated effect of 1.1% from the baseline myopic model.

The exponential discounting model assumes that treatment has a constant ex post effect over time and we estimate it assuming that agents have rational expectations. The first assumption is supported by **Fig. 1**. Although it displays a small visible decline after adoption of treatment, statistical tests (not reported) indicate that the decline is not significant.

**Table 7**  
OLS and IV estimates of exponential discounting model: exclude endogenous states, all controls, no trends.

|   | (1)                | (2)                | (3)                |
|---|--------------------|--------------------|--------------------|
| Discount rate ( $\hat{\theta}$ )  | 0.693**<br>(0.052) | 0.543**<br>(0.121) | 0.696**<br>(0.073) |
| Punitive damage caps ( $\hat{\beta}$ )                                    | 0.019*<br>(0.010)  | 0.035**<br>(0.015) | 0.023<br>(0.015)   |
| Calculated effects  |                    |                    |                    |
| Ex post effect ( $\hat{\beta}/(1-\hat{\theta})$ )                         | 0.061              | 0.077              | 0.075              |
| Ex ante effect (t – 1) ( $\hat{\beta}\sum_{i=1}^{\infty}\hat{\theta}^i$ ) | 0.042              | 0.042              | 0.052              |
| Ex ante effect (t – 2) ( $\hat{\beta}\sum_{i=2}^{\infty}\hat{\theta}^i$ ) | 0.029              | 0.023              | 0.036              |
| Ex ante effect (t – 3) ( $\hat{\beta}\sum_{i=3}^{\infty}\hat{\theta}^i$ ) | 0.020              | 0.012              | 0.025              |
| Ex ante effect (t – 4) ( $\hat{\beta}\sum_{i=4}^{\infty}\hat{\theta}^i$ ) | 0.014              | 0.007              | 0.018              |
| Wald test: $\hat{\beta}\hat{\theta} = 0$ (p-value)                        | 0.087              | 0.032              | 0.123              |
| IV  | None               | Leads              | Lags               |
| Exclude IL, PA, OR, VA, WI  | Yes                | Yes                | Yes                |
| Exclude tort controls   | No                 | No                 | No                 |
| State trends  | No                 | No                 | No                 |
| Observations  | 5139               | 4218               | 4859               |
| AR(3) test (p-value)  | N/A                | 0.897              | 0.956              |

This table reports OLS and IV estimates of Eq. (11). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the estimated coefficient  $\hat{\beta}$  or  $\hat{\theta}$  indicates significance at the 10/5% level. Null hypothesis of the AR(3) test is that the estimated residuals are not third-order autocorrelated.

**Table 8**

OLS and IV estimates of exponential discounting model: Exclude endogenous states, no controls, trends.

|   | (1)                | (2)                | (3)                |
|---|--------------------|--------------------|--------------------|
| Discount rate ( $\hat{\theta}$ )  | 0.598**<br>(0.044) | 0.495**<br>(0.117) | 0.489**<br>(0.080) |
| Punitive damage caps ( $\hat{\beta}$ )                                    | −0.001<br>(0.007)  | 0.009**<br>(0.004) | 0.001<br>(0.005)   |
| Calculated effects  |                    |                    |                    |
| Ex post effect ( $\hat{\beta}/(1-\hat{\theta})$ )                         | −0.002             | 0.018              | 0.003              |
| Ex ante effect (t − 1) ( $\hat{\beta}\sum_{i=1}^{\infty}\hat{\theta}^i$ ) | −0.001             | 0.009              | 0.001              |
| Ex ante effect (t − 2) ( $\hat{\beta}\sum_{i=2}^{\infty}\hat{\theta}^i$ ) | −0.001             | 0.004              | 0.001              |
| Ex ante effect (t − 3) ( $\hat{\beta}\sum_{i=3}^{\infty}\hat{\theta}^i$ ) | −0.000             | 0.002              | 0.000              |
| Ex ante effect (t − 4) ( $\hat{\beta}\sum_{i=4}^{\infty}\hat{\theta}^i$ ) | −0.000             | 0.001              | 0.000              |
| Wald test: $\hat{\beta}\hat{\theta} = 0$ (p-value)                        | 0.925              | 0.012              | 0.789              |
| IV  | None               | Leads              | Lags               |
| Exclude IL, PA, OR, VA, WI  | Yes                | Yes                | Yes                |
| Exclude tort controls   | Yes                | Yes                | Yes                |
| State trends  | Yes                | Yes                | Yes                |
| Observations  | 5139               | 4218               | 4859               |
| AR(3) test (p-value)  | N/A                | 0.868              | 0.913              |

This table reports OLS and IV estimates of Eq. (11). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the estimated coefficient  $\hat{\beta}$  or  $\hat{\theta}$  indicates significance at the 10/5% level. Null hypothesis of the AR(3) test is that the estimated residuals are not third-order autocorrelated.

Assessing the plausibility of the rational expectations assumption is more difficult, though our estimates do not cast doubt on it. A barometer of whether a rational expectations assumption is reasonable is the size of the estimated discount rate,  $\hat{\theta}$ . In many standard economic models, the discount rate is equal to  $1/(1+r)$ , where  $r$  is the interest rate. Although it is difficult to estimate precisely, an estimate of  $\hat{\theta}$  near 0 (implying an enormous interest rate) or greater than 1 (implying negative interest rates) indicates model misspecification, perhaps because agents do not have rational expectations. Column 6 from Table 9 estimates  $\hat{\theta} = 0.735$ , implying an interest rate of 36%. This is in the upper range of estimates from the literature (Warner and Pleeter, 2001).

**Table 9**

Estimates of exponential discounting model when employing only one IV.

|   | (1)                | (2)                | (3)                | (4)                | (5)                | (6)                |
|---|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Discount rate ( $\hat{\theta}$ )  | 0.593**<br>(0.144) | 0.719**<br>(0.176) | 0.563**<br>(0.172) | 0.673**<br>(0.229) | 0.639**<br>(0.133) | 0.735**<br>(0.151) |
| Punitive damage caps ( $\hat{\beta}$ )                                    | 0.032**<br>(0.012) | 0.027**<br>(0.012) | 0.034**<br>(0.015) | 0.026<br>(0.018)   | 0.008**<br>(0.003) | 0.007**<br>(0.003) |
| Calculated effects  |                    |                    |                    |                    |                    |                    |
| Ex post effect ( $\hat{\beta}/(1-\hat{\theta})$ )                         | 0.078              | 0.095              | 0.077              | 0.079              | 0.022              | 0.026              |
| Ex ante effect (t − 1) ( $\hat{\beta}\sum_{i=1}^{\infty}\hat{\theta}^i$ ) | 0.046              | 0.068              | 0.043              | 0.053              | 0.014              | 0.019              |
| Ex ante effect (t − 2) ( $\hat{\beta}\sum_{i=2}^{\infty}\hat{\theta}^i$ ) | 0.027              | 0.049              | 0.024              | 0.036              | 0.009              | 0.014              |
| Ex ante effect (t − 3) ( $\hat{\beta}\sum_{i=3}^{\infty}\hat{\theta}^i$ ) | 0.016              | 0.035              | 0.014              | 0.024              | 0.006              | 0.010              |
| Ex ante effect (t − 4) ( $\hat{\beta}\sum_{i=4}^{\infty}\hat{\theta}^i$ ) | 0.010              | 0.025              | 0.008              | 0.016              | 0.004              | 0.008              |
| Wald test: $\hat{\beta}\hat{\theta} = 0$ (p-value)                        | 0.019              | 0.014              | 0.028              | 0.052              | 0.004              | 0.011              |
| IV  | $Y_{t+2}$          | $Y_{t+3}$          | $Y_{t+2}$          | $Y_{t+3}$          | $Y_{t+2}$          | $Y_{t+3}$          |
| Exclude IL, PA, OR, VA, WI  | No                 | No                 | Yes                | Yes                | Yes                | Yes                |
| Exclude tort controls   | No                 | No                 | No                 | No                 | Yes                | Yes                |
| State trends  | No                 | No                 | No                 | No                 | Yes                | Yes                |
| Observations  | 4488               | 4488               | 4218               | 4218               | 4218               | 4218               |
| AR(3) test (p-value)  | 0.899              | 0.923              | 0.904              | 0.926              | 0.916              | 0.931              |

This table reports IV estimates of Eq. (11). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the estimated coefficient  $\hat{\beta}$  or  $\hat{\theta}$  indicates significance at the 10/5% level. Null hypothesis of the AR(3) test is that the estimated residuals are not third-order autocorrelated.

**Table 10**

OLS estimates of myopic and quasi-myopic (QM) models: All states, all controls, no trends.

|  | (1)                | (2)               | (3)               | (4)                | (5)                |
|--|--------------------|-------------------|-------------------|--------------------|--------------------|
| Ex post effect ( $\hat{\beta}_0^{quasi}$ )         | 0.045**<br>(0.022) | 0.048*<br>(0.024) | 0.053*<br>(0.026) | 0.059**<br>(0.028) | 0.063**<br>(0.031) |
| Ex ante effect (t − 1) ( $\hat{\beta}_1^{quasi}$ ) |                    | 0.013<br>(0.012)  | 0.019<br>(0.016)  | 0.025<br>(0.018)   | 0.029<br>(0.021)   |
| Ex ante effect (t − 2) ( $\hat{\beta}_2^{quasi}$ ) |                    |                   | 0.019<br>(0.012)  | 0.025*<br>(0.014)  | 0.030<br>(0.018)   |
| Ex ante effect (t − 3) ( $\hat{\beta}_3^{quasi}$ ) |                    |                   |                   | 0.017**<br>(0.008) | 0.022*<br>(0.011)  |
| Ex ante effect (t − 4) ( $\hat{\beta}_4^{quasi}$ ) |                    |                   |                   |                    | 0.010<br>(0.010)   |
| Model  | Myopic             | QM                | QM                | QM                 | QM                 |
| Exclude IL, PA, OR, VA, WI                         | No                 | No                | No                | No                 | No                 |
| Exclude tort controls                              | No                 | No                | No                | No                 | No                 |
| State trends                                       | No                 | No                | No                | No                 | No                 |
| Observations                                       | 6493               | 6493              | 6493              | 6493               | 6493               |
| R-squared  | 0.990              | 0.990             | 0.990             | 0.990              | 0.990              |

This table reports estimates of Eq. (10). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the coefficient indicates significance at the 10/5% level.

We explained in Section 3.1 that the estimated treatment effect in a myopic model often will typically underestimate the true ex post effect of reform. Including leads can mitigate this problem by reducing omitted variable bias. Tables 5 and 6 show that the estimated effects from a quasi-myopic model are larger than the corresponding estimates from the myopic model, as predicted. The fact that these estimates increase as we keep adding leads to the quasi-myopic model suggests that each additional lead moves us closer to an unbiased estimate of the ex post effect. The exponential discounting model yields even larger estimates, suggesting that anticipation effects may matter for longer than the data permit in the quasi-myopic model.

That said, the differences between the estimated ex post effects for any two particular specifications are not statistically significant due to the large standard errors. Furthermore, the similarity of the point estimates from the quasi-myopic model and the exponential discounting model suggests that the choice between these two parameterizations of expectations is not critical to estimates of anticipation effects from punitive damage caps. However, some of the quasi-myopic models,

and nearly all the exponential discounting models, estimate a statistically significant ex ante effect. This differs from the myopic model, which implicitly assumes that it is equal to zero.

Like other prior studies on this topic, we do not account for general equilibrium effects. A physician fleeing one state necessarily enters another, magnifying the relative supply differences between the two states. Kessler et al. (2005) have previously demonstrated, however, that most of the equilibrium adjustment comes from newly graduated residents deciding where to practice and retirees leaving practice. Furthermore, we are primarily interested in the *relative* differences between our model estimates, for it is these relative differences that reveal the importance of anticipation effects.

Our results are consistent with Currie and MacLeod (2008), who find that joint and several liability and punitive damage caps have a significant effect on birth outcomes. Our estimated effects, which range from about 2 to 6%, are similar in magnitude to other studies of tort reform and physician supply. For example, Kessler et al. (2005) estimate an effect of 3% for a broad set of reforms, Klick and Stratmann (2007) estimate an effect of 6% for noneconomic damage caps, and Matsa (2007) estimates an effect of 10% for total damage caps.

## 6. Conclusion

Researchers are confronted with forward-looking agents in a variety of different applications, including durable goods purchases, rational addiction, and human and physical capital investments. In all these cases, anticipation causes outcomes to respond to future treatment. Incorrectly assuming that this change is due to endogeneity rather than anticipation will produce biased estimation. While not all economic decisions are made with an eye towards the future and not all shocks are anticipated, enough are that empirical work should consider how to define and estimate treatment effects in the context of anticipation effects.

The framework that we introduce in this paper for comparing and estimating different models of anticipation effects has a number of limitations. Foremost, we offer no clean test that distinguishes between anticipation and endogeneity. Our application merely provides informal evidence in favor of anticipation and thus our framework. Further, within our framework, we offer no formal way to discriminate between the different sets of parametric restrictions that we discuss. There may be

**Table 11**  
OLS and IV estimates of exponential discounting model: all states, all controls, no trends.

|   | (1)                | (2)                | (3)                |
|---|--------------------|--------------------|--------------------|
| Discount rate ( $\hat{\theta}$ )  | 0.692**<br>(0.050) | 0.544**<br>(0.117) | 0.706**<br>(0.070) |
| Punitive damage caps ( $\hat{\beta}$ )                                    | 0.021**<br>(0.009) | 0.034**<br>(0.012) | 0.023*<br>(0.012)  |
| Calculated effects  |                    |                    |                    |
| Ex post effect ( $\hat{\beta}/(1-\hat{\theta})$ )                         | 0.067              | 0.074              | 0.078              |
| Ex ante effect (t – 1) ( $\hat{\beta}\sum_{i=1}^{\infty}\hat{\theta}^i$ ) | 0.046              | 0.040              | 0.055              |
| Ex ante effect (t – 2) ( $\hat{\beta}\sum_{i=2}^{\infty}\hat{\theta}^i$ ) | 0.032              | 0.022              | 0.039              |
| Ex ante effect (t – 3) ( $\hat{\beta}\sum_{i=3}^{\infty}\hat{\theta}^i$ ) | 0.022              | 0.012              | 0.028              |
| Ex ante effect (t – 4) ( $\hat{\beta}\sum_{i=4}^{\infty}\hat{\theta}^i$ ) | 0.015              | 0.006              | 0.019              |
| Wald test: $\hat{\beta}\hat{\theta} = 0$ (p-value)                        | 0.036              | 0.021              | 0.082              |
| IV  | None               | Leads              | Lags               |
| Exclude IL, PA, OR, VA, WI  | No                 | No                 | No                 |
| Exclude tort controls   | No                 | No                 | No                 |
| State trends  | No                 | No                 | No                 |
| Observations  | 5479               | 4488               | 5189               |
| AR(3) test (p-value)  | N/A                | 0.883              | 0.955              |

This table reports OLS and IV estimates of Eq. (11). Dependent variable is log of count of high-risk physicians per 100,000 population. All estimates include state-specialty and specialty-year fixed effects. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the estimated coefficient  $\hat{\beta}$  or  $\hat{\theta}$  indicates significance at the 10/5% level. Null hypothesis of the AR(3) test is that the estimated residuals are not third-order autocorrelated.

**Table 12**  
Coefficient estimates on the time-t treatment effect for different pre-post windows.

|  | (1)                | (2)                | (3)                | (4)                | (5)                | (6)               |
|--|--------------------|--------------------|--------------------|--------------------|--------------------|-------------------|
| Myopic ( $\hat{\beta}^{myopic}$ )                              | 0.019**<br>(0.008) | 0.011*<br>(0.007)  | 0.011*<br>(0.006)  | 0.009<br>(0.006)   | 0.008<br>(0.005)   | –0.001<br>(0.007) |
| Quasi-myopic ( $\hat{\beta}_0^{quasi}$ )                       | 0.021*<br>(0.011)  | 0.012<br>(0.007)   | 0.019**<br>(0.008) | 0.011<br>(0.012)   | 0.009<br>(0.011)   | 0.007<br>(0.014)  |
| Exponential discounting<br>(IV = leads) ( $\hat{\beta}$ )      | 0.010**<br>(0.002) | 0.008**<br>(0.003) | 0.008**<br>(0.003) | 0.009**<br>(0.003) | 0.008**<br>(0.004) | –0.022<br>(0.018) |
| Exponential discounting<br>(IV = lags) ( $\hat{\beta}$ )       | 0.009**<br>(0.003) | 0.000<br>(0.006)   | 0.001<br>(0.007)   | 0.002<br>(0.008)   | –0.005<br>(0.009)  | –0.008<br>(0.012) |
| Exponential discounting<br>(IV = $y_t + 2$ ) ( $\hat{\beta}$ ) | 0.008**<br>(0.002) | 0.006**<br>(0.003) | 0.006**<br>(0.003) | 0.008**<br>(0.002) | 0.007<br>(0.004)   | –0.025<br>(0.021) |
| Exponential discounting<br>(IV = $y_t + 3$ ) ( $\hat{\beta}$ ) | 0.008**<br>(0.003) | 0.006*<br>(0.003)  | 0.005<br>(0.004)   | 0.006*<br>(0.003)  | 0.006<br>(0.005)   | –0.026<br>(0.022) |
| Exclude IL, PA, OR, VA, WI                                     | Yes                | Yes                | Yes                | Yes                | Yes                | Yes               |
| Exclude tort controls  | Yes                | Yes                | Yes                | Yes                | Yes                | Yes               |
| State trends   | Yes                | Yes                | Yes                | Yes                | Yes                | Yes               |
| Pre-post window  | All<br>years       | 4–4                | 4–3                | 3–3                | 3–2                | 2–2               |

This table displays estimates for different pre-post windows. Column (1) reports estimates when we employ all available years. Columns (2) through (6) limit the pre- and post-treatment periods to varying degrees ranging from 2 to 4 years. Standard errors, given in parentheses, are clustered by state. An \*/\*\* next to the coefficient indicates significance at the 10/5% level.

other restrictions a researcher might employ or estimation strategies that do not require any restrictions at all. For example, if two agents are both treated but one found out about the treatment earlier than the other, one could estimate anticipation effects with a difference-in-differences estimator that would eliminate many expectations terms. Likewise, there may be alternative models of updating or belief formation that can be employed. Ideally the researcher would directly survey agents about their expectations or at least survey a subsample to empirically estimate the relationship between expectations and observables. Each of these limitations is a useful topic for future research.

## Appendix A. Adaptive expectations

In this section we derive a model under the assumption that agents have adaptive expectations and show how to estimate it. One can show that the object of the agent's expectations (outcomes or treatment) does not affect identification of  $\beta$  or  $\theta$ . For ease of exposition we assume that agents have adaptive expectations about outcomes:

$$E_t[y_{t+1}] = E_t[y_{t+j}] = \phi y_t + (1-\phi)E_{t-1}[y_t].$$

Plugging these equations into the estimation equation

$$y_t = \theta E_t[y_{t+1}] + \beta d_t + e_t \quad (12)$$

and simplifying yields

$$y_t = \theta \phi y_t + \theta(1-\phi)E_{t-1}[y_t] + \beta d_t + \varepsilon_t. \quad (13)$$

The one-step back version of Eq. (12) is:

$$y_{t-1} = \theta E_{t-1}[y_t] + \beta d_{t-1} + \varepsilon_{t-1}.$$

Solve this for  $E_{t-1}[y_t]$  and plug the result into Eq. (13). Simplifying then produces the estimable equation

$$y_t = \gamma(1-\phi)y_{t-1} + \gamma\beta d_t - \gamma\beta(1-\phi)d_{t-1} + \gamma\varepsilon_t - \gamma(1-\phi)\varepsilon_{t-1}$$

where  $\gamma \equiv 1/(1-\theta\phi)$ . Time- $t$  outcomes are now a function of past rather than future outcomes. The reason is that adaptive

expectations is a backward-looking model of learning. The coefficient on current treatment no longer directly identifies  $\beta$ , though that parameter can be identified. Finally, the only source of endogeneity is previous period model error:  $E[y_{t-1}\varepsilon_{t-1}] \neq 0$ . Estimation is therefore straight-forward (so long as there is no serial correlation in  $\varepsilon_t$ ): use lags of order three or deeper and/or leads of order one or higher as instruments.

## B. Rational expectations about outcomes

Consider the case where the agent is able to form rational expectations about outcomes.<sup>23</sup> Initially we assume that realizations are a function of expectations:  $y_{t+j} = E_t[y_{t+j}] + v_{t,t+j}^y$ , where  $v_{t,t+j}^y$  indicates the error given time  $t$  expectations about outcomes in time  $t+j$  and  $E_t[E_t[y_{t+j}]v_{t,t+j}^y] = 0$ . This model is appropriate, for example, when outcomes are stock prices since realizations of stock prices are a composite of expectations (e.g., Chow, 1989). Expectations at time  $t$  about  $\theta_{y+1}$  are

$$\theta E_t[y_{t+1}] = E_t\left[\theta\beta\sum_{j=0}^{\infty}\theta^j E_{t+1}[d_{t+1+j}]\right] \quad (14)$$

since  $E_t[e_{t+1}] = 0$ . Subtracting Eq. (14) from Eq. (4) yields

$$y_t = \theta E_t[y_{t+1}] + \beta d_t + e_t. \quad (15)$$

Plugging in our rational expectations assumption produces the estimation equation

$$y_t = \theta y_{t+1} + \beta d_t + w_t \quad (16)$$

where  $w_t = e_t - \theta v_{t,t+1}^y$ .

The error term has two components, model error ( $e_t$ ) and unexpected, mean-zero forecast error ( $v_{t,t+1}^y$ ), which cause outcomes to deviate from forecasts. Thus rational expectations introduces measurement error. The result is endogeneity between next period's outcome  $y_{t+1}$  and  $v_{t,t+1}^y$ . Furthermore, if  $\{d_t\}$  are serially correlated, then  $d_t$  would be correlated with  $y_{t+1}$  and thus  $v_{t,t+1}^y$  through  $d_{t+1}$ .

If we had instead assumed that expectations about outcomes were a function of actual outcomes, i.e.,  $E_t[y_{t+j}] = y_{t+j} + v_{t,t+j}^y$  where  $E[y_{t+j}v_{t,t+j}^y] = 0$ , then there would be no endogeneity. The Euler equation would look the same, but  $w_t = e_t + \theta v_{t,t+1}^y$ . By assumption  $y_{t+1}$  is exogenous, and even with serial correlation in  $\{d_t\}$  we would have  $E[d_t v_{t+1}^y] \neq 0$ .

Estimating this model is straightforward. Suppose that expectations are a function of outcomes and  $E_t[y_{t+1}] = y_{t+1} + v_{t,t+1}^y$ .<sup>24</sup> Our error term is

$$w_t = e_t + \theta v_{t,t+1}^y.$$

The analogue to Assumption A2' is that, for some constant  $H$ ,

$$E[w_t w_{t+j}] = E\left[(e_t - \theta v_{t,t+1}^y)(e_{t+j} - \theta v_{t+j,t+1}^y)\right] = 0 \forall j > H.$$

If  $e_t$  and  $v_t^y$  are serially and mutually uncorrelated then the usual difference and system GMM estimators can be used so long as Assumptions A1 and A3 hold. Limited correlation in the error term can be accommodated by using higher order leads.

<sup>23</sup> We ignore the role of covariates  $x_t$  to simplify the exposition. However, it is straight-forward to incorporate covariates into the analysis.

<sup>24</sup> If outcomes are a function of expectations ( $y_{t+1} = E_t[y_{t+1}] + v_{t,t+1}^y$ ), then  $y_{t+1}$  is exogenous and no instruments are required.

## C. Binary treatment variables

In many applications the treatment variable,  $d_t$ , is binary. The forecast error corresponding to rational expectations of a binary variable is necessarily mean reverting, which induces a negative correlation between  $d_{t+j}$  and  $v_{t,t+j}^d$ .<sup>25</sup> If treatment states are correlated over time then endogeneity occurs because  $E[d_t v_{t,t+j}^d] \neq 0$  and the error term will be serially correlated, violating Assumption A2' from Section 4.1. This problem can be resolved if agent forecasts follow a Markov process because then future treatment states can be used to absorb the endogeneity. More specifically, suppose that

$$\text{Cov}[d_t, v_{t,t+j}^d | d_{t+1}] = 0 \forall t, \forall j > 1.$$

For rational expectations, this assumption can be tested by regressing  $d_t$  on lags of  $d_t$  and showing that, conditional on  $d_{t-1}$ , deeper lags are not correlated with  $d_t$ . If the assumption holds, one can then consistently estimate the equation

$$y_t = \theta y_{t+1} + \beta d_t + \delta d_{t+1} + w_t$$

where  $d_t$  is binary and  $\delta$  is a nuisance parameter. For interested readers, Monte Carlo simulations demonstrating the consistency of this estimator are available upon request.

We verified that adoption of tort reform follows a Markov process by estimating the following state-level regression for each tort reform listed in Table 2:  $d_{st} = \alpha_1 x_{st} + \alpha_2 d_{st-1} + \alpha_3 d_{st-2} + e_{st}$ , where  $d_{st}$  is a reform and  $x_{st}$  is a vector of controls that includes all other tort reforms. Our results (not reported) show that, at a 5% level of significance,  $\alpha_2$  is significant for all ten tort reforms while  $\alpha_3$  is insignificant for nine of them. These regressions clustered standard errors at the state level and were unweighted.

## D. Robustness checks

Tables 10 and 11 report results for the quasi-myopic and exponential models when our specification includes all states, all tort controls, and no state trends. The estimates are similar to the other tables that focus on the states where reforms are more likely to be exogenous (Tables 5 and 7).

The model estimated in the main text employed the widest window that permitted full pre and post coverage for each treated state: a 9-year pre-post moving window that included the 5 years preceding adoption of punitive damage caps and the 4 years following adoption. In this section, we show that our results are not substantively affected if we employ alternative windows.

Table 12 displays estimates from all of our models using five different definitions for the window. For compactness, we present results only for our preferred specification (exclude endogenous states, omit other tort reforms from the regression, and include state trends) and only for the coefficient on the time- $t$  treatment variable. The first column of Table 12 presents results when we employ all available years for both the pre and the post period. The second column presents results when we employ four years of data in both the pre and post periods. Columns (3)–(6) sequentially shorten the number of years in both the pre and post periods.

Estimates from the quasi-myopic model continue to be larger in magnitude than the corresponding estimates from the myopic model, regardless of whether the pre-post window includes all available years (Column 1) or fewer years (Columns 2–6). As we decrease the number of available years in the window, the standard errors increase, which eventually causes estimates to become insignificant.

<sup>25</sup> Recall that  $v_{t,t+j}^d$  is defined as the forecast error from the time- $t$  forecast of  $d_{t+j}$ .



## References

- Acemoglu, D., Linn, J., 2004. Market size in innovation: theory and evidence from the pharmaceutical industry. *Q. J. Econ.* 119 (3), 1049–1090.
- Alpert, A., 2010. The anticipatory effects of medicare part D on drug utilization. Working Paper. University of Maryland.
- AMA, 1997. Appendix II: graduate medical education. *J. Am. Med. Assoc.* 278 (9), 775–776.
- Anderson, S.T., Kellogg, R., Sallee, J.M., 2013. What do consumers believe about future gasoline prices? *J. Environ. Econ. Manag.* 66 (3), 383–403.
- Angrist, J.D., Pischke, J.-S., 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.* 58 (2), 277–297.
- Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. *J. Econ.* 68 (1), 29–51.
- Ashenfelter, O., 1978. Estimating the effect of training programs on earnings. *Rev. Econ. Stat.* 47–57.
- Autor, D.H., Donohue, J.J., Schwab, S.J., 2006. The costs of wrongful-discharge laws. *Rev. Econ. Stat.* 88 (2), 211–231.
- Avraham, R., 2007. An empirical study of the impact of tort reforms on medical malpractice settlement payments. *J. Leg. Stud.* 36 (S2), S183–S229.
- Avraham, R., 2010. Database of State Tort Law Reforms (DSTLR 3rd). SSRN eLibrary.
- Avraham, R., Dafny, L., Schanzenbach, M., 2010. The impact of tort reform on employer-sponsored health insurance premiums. *J. Law Econ. Org.* <http://jleo.oxfordjournals.org/content/early/2010/12/30/jleo.ewq017.short>.
- Ayers, B.C., Cloyd, C.B., Robinson, J.R., 2005. “Read my lips...”: does the tax rhetoric of presidential candidates affect security prices? *J. Law Econ.* 48 (1), 125–148.
- Becker, G.S., Grossman, M., Murphy, K.M., 1994. An empirical analysis of cigarette addiction. *Am. Econ. Rev.* 84 (3), 396–418.
- Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? *Q. J. Econ.* 119 (1), 249–275.
- Bhattacharya, J., Vogt, W.B., 2003. A simple model of pharmaceutical price dynamics. *J. Law Econ.* 46 (2), 599–626.
- Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *J. Econ.* 87 (1), 115–143.
- Blundell, R., Francesconi, M., Van der Klaauw, W., 2010. Anatomy of policy reform evaluation: announcement and implementation effects. Technical Report, Working Paper.
- Born, P., Viscusi, W.K., Baker, T., 2006. The effects of tort reform on medical malpractice insurers' ultimate losses. Working Paper 12086. National Bureau of Economic Research.
- Carroll, C.D., 2003. Macroeconomic expectations of households and professional forecasters. *Q. J. Econ.* 118 (1), 269–298.
- Chou, C., Lo Sasso, A., 2009. Practice location choice by new physicians: the importance of malpractice premiums, damage caps, and health professional shortage area designation. *Health Serv. Res.* 44 (4), 1271–1289.
- Chow, G.C., 1989. Rational versus adaptive expectations in present value models. *Rev. Econ. Stat.* 71 (3), 376–384.
- Clemens, M., Bazzi, S., 2009. Blunt instruments: on establishing the causes of economic growth. Center for Global Development Working Paper (171).
- Cohen, T., 2004. Medical Malpractice Trials and Verdicts in Large Counties, 2001. US Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics.
- Cohen, T., Harbacek, K., 2011. Punitive Damage Awards in State Courts, 2005. US Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics.
- Currie, J., MacLeod, W.B., 2008. First do no harm? Tort reform and birth outcomes. *Q. J. Econ.* 123 (2), 795–830.
- Doran, H.E., Schmidt, P., 2006. GMM estimators with improved finite sample properties using principal components of the weighting matrix, with an application to the dynamic panel data model. *J. Econ.* 133 (1), 387–409.
- Finkelstein, A., 2004. Static and dynamic effects of health policy: evidence from the vaccine industry. *Q. J. Econ.* 119 (2), 527–564.
- Gruber, J., Koszegi, B., 2001. Is addiction “rational”? Theory and evidence. *Q. J. Econ.* 116 (4), 1261–1303.
- Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50 (4), 1029–1054.
- Helland, E., Klick, J., Tabarrok, A., 2005. Data watch: tort-uring the data. *J. Econ. Perspect.* 19 (2), 207–220.
- Holtz-Eakin, D., 2004. The Effects of Tort Reform: Evidence from the States. Congressional Budget Office.
- Hubbard, F., 2006. The nature and impact of the ‘tort reform’ movement. *Hofstra Law Rev.* 35, 437–538.
- Judson, R.A., Owen, A.L., 1999. Estimating dynamic panel data models: a guide for macroeconomists. *Econ. Lett.* 65 (1), 9–15.
- Kahn, J.A., 1986. Gasoline prices and the used automobile market: a rational expectations asset price approach. *Q. J. Econ.* 323–339.
- Kessler, D.P., Sage, W.M., Becker, D.J., 2005. Impact of malpractice reforms on the supply of physician services. *J. Am. Med. Assoc.* 293, 2618–2625.
- Klick, J., Stratmann, T., 2007. Medical malpractice reform and physicians in high-risk specialties. *J. Leg. Stud.* 36 (S2), S121–S142.
- Lueck, D., Michael, J.A., 2003. Preemptive habitat destruction under the endangered species act. *J. Law Econ.* 46 (1), 27–60.
- Matsa, D.A., 2007. Does malpractice liability keep the doctor away? Evidence from tort reform damage caps. *J. Leg. Stud.* 36 (S2), S143–S182.
- McCallum, B.T., 1976. Rational expectations and the natural rate hypothesis: some consistent estimates. *Econometrica* 44 (1), 43–52.
- McCullough, Campbell, Lane, L.L.P., 2004. The Insurability of Punitive Damages.
- Mertens, K., Ravn, M.O., 2011. Understanding the aggregate effects of anticipated and unanticipated tax policy shocks. *Rev. Econ. Dyn.* 14, 27–54.
- Nockleby, J., Curreri, S., 2005. 100 Years of Conflict: The Past and Future of Tort Retrenchment. *Loyola of Los Angeles Law Review*, p. 38.
- Phillips, P.C., Park, J.Y., 1988. On the formulation of Wald tests of nonlinear restrictions. *Econometrica J. Econ. Soc.* 1065–1083.
- Poterba, J.M., 1984. Tax subsidies to owner-occupied housing: an asset-market approach. *Q. J. Econ.* 99 (4), 729–752.
- Rabin, R., 1988. Some reflections on the process of tort reform. *San Diego L. Rev.* 25, 13–48.
- Roodman, D., 2009. How to do xtabond2: an introduction to difference and system GMM in stata. *Stata J.* 9 (1), 86–136.
- Ryoo, J., Rosen, S., 2004. The engineering labor market. *J. Polit. Econ.* 112 (S1), S110–S140.
- Siegel, R., 1995. Aarp, Madd Oppose Tort-Reform Proposals They're Against Punitive Damage Caps, Changes in Liability. *Whitman Favors the Controversial Reforms*. *The Philadelphia Inquirer*.
- The Economist, 2005. Scalpel, scissors, lawyer. December 14. *De Economist*.
- Viscusi, W., Born, P., 2005. Damages caps, insurability, and the performance of medical malpractice insurance. *J. Risk Insur.* 72 (1), 23–43.
- Warner, J.T., Pleeter, S., 2001. The personal discount rate: evidence from military downsizing programs. *Am. Econ. Rev.* 33–53.
- Wilson, Elser, Moskowitz, Edelman, Dicker, L.L.P., 2008. Punitive Damages Review.
- Zeiler, K., Silver, C., Black, B., Hyman, D., Sage, W., 2007. Physicians' insurance limits and malpractice payments: evidence from Texas closed claims, 1990–2003. *J. Leg. Stud.* 36 (2), 9.