

**Research Article****Special Issue on Econometrics and Business Analytics**

Galina Besstremyannaya\* and Sergei Golovan

# Disentangling the impact of mean reversion in estimating policy response with dynamic panels

<https://doi.org/10.1515/demo-2022-0104>

received July 17, 2021; accepted March 29, 2022

**Abstract:** This article accounts for multivariate dependence of the variable of policy interest in dynamic panel data models by disentangling the two sources of intertemporal dependence: one from the effect of the policy variable and the other from mean reversion. In a situation where intensity of the policy varies over time, we estimate the unconditional mean in the autoregressive process as a function of the agent's characteristics and the policy intensity. Comparison of the fitted values of the unconditional mean under different values of the policy intensity enables identification of the policy effect cleared of mean reversion. The approach is relevant for measuring the effect of reforms, which use an intertemporal incentive where intensity of the reform varies over time. The empirical part of the article assesses the effect of hospital financing reform based on incentive contracts, related to the observed quality of services at Medicare hospitals in 2013–2019. We find a direct association between prior quality and quality improvement owing to the reform. Our result reassesses a stylized fact in the literature, which asserts that a pay-for-performance incentive leads to greater improvements at hospitals with lower baseline quality.

**Keywords:** policy evaluation, autoregressive process, intertemporal optimization, incentive contract, hospitals

**MSC 2020:** 62J05, 91B55, 91B69, 91B74

## 1 Introduction

The phenomenon of regression toward the mean (mean reversion) is observed in case of longitudinal observations of a variable, which is susceptible to random variations. In this case, exceptionally low or high values of the variable in initial measurement tend to be closer to the center of the distribution in subsequent measurements [24]. In short, mean reversion is an inherent part of the stationary process and implies the return of the process to its mean value [25,31].<sup>1</sup>

Historically, the appearance of the term “mean reversion” is associated with the seminal works by Galton, who discovered an inverse relationship between the height of parents and children [30] and hence framed the term “regression” as the tendency of the dependent variable to revert to the mean value. Recent examples of the analysis of processes which exhibit mean reversion in various fields of economics include

---

<sup>1</sup> For instance, mean reversion may manifest itself in the intersection between the mean and the trajectory for the process as it varies over time [25] or in the return of the autoregressive process to the long-term mean [31].

---

\* **Corresponding author: Galina Besstremyannaya**, National Research University Higher School of Economics, Moscow, Pokrovsky Blvd 11, Russian Federation, e-mail: gbestremyannaya@hse.ru

**Sergei Golovan:** New Economic School, Skolkovskoe shosse 45, Moscow, Russia, e-mail: sgolovan@nes.ru

the current account of countries [81] and their productivity [29], profitability of banks [48], housing prices [31], tax avoidance by companies [3], blood pressure and cholesterol level of patients [5], and birthweight of children in successive pregnancies of the same mother [79].

Mean reversion contaminates judgment about the time profile of the dependent variable in case of groupwise estimations. If the value of the dependent variable for a certain observation is lower than average in period  $t$ , it is likely to be higher in period  $t + 1$  than in period  $t$ . Similarly, observations with high values in period  $t$  tend to be followed by lower values in period  $t + 1$ . Accordingly, mean reversion leads to an increase in the expected value of the dependent variable in the group of observations belonging to the lower percentiles of  $y$ , and to a decrease in the expected value – in the higher percentiles of  $y$ . Therefore, the impact of mean reversion needs to be excluded in econometric analysis, which evaluates the longitudinal impact of policy interventions on groups of economic agents.

The purpose of this article is to model multivariate dependence of the variable of policy interest by disentangling the two sources of intertemporal dependence: one from the effect of the policy of interest per se and the other from mean reversion. Specifically, we show a way of separating the effect of mean reversion from the policy effect when evaluating the impact of an incentive scheme with intertemporal stimuli and intertemporal variation of the parameter of the reform intensity.

Although mean reversion is inherent to any stationary process, it is most often noted in the analysis of dynamic panels. The dynamic panel data model is a generalization of the panel data fixed effect regression when the dynamic structure of the process needs to be introduced. In our article we use the example of Medicare's incentive contract applied to the observed quality of services, which has to be described as an autoregressive process. Hence, in evaluating the effect of this incentive scheme on hospital quality, we follow a handful of articles which deal with mean reversion in dynamic panels [25,31,48,81].

We focus on the pay-for-performance mechanism – an innovative method of remuneration, which originally emerged in corporate finance and managerial economics, and has since been much used in the public sector (civil service, education, social work, and healthcare). In order to quantify the unobserved quality of work, the incentive scheme computes the performance level using imprecisely measured proxies for various dimensions of quality. Next, the regulator imposes an incentive contract, which relates remuneration to performance, so that agents with higher performance in the current period receive higher payment for their services in future periods than agents with lower performance. The reform intensity parameter in this context is the share of the agent's income, which is “at risk” under the incentive contract.

Assuming a direct association between demand for services and quality of work, higher payment to agents with high performance incentivizes agents to improve their level of quality in order to raise demand for their services. In such a setting, if the unobserved quality could be measured precisely, each agent would have sustained their fixed level of performance.

However, performance is in fact a noisy signal. First, there is an imprecision in measuring performance, since it is only a proxy for true quality. Second, in case of healthcare, the unobserved true quality of services is itself subject to a random variation, due, for instance, to patient non-compliance with medical treatment [62]. So it is plausible to assume that performance contains a random error. Hence, performance may unexpectedly be valued as having improved in period  $t$  due to this random error, and then the payment in period  $t + 1$  (which is a function of current performance) will increase. Accordingly, the incentive to improve quality in the future period becomes stronger for agents with higher performance. So the performance of these agents in period  $t + 1$  will be *on average* higher than their performance in period  $t$ . The reverse argument applies in case of unexpected lowering of performance valuation in period  $t$ .

What therefore happens is that performance of the economic agent becomes a process with serial correlation. So the evolution of the variable of policy interest when such incentives are applied can be viewed as an autoregressive process. In a situation where the policy variable changes over time, we estimate the unconditional mean in the autoregressive process as a function of the agent's characteristics and of policy intensity. Comparison of the fitted values of the unconditional mean under different values of the reform intensity enables us to identify the reform effect cleared of mean reversion. For instance, we contrast the unconditional means estimated under the values of the policy variable in two consecutive time periods. Alternatively, we compare the fitted value of the unconditional mean in period  $t$  with its counterfactual analogue: the unconditional mean at zero value of policy intensity. The article which is closest to our latter approach in assessing the

policy effect in dynamic panels is [48]: the actual value of return on equity (ROE) at merged banks is compared with the fitted value of ROE, measured as the unconditional mean of the AR(1) process for the whole banking industry (i.e., the counterfactual value of ROE in the absence of the merger).

It should be noted that our identification strategy is close to difference-in-difference analysis in a non-binary treatment: the intensity of reform is the analogue of the treatment variable and the share of Medicare's patients at the hospital is the analogue of the variable for the treatment/control groups.<sup>2</sup>

We use the example of Medicare's value-based purchasing, implemented at national level in the US since 2013 on the basis of a reward function that relates the aggregate measure of hospital performance to remuneration. Overall, applications of pay-for-performance are very numerous in healthcare, since healthcare is the classic example of an industry with asymmetric information where sustained quality of service is extremely important. It should be noted that the research in health economics is vulnerable to random shocks in the dependent variable and hence, to the phenomenon of mean reversion. Yet, as regards incentive schemes, to the best of our knowledge, only one article explicitly discusses the impact of random variation of quality [62] and only a few articles point to the need for reassessing the impact of Medicare's pay-for-performance incentive mechanisms in view of the potential impact of mean reversion [58,63].

Our estimations of the association between the observed level of prior quality and measured quality improvement employ nationwide data for 2,984 acute-care Medicare hospitals which are financed according to the quality-incentive mechanism in 2013–2019. The empirical approach uses annual variation in the size of quality incentives in order to estimate the effect of pay-for-performance cleansed of mean reversion. We control for other potential channels of quality improvement by Medicare hospitals, using data on the Hospital Readmissions Reduction Program (HRRP) and on the meaningful use of Electronic Health Records (EHR).

We find that the higher the quintile of the composite quality measure at Medicare hospitals, the larger the estimated effect of the reform. Our empirical results suggest that the stylized fact of inverse relationship between improvement owing to the incentive scheme and the baseline performance should be revisited. This inverse relationship has been found by most empirical assessments of the impact of incentive contracts on healthcare quality and seems to hold for various designs of pay-for-performance: it is observed for general practitioners in the UK; physician groups in California, Chicago, and Ontario; US hospitals in Michigan, New York, and Wisconsin; and hospitals involved in Medicare's pilot project for quality improvement [19,26,34,40,42,51,52,63,67,76]. However, we argue that the finding of an inverse relationship may be incorrect when the empirical approach fails to account for the impact of the random shocks on the time profile of quality under the intertemporal incentive scheme.

The remainder of the article is structured as follows. Section 2 reviews the design of Medicare's quality incentive and sets up the framework for evaluating its outcomes. Section 3 outlines the empirical methodology, and Section 4 describes the data for Medicare hospitals. The results of the empirical analysis are presented in Section 5. Section 6 contains a discussion of our approach in view of conventional methods for policy evaluation, and Section 7 supports the quantitative findings of our analysis by suggesting potential channels used for quality improvement at hospitals.

## 2 Medicare's incentive contract

### 2.1 Policy setting

The mechanism provides an incentive proportional to measured quality and has been applied to discharges in the inpatient prospective payment system at acute-care Medicare hospitals since 2013.<sup>3</sup> The scheme

<sup>2</sup> If the treatment is binary, our approach does not differ from the conventional difference-in-difference estimation with the interaction term of the pre/post treatment dummy and the dummy for the treatment/control group.

<sup>3</sup> Two US states are exceptions to the rule: Puerto Rico, which only started innovating its healthcare system in 2015 and Maryland, which has a unique model for hospital financing.

reduced Medicare's base payment<sup>4</sup> to each hospital by a factor  $\alpha_t$  which equaled 0.01 in 2013. The amount of the reduction was increased annually by 0.0025 in 2014–2017 and has remained flat at 0.02 since 2017. Note that  $\alpha$  is the parameter of the reform intensity, varying over time, and  $\alpha = 0$  would correspond to a counterfactual setting with the absence of the reform.

The accumulated saving from reduction in base payment is redistributed across hospitals according to an adjustment coefficient, which is computed as a linear function of the composite quality measure:  $1 + \left(\kappa_t \frac{m_{it}}{100} - 1\right) \cdot \alpha_t$ , where  $i$  is the index of a hospital,  $t$  indicates year, and  $m_{it}$  is the hospital's *total performance score (TPS)*, ( $0 \leq m_{it} \leq 100$ ). A hospital is rewarded in period  $t + 2$  if the adjustment coefficient based on  $m_{it}$  is above one and is penalized otherwise. The quality incentive scheme is budget-neutral and the value of the slope  $\kappa_t$  is chosen to ensure budget neutrality, so that hospitals with value of *TPS* above the empirical mean gained under the reform. In the first years of the reform  $\kappa_t$  was close to 2, so hospitals with value of the composite quality measure above 50 were winners from the incentive scheme.

The *TPS* is a weighted sum of scores for measures in several domains: timely implementation of recommended medical interventions (*clinical process* of care), quality of healthcare as perceived by patients (*patient experience* of care), survival rates for AMI, heart failure and pneumonia patients and other proxies for *outcome* of care, healthcare-associated infections and other measures of *safety* of care, and spending per beneficiary as a measure of *efficiency* of care.<sup>5</sup>

A hospital's intertemporal incentive in Medicare's scheme is based on the expectation that the quality payments will continue over a long term, so the hospital's executives and physicians realize that demand is proportionate to quality and that their current policies toward quality of care will influence future reimbursement [46,73].

## 2.2 Autoregressive process and quality convergence

The evolution of the measured quality constitutes a process with serial correlation. If the process for the measured quality is stationary, then it may be treated as an autoregressive process  $m_t - \mu(\theta) = \phi_1(m_{t-1} - \mu(\theta)) + \dots + \phi_p(m_{t-p} - \mu(\theta)) + \varepsilon_t$ . Here  $\mu(\theta) = E(m_t|\theta)$  denotes the mean value of the measured quality for a hospital with type  $\theta$ . As the absolute values of the reciprocals of the roots of the characteristic equation of AR( $p$ ) processes are less than one, the maximum absolute value of these reciprocals (denoted  $\lambda$ ) may be used as the measure of persistence for the process of measured quality [74].

Using definitions in [29], we can disentangle a permanent component in  $m_t$ , which is related to economic impact of pay-for-performance from a transient component (a pure dynamic effect), which may be referred to as “mean reversion” or “regression toward the mean” [30].

The reason for the phenomenon of mean reversion is the existence of the random error  $\varepsilon_t$  in the measured quality  $m_t$ . Indeed, in the absence of  $\varepsilon_t$  the process quickly converges to its mean  $\mu(\theta)$  and does not exhibit mean reversion because it always sits at the mean. The random error in the measured quality is largely attributed to imprecision in quality measurement: it is hard to reveal true quality using observable proxies. Another reason is random variation in true quality, which may be explained by the fact that patients do not always comply with the prescribed treatment [62]. Combined with the fact that hospitals make an intertemporal decision in respect of the quality-based reimbursement, the random error leads to the autoregressive form of measured quality  $m_t$ .

<sup>4</sup> The base payment is linked to each diagnosis-related group.

<sup>5</sup> The domain score is the sum of the scores for its measures. Higher score of the measure reflects higher position of the hospital in the empirical distribution of the quality measure in a given year or greater improvement of the quality measure relative to the baseline period. Specifically, achievement points are computed for each measure evaluating a hospital's performance relative to other hospitals in a given year, and improvement points for each measure are computed to assess change in the hospital's own performance in the given year relative to the baseline period. Then, for each measure, the highest of the two (achievement points or improvement points) is used as the hospital's score for that measure.

The autoregressive specification can be taken as equivalent to convergence of the measured quality toward the value  $\mu(\theta)$  and  $\lambda$  is associated with the speed of quality convergence. The persistence parameter  $\lambda$  essentially describes how quickly the effect of any unexpected shock in value of the dependent fades over time. For example, consider a simple AR(1) process with  $0 < \lambda < 1$  and the conditional mean  $E(m_t|m_{t-1}, \theta) = \mu(\theta) + \lambda(m_{t-1} - \mu(\theta))$ . Here the expected value of current measured quality  $Em_t$  is closer to the mean value  $\mu(\theta)$  than is the value of the measured quality in the previous period, i.e.,  $m_{t-1}$ . The expression for  $E(m_t|m_{t-1}, \theta)$  becomes more complicated for AR( $p$ ) processes with  $p > 1$ , but  $\lambda$  can still be used as a measure of persistence of the process.

The hospital receives higher profits for improvement of performance under higher values of  $\alpha$  than under lower values of  $\alpha$ . This, combined with the serial correlation between performance in consecutive periods, implies a direct association between the persistence parameter  $\lambda$  and  $\alpha$ . Higher values of  $\lambda$  imply a lower rate of convergence of quality and hence a weaker effect of mean reversion.

## 2.3 Expected outcomes of the reform and time profiles of the quality measure

### 2.3.1 Mean effect of the reform

The payment schedule makes the hospital adjustment coefficient a linear function of TPS, so each hospital has an incentive to raise the value of the observed composite quality measure. Hence, the introduction of pay-for-performance is expected to have a positive effect on mean value of the composite quality measure. Indeed, the mean level of hospital performance was improved even in case of a continuous reward function applied to hospitals above the threshold values of quality indicators (Medicare's pilot program, Phase I) [18,34,37,52,68]. Specifically, the value of the composite performance score in Medicare's pay-for-performance hospitals was higher than in the control group of hospitals [52,78]. Moreover, sociological evidence points to the fact that hospitals participating in incentive schemes are likely to improve performance as they implement a larger number of quality improving activities that non-incentivized hospitals do not carry out [41].

The higher the value of  $\alpha$ , the higher may be the hospital's loss under the reform in case of insufficient value of TPS. Indeed, the empirical evidence points to larger incentives being more effective than smaller ones in such reforms [8,15,60].

Accordingly, the expected mean effects of the reform may be formulated as follows:

Hypothesis *H1a*: The introduction of pay-for-performance and the increase of parameter  $\alpha$  in the context of pay-for-performance lead to a positive mean effect on observed quality.

Hypothesis *H1a* implies that hospitals can be treated as agents which take their future payments into account. The intertemporal stimuli result in mean reversion with respect to observed quality. However, the strength of mean reversion is interrelated with parameter  $\alpha$  as follows:

Hypothesis *H1b*. The increase in the share of hospital funds at risk in pay-for-performance weakens the effect of convergence of the measured quality to the mean value.

### 2.3.2 Groupwise effects of the reform

We assume that the effect of Medicare's reform will be larger at hospitals with higher quality, based on findings in the health policy literature that emphasis on quality improvement in incentive schemes is greater at high-quality hospitals or among high-quality physicians in comparison with low-quality hospitals and physicians [21,37,69,77,78].

For instance, [77] conducted structural surveys at hospitals in the top two and bottom two deciles of performance measure in Medicare's pilot program and discovered stronger involvement in quality improving activities among top performing hospitals. The statistically significant differences between top- and bottom-performing hospitals were observed in case of the numerical values, assigned to the following components of quality improvement: organizational culture, multidisciplinary teams, "adequate

human resources for projects to increase adherence to quality indicators” and “new activities or policies related to quality improvement” (Tables 3 and P on pp. 836–837).

Interviews with the leaders of California physician organizations [21] similarly discovered that physicians with high performance placed higher emphasis on the support that “the organization dedicates to addressing quality issues” than medium- and low-performing physicians (Exhibit 3, p. 521).

Moreover, papers that use policy evaluation techniques applied to assessment of the effect of the pilot pay-for-performance program at Medicare hospitals report that hospitals in the top two deciles of quality measures showed the fastest improvement, while hospitals in the lowest deciles raised their quality to a much lesser extent or may even have failed to improve [69,78].

To sum up, the hypothesis on groupwise effects of pay-for-performance is as follows:

Hypothesis *H2*. The introduction of pay-for-performance leads to a larger boost of measured quality at high-quality hospitals than at low-quality hospitals.

### 2.3.3 Net total effect over time at groups of hospitals

Consider the multivariate dependence of the variable of interest on two sources of intertemporal dependence: the policy reform and mean reversion. The effect of mean reversion implies a differential time profile of measured quality: measured quality increases at hospitals in low percentiles of the quality distribution and decreases at hospitals in high percentiles. Combined with the positive effect of pay-for-performance on the mean value of measured quality (Hypothesis *H2*), mean reversion is likely to result in heterogeneous net total effect of change in measured quality over time.

Hypothesis *H3a*. High-quality hospitals experience decrease of measured quality owing to regression toward the mean. However, the introduction of pay-for-performance and increase of the share of hospital funds at risk in pay-for-performance lead to improvements in measured quality at these hospitals. The net total effect may vary.

Hypothesis *H3b*. Low-quality hospitals increase their measured quality owing to regression toward the mean. The introduction of pay-for-performance and increase of the share of hospital funds at risk in pay-for-performance also cause a rise in measured quality, so the net total effect at these hospitals is positive.

If  $\alpha$  is gradually raised in the course of implementation of the incentive scheme, then, according to *H1b*, convergence of measured quality weakens over time. The net total effect at high-quality hospitals is the sum of the positive effect of the quality incentive and negative effect of the quality convergence. With increase in  $\alpha$ , the number of hospitals where the positive effect outweighs the negative becomes larger.

Hypothesis *H3c*. The increase of hospital funds at risk under pay-for-performance weakens the effect of convergence of measured quality, so the number of high-quality hospitals with negative net total effect decreases.

## 3 Empirical approach

### 3.1 Specification

The dependent variable  $y_{it}$  is the TPS of hospital  $i$  in year  $t$ . The value of  $y_{it}$  is used for remuneration of Medicare hospitals at time  $t + 2$ , so we employ the second-order dynamic panel,<sup>6</sup>

$$y_{it} = \phi_0 + \phi_1 y_{it-1} + \phi_2 y_{it-2} + \phi_3 \alpha_t s_{it} + \phi_4 \alpha_t s_{it} y_{it-1} + \phi_5 \alpha_t s_{it} y_{it-2} + \delta_0 s_{it} + z_{it}' \delta_1 + \alpha_t s_{it} \cdot z_{it}' \delta_2 + d_t' \delta_3 + u_i + \varepsilon_{it}, \quad (1)$$

<sup>6</sup> Formally, the model with *at least* two lags must be used for describing the evolution of the TPS. However, the coefficients for the variables with the third lag of TPS turned out insignificant in our empirical estimation, so we treat TPS as an AR(2) process.

where  $z_{it}$  are hospital time-varying characteristics,  $u_i$  are individual hospital effects (in particular, they incorporate the altruistic effects), the size of quality incentives  $\alpha_t$  varies in different years and enters the equation multiplied by the share of Medicare discharges  $s_{it}$ , which indicates that the quality incentives apply only to treatment of Medicare patients, and  $d_t$  is a set of dummy variables which capture external time effects (effects unrelated to hospital decisions). The following restrictions are used to identify the constant term  $\phi_0$ : the sum of the coefficients for components of  $d_t$  is normalized to zero, and the expected value  $E(u_i) = 0$ . Hospital time-varying characteristics are disproportionate share index, casemix index, number of hospital beds,<sup>7</sup> physician-to-bed ratio, and nurse-to-bed ratio. The posterior analysis of the effect of quality incentives deals with hospital grouping according to the time-invariant characteristics, which could not be incorporated in the empirical specification with fixed effects: geographic region where the hospital is located, public ownership, urban location, and teaching status.

We use two hospital control variables which affect quality improvement and allow us to mitigate potential biases, which might occur if the pay-for-performance effect is identified based only on the variation of  $\alpha$  in time. The HRRP penalty captures the impact of a simultaneously adopted incentive program with similar incentives. Moreover, the readmission reduction program targets improvement of quality measures which are components of TPS.<sup>8</sup> The binary variable for successful attestation of meaningful usage of EHR accounts for the effect of another compulsory program, which provides bonuses to attested hospitals. The variable controls for the fixed cost incurred by a hospital to improve its quality through installing and using health information technology systems.

Eq. (1) can be estimated using the generalized method of moments: the [2] and [12] methodology for dynamic panel data. Examples of use of the methodology in health economics include analysis of the quality of care at Medicare's hospitals in [56], study of the length of stay at Japanese hospitals in [10], investigation of labor supply by Norwegian physicians in [4], and of health status of individuals in the US in [57].

The first set of moment conditions in GMM comes from the approach of [2] and [12]. We take the first difference of the right-hand side and left-hand side of Eq. (1):

$$\begin{aligned} \Delta y_{it} = & \phi_1 \Delta y_{it-1} + \phi_2 \Delta y_{it-2} + \phi_3 \Delta(\alpha_t s_{it}) + \phi_4 \Delta(\alpha_t s_{it} y_{it-1}) + \phi_5 \Delta(\alpha_t s_{it} y_{it-2}) + \delta_0 \Delta s_{it} + \Delta(z_{it})' \delta_1 \\ & + \Delta(\alpha_t s_{it} \cdot z_{it})' \delta_2 + \Delta d_t' \delta_3 + \Delta \varepsilon_{it}. \end{aligned} \quad (2)$$

Since  $\varepsilon_{it}$  cannot be predicted using the information available at period  $t - 1$ ,  $\varepsilon_{it}$  is uncorrelated with any variable known at time  $t - 1$ ,  $t - 2$  etc. Therefore,  $\Delta \varepsilon_{it}$  is uncorrelated with any variable known at time  $t - 2$ ,  $t - 3$  etc. Hence, the following set of moment conditions can be imposed to estimate the model parameters in Eq. (2), see [2] and [12]:

$$\begin{aligned} t = 3: & E(\Delta e_{i3} Z_{i1}) = 0, \\ t = 4: & E(\Delta e_{i4} Z_{i1}) = 0, \quad E(\Delta e_{i4} Z_{i2}) = 0, \\ t = 5: & E(\Delta e_{i5} Z_{i1}) = 0, \quad E(\Delta e_{i5} Z_{i2}) = 0, \quad E(\Delta e_{i5} Z_{i3}) = 0, \\ & \text{etc.} \end{aligned}$$

where  $e_{it}$  is the regression residual and  $Z_{it}$  is any variable known at time  $t$ .<sup>9</sup>

Another set of moment conditions comes from [12] for the level Eq. (1):  $u_i + \varepsilon_{it}$  has to be uncorrelated with  $\Delta Z_{it-1}$  for any stationary variable  $Z_{it}$ , where  $Z_{it-1}$  is known at time  $t - 1$ .

$$E((u_i + \varepsilon_{it}) \Delta Z_{it-1}) = 0, \quad t = 3, 4, \dots \quad (3)$$

<sup>7</sup> As the distribution of the number of hospital beds is extremely skewed, we take the log of hospital beds. This approach is in line with [22] and makes it possible to account for nonlinear effect of hospital beds. It is less restrictive than the alternative approach, employing a list of dummies based on the ranges of hospital beds (e.g., less than 99, 100–199 etc.). Use of a list of dummies condemns the effect to be piece-wise, prohibiting variation within the category of hospitals with a given range of beds.

<sup>8</sup> 30-day unplanned readmission rates for acute myocardial infarction, heart failure, and pneumonia.

<sup>9</sup> For instance,  $y_{it}$  may serve as  $Z_{it}$ .

So  $Z_{it}$  includes lagged values of predetermined and endogenous variables (the first set of moment conditions) and differenced predetermined and endogenous variables (the second set of the moment conditions). All moment conditions are formulated separately for different years, so the number of observations for asymptotics equals the number of hospitals.<sup>10</sup>

More specifically, lagged value of TPS and other hospital control variables in  $z_{it}$  (beds, physician-to-bed and nurse-to-bed ratios, HRRP penalty, and the binary variable for hospital EHR attestation) are taken as predetermined and do not require the use of instruments in estimations. Casemix and the disproportionate share index are assumed to be endogenous: we rely on the empirical evidence of manipulation by hospitals of patient diagnoses (i.e., with casemix) and reluctance to admit low-income patients under quality-incentive schemes [17,23,28]. We assume that the Medicare share is endogenous, too: the fact may be explained by demand-side response from Medicare patients to publicly reported hospital quality [44,53,72].

It should be noted that the use of dynamic panel data methodology requires justification on economic grounds. This is because the approach uses lags and lagged differences as instruments, and there are potential problems with using lags as instruments even though they pass the Arellano-Bond tests. Specifically, lags may prove to be weak and invalid [7]: the weakness may occur when lags are distant [59], and invalidity happens due to overfitting of the endogenous variable under large  $T$  [66]. However, neither of these problems (weakness and invalidity) are likely to be present in our analysis since we restrict our instruments only by the first appropriate lag.

The validity of instruments is assessed through statistics of the Arellano-Bond test. We employ [80] robust standard errors for estimation.<sup>11</sup> But formal tests are insufficient for establishing the causal relationship in models, which use an instrumental variable approach [1,7]. Accordingly, it is necessary to provide an economic justification for the assumption of the exclusion restriction of the instruments, i.e., that the instruments are exogenous and impact the dependent variable through no channels other than the endogenous variable and, possibly, also through exogenous covariates. An example of such justification on theoretical grounds can be found in [6], who uses lags of GDP and lags of the inflation rate as instruments for GDP and inflation. Another way of arguing for the exclusion restriction is given in [38], which estimates per capita output in various countries as a function of social infrastructure. Owing to endogeneity of social infrastructure, variables related to exposure to Western culture are used as instruments, and there is a discussion of the absence of any direct channels through which these variables could impact a country's per capita output.

We follow the latter approach to provide an economic justification for the validity of instruments in the dynamic panel data model for the composite quality measure at Medicare hospitals. Our arguments below, which advocate the applicability of lagged first differences as instruments for the level Eq. (1) and first lagged levels as instruments for the difference Eq. (2), are based on the plausible assumption of a short adjustment period in the values of the dependent variable. Specifically, we assume that hospital managers take prompt action upon learning the TPS in year  $t$ , so that adjustment is observed in the next period and is not delayed until a more remote future. This assumption is supported by interviews with hospital managers [21,46,37,55,73,77], which show real-time assessment of performance of hospital personnel and immediate feedback initiatives aimed at correcting possible lack of quality. For instance, at Medicare hospitals which participated in the pilot pay-for-performance program, “progress reports were routinely delivered to hospital leadership and regional boards” ([37], p. 45S). Hospital-specific and physician-specific compliance reports were collected at least every 1.5 months on average, and the results of these reports were delivered to individual physicians once in 5 months on average at both top-performing and bottom-performing hospitals ([77], Table 4, p. 837). As regards nationwide implementation of pay-for-performance at Medicare hospitals, the TPS is calculated annually, but values for the quality dimensions of the TPS are made publicly

<sup>10</sup> The separate formulation of moment conditions for different years makes it impossible to apply the exclusion restrictions, which are commonly used in the instrumental variables approach.

<sup>11</sup> The Sargan statistic may be used in dynamic panels for assessing validity of instruments under the homoskedasticity assumption. But it is not applicable to our specification with robust standard errors.



available on a quarterly basis.<sup>12</sup> Frequent announcements of quality scores make it possible to expedite quality adjustment at each hospital and improve the value of the TPS within a year. For instance, the survey of hospital CEOs, physicians, nurses, and board members showed that, since implementation of the value-based purchasing program, “data were shared with their board and discussed at least quarterly with senior leadership” ([55], p. 435).

As regards our formal analysis, Eq. (1) has TPS as a dependent variable and its first and second lags as explanatory variables.  $\Delta y_{t-1}$  is used as an instrument for  $y_{t-1}$ . We assume that the change in TPS from period  $t - 2$  to  $t - 1$ , i.e.,  $\Delta y_{t-1}$ , which is observed at a hospital at  $t - 1$ , is immediately followed by the hospital’s action in period  $t - 1$ . So the instrument  $\Delta y_{t-1}$  affect the dependent variable  $y_t$  through the endogenous variable  $y_{t-1}$ , i.e., through improved quality in period  $t - 1$  (and potentially also through the predetermined variable  $y_{t-2}$ , i.e., quality adjustment may start as early as in period  $t - 2$ ) but not through other channels. Without the short adjustment period, these other channels might have included some postponed effects which only come into effect in period  $t$ . Note that the equation has hospital control variables, and we follow the empirical literature on the US Medicare reform by treating some of them as endogenous. One of such variables, the share of Medicare patients, reflects the desire of the regulator to sign contracts with the hospital to treat Medicare patients, and it is a function of the hospital’s quality enhancing efforts [46]. Our empirical strategy relies on the fact that  $\Delta x_{t-1}$  is an excludable instrument for  $x_t$ . It is, indeed, plausible to assume that increase of quality efforts from period  $t - 2$  to period  $t - 1$  results in positive value of  $\Delta s_{t-1}$  (where  $s_t$  denotes the share of Medicare patients) and impacts the value of the TPS in the period  $t$ . A similar argument applies to another endogenous control variable – casemix – which reflects the share of patients with complicated diagnoses. If we ignore potential dumping of patients by hospitals, hospitals are interested in treating patients with complicated diagnoses, since compensation in the system of diagnosis-related groups is higher for severe cases. But patient demand responds to public reports on hospital quality [20,42,44], so the share of Medicare cases becomes a function of hospital quality.

Another equation is (2) and it models first differences, i.e., changes in quality. The dependent variable is  $\Delta y_t$  and it is a function of the endogenous variable  $\Delta y_{t-1}$ , a predetermined variable  $\Delta y_{t-2}$ , and the difference in the values of hospital control variables  $\Delta x_t$ . The instrument for  $\Delta y_{t-1}$  is  $y_{t-2}$  and the instrument for each endogenous hospital control variable is  $x_{t-2}$ . Following the above logic about prompt response of TPS to its values in the previous period, we presume that  $y_{t-2}$  will affect the change in the value of the TPS from period  $t - 2$  to period  $t - 1$ . So  $y_{t-2}$  impacts  $y_t$  through  $\Delta y_{t-1}$  (and potentially even through the predetermined variable  $\Delta y_{t-2}$ ) but not through other channels (i.e., not through processes that occur as late as in period  $t$ ). Similarly, upon learning the value of  $x_{t-2}$ , hospitals speedily adjust their quality to change  $\Delta x_t$  and it affects  $\Delta y_t$ .

Note that [56] used similar arguments in discussing applicability of the dynamic panel data model to analyze in-hospital mortality and the complication rate, which are used as measures of hospital quality in US Medicare hospitals. They write: “We believe our approach is appropriate because (i) changes to in-hospital mortality and complications should be immediately affected by changes in staffing levels, not after a long adjustment period, and (ii) the influence of the past is incorporated through the lagged value of the dependent variable.” (p. 296, Footnote 3).

A related study applying dynamic panel data models to hospital performance indicators deals with average length of stay at Japanese acute-care hospitals that plan to introduce a prospective payment system [10]. The variable is treated in Japan as a proxy for hospital efficiency. It is regularly monitored and analyzed by the regulator and by hospital management, with feedback actions by hospital personnel in response to annual updates on levels of the variable [9,10,43,45,75]. Accordingly, the assumption of a short adjustment period for the length of stay is likely to hold at Japanese hospitals and the use of lagged levels and lagged differences as instruments is justified.

<sup>12</sup> Exceptions are one measure in the clinical process of care domain (influenza immunization), one measure in the safety domain (PSI-90), and a measure in the efficiency domain – Medicare spending per beneficiary, which are updated annually. See measure dates in the quarterly data archives available at <https://data.cms.gov/provider-data/archived-data/hospitals>.

Note that potential violations of the exclusion restriction may occur in instances where the quality measure requires long periods to adjust. In such instances, causal impact of the Medicare reform on the quality of care cannot be established [1,7].

We note other limitations of our approach. First, the analysis deals with the composite quality measure. While quality-related efforts of a hospital and the TPS composite quality measure are multi-dimensional, we do not touch upon multi-tasking in the empirical estimations. Our approach considers a one-dimensional effort, a one-dimensional true quality, and its measurable proxy.<sup>13</sup>

Second, we do not touch on the rules for computing the scores of each dimension of the composite measure or on aggregation of dimension scores. It is important to note that Medicare uses whichever is highest, improvement points or achievement points, as the score for each dimension. The choice between achievement and improvement points stimulates low-performing hospitals, and the uniform formula assumes that all groups of hospitals have equal margin for improvement. A minor exception is protection of hospitals above the benchmark value of the 95th percentile of a corresponding measure score: these hospitals receive 10 points for their achievement on a [0, 10] scale, while the maximum number of points for improvement by any hospital is 9.<sup>14</sup>

Third, weighting of scores across domains is another feature of the design of the incentive mechanism which is not analyzed in our article. So the dichotomous variables for annual periods in the empirical specification capture time effects unrelated to Medicare's value-based purchasing as well as time effects not associated with the size of incentives but potentially linked to changes in other elements of the reform design (i.e., changes in weights).

Finally, conventional policy evaluation using a control group of hospitals is not possible because quality measures for non-Medicare hospitals are not available.<sup>15</sup> The empirical part of the article therefore focuses solely on pay-for-performance hospitals and identifies the effect of quality incentives based on variation in  $\alpha_t$  and the share of Medicare patients in the hospital  $s_{it}$ . Variation in  $\alpha_t$  plays the role of the dummy for treatment/pre-treatment periods, and variations in  $s_{it}$  act similar to the dummy for the treatment/control groups.

## 3.2 Multivariate dependence of the quality variable

### 3.2.1 Calculation of the mean in the autoregressive process

We interpret the second-order dynamic panel (1) as a second-order autoregressive process. The coefficients for the first and the second lags of  $y_{it}$  in this AR(2) process are equal to  $\phi_1 + \phi_4\alpha_t s_{it}$  and  $\phi_2 + \phi_5\alpha_t s_{it}$ ,

---

**13** Note that in case of Medicare's formula, the true multi-dimensional quality of hospitals (and hence quality-related efforts) is transformed into measured quality (i.e., TPS) in a non-linear manner, owing to the step-wise scale used for computing the points for each measure. We might nonetheless assume that quality is transformed into TPS monotonically and can be linearized in the empirical part of the article. Several arguments can be listed to support the conjecture. First, the data for Medicare hospitals show that no hospital has the highest possible step-wise values for all its measures. So even the best hospitals have an incentive to work to increase at least one of their measure scores in order to improve TPS. Second, we can neglect disincentives within the step-wise scale used for aggregating measure scores, which may cause deterioration of quality for hospitals that are already positioned at the highest step. Such hospitals could afford only a slight decrease of their quality (due to slacking efforts) while remaining at this step, and the impact on the value of TPS of a fall in quality in only one quality measure will be negligible. Third, interviews with executives of hospitals using value-based purchasing show that a hospital rarely gives special attention to a given subset of measures or shifts its administrative and other efforts across measures. All dimensions of TPS are monitored and actions to improve each dimension are implemented [46,73].

**14** The approach used by Medicare is in contrast with the methodology used in France, where all providers are stimulated according to improvement while only providers with quality above the mean value are also rewarded for their achievement [14].

**15** The TPS or all its components are only available for hospitals in the Hospital Compare database. The Hospital Compare database does include a small group of non-incentivized hospitals together with value-based purchasing hospitals. These are children's hospitals and critical-access hospitals. But both groups offer a special type of healthcare and are not comparable with acute-care hospitals. Moreover, critical-access hospitals usually have no more than 20 beds, which makes it impossible to find a close match with acute-care hospitals. See [70] for an attempt of matching acute-care and critical-access hospitals.

respectively. Note that both coefficients are linear functions of  $\alpha_t$ . While the standard form of the AR(2) process contains only the lags of the dependent variable, the right-hand side of our empirical equation includes various hospital characteristics and control variables.

To test the hypotheses which concern the mean value of the measured quality  $\mu$ , we measure the mean fitted value of  $y_{it}$  as follows.

For a fixed value of  $\alpha$  we take the unconditional expected values of both sides of (1) and denote  $\mu(\alpha) = E(y_{it})$ :

$$\begin{aligned} \mu(\alpha) = & \phi_0 + \phi_1\mu(\alpha) + \phi_2\mu(\alpha) + \phi_3\alpha E(s_{it}) + \phi_4\alpha E(s_{it})\mu(\alpha) + \phi_4\alpha\text{cov}(s_{it}, y_{it-1}) + \phi_5\alpha E(s_{it})\mu(\alpha) \\ & + \phi_5\alpha\text{cov}(s_{it}, y_{it-2}) + \delta_0 E(s_{it}) + E(z_{it})'\delta_1 + \alpha E(s_{it}z_{it})'\delta_2 + E(d_t)'\delta_3, \end{aligned} \quad (4)$$

where  $E(d_t)'\delta_3 = 0$  because of the normalization of coefficients  $\delta_3$  in (1). After collecting the terms with  $\mu$  and rearranging them, we obtain:

$$\mu(\alpha) = \frac{\phi_0 + \phi_3\alpha E(s_{it}) + \delta_0 E(s_{it}) + E(z_{it})'\delta_1 + \alpha E(s_{it}z_{it})'\delta_2 + \phi_4\alpha\text{cov}(s_{it}, y_{it-1}) + \phi_5\alpha\text{cov}(s_{it}, y_{it-2})}{1 - \phi_1 - \phi_2 - \phi_4\alpha E(s_{it}) - \phi_5\alpha E(s_{it})}.$$

Since  $\alpha$  differs across  $t$ , we use sample means across the hospitals for fixed  $t$  to obtain estimates of expectations.

The estimate of  $\mu(\alpha)$  is constructed by replacing the expected values and covariances by corresponding sample means and sample covariances:

$$\mu(\alpha) = \frac{\phi_0 + \phi_3\alpha\bar{s} + \delta_0\bar{s} + \bar{z}'\delta_1 + \alpha\bar{s}\bar{z}'\delta_2 + \phi_4\alpha\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha\bar{s} - \phi_5\alpha\bar{s}}.$$

Note that the expression for  $\mu(\alpha)$  does not contain the time effects  $d_t'\delta_3$ , as they represent shifts in quality which are common to all the hospitals and are caused by external circumstances.

### 3.2.2 Intertemporal dependence due to the policy reform

The policy parameter  $\alpha$  increases in 2013–2017 and remains unchanged in 2017–2019. As follows from the hypothesis  $H1a$ , the value of  $\mu(\alpha_t)$  is expected to increase through 2013–2017 and should become flat in 2017–2019.

Accordingly, we examine the difference between  $\mu(\alpha_t)$  and  $\mu(\alpha_{t-1})$ :

$$\begin{aligned} \mu(\alpha_t) - \mu(\alpha_{t-1}) = & \frac{\phi_0 + \phi_3\alpha_t\bar{s} + \delta_0\bar{s} + \bar{z}'\delta_1 + \alpha_t\bar{s}\bar{z}'\delta_2 + \phi_4\alpha_t\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha_t\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha_t\bar{s} - \phi_5\alpha_t\bar{s}} \\ & - \frac{\phi_0 + \phi_3\alpha_{t-1}\bar{s} + \delta_0\bar{s} + \bar{z}'\delta_1 + \alpha_{t-1}\bar{s}\bar{z}'\delta_2 + \phi_4\alpha_{t-1}\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha_{t-1}\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha_{t-1}\bar{s} - \phi_5\alpha_{t-1}\bar{s}}. \end{aligned}$$

The null hypothesis is as follows:

$$H_0 : \mu(\alpha_t) - \mu(\alpha_{t-1}) = 0,$$

and it is tested against the positive alternative.

Equivalently, we compute the difference between  $\mu(\alpha)$  and  $\mu(0)$ :

$$\mu(\alpha) - \mu(0) = \frac{\phi_0 + \phi_3\alpha\bar{s} + \delta_0\bar{s} + \bar{z}'\delta_1 + \alpha\bar{s}\bar{z}'\delta_2 + \phi_4\alpha\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha\bar{s} - \phi_5\alpha\bar{s}} - \frac{\phi_0 + \bar{z}'\delta_1}{1 - \phi_1 - \phi_2}.$$

Note that  $\mu(0)$  represents the mean value in the pre-reform years when  $\alpha = 0$  and is obtained analytically by plugging  $\alpha = 0$  into the expression for  $\mu(\alpha)$ .

The null hypothesis is as follows:

$$H_0 : \mu(\alpha_t) - \mu(0) = 0,$$

and it is tested against the positive alternative.

In conjunction with hypothesis *H1a*,  $\mu(\alpha_t) - \mu(\alpha_{t-1})$  should be positive in 2013–2017 and should be close to zero in 2017–2019. Equivalently,  $\mu(\alpha_t) - \mu(\alpha_0)$  should be positive in 2013–2019 and should increase over the period 2013–2017.

Now consider hypothesis *H1b*. The persistence parameter  $\lambda(\alpha)$  describes how quickly the effect of random shock in quality fades over time. For a second-order autoregressive process the rate of convergence of the conditional expected value of  $y_{it}$  decays exponentially at a rate equal to the reciprocal value of the smallest root of the characteristic equation for the AR(2) process:

$$1 - (\phi_1 + \phi_4\alpha_t s_{it})\lambda - (\phi_2 + \phi_5\alpha_t s_{it})\lambda^2 = 0$$

([39], Section 2.3). Again, for a fixed value of  $\alpha$  we take expectations

$$1 - (\phi_1 + \phi_4\alpha E(s_{it}))\lambda - (\phi_2 + \phi_5\alpha E(s_{it}))\lambda^2 = 0.$$

Then we replace the expected values by sample means and solve this quadratic equation to obtain the following formula for  $\lambda(\alpha)$ :

$$\lambda(\alpha) = \frac{\phi_1 + \phi_4\alpha\bar{s} + \sqrt{(\phi_1 + \phi_4\alpha\bar{s})^2 + 4(\phi_2 + \phi_5\alpha\bar{s})}}{2},$$

where  $\bar{s}$  is the mean value of the share of Medicare cases for a given year.

An alternative approach considers the value of the autocorrelation function (ACF(1)) (the correlation coefficient between  $y_{it}$  and  $y_{it-1}$ ) as the persistence parameter  $\lambda$ . Specifically, for the second-order autoregressive process (1) the estimated value of ACF(1) becomes

$$\lambda(\alpha) = \frac{\phi_1 + \phi_4\alpha\bar{s}}{1 - \phi_2 - \phi_5\alpha\bar{s}}$$

([39], Section 3.4).

Testing *H1b* implies analyzing whether  $\lambda(\alpha)$  is an increasing function of  $\alpha$ . So, similar to *H1a*, the null hypothesis:

$$H_0 : \lambda(\alpha_t) - \lambda(\alpha_{t-1}) = 0$$

is tested against the positive alternative.

Alternatively, we assess whether  $\lambda(\alpha) - \lambda(0)$  is positive, whether it increases in 2013–2017 and changes only negligibly in 2017–2019.

To assess *H2* we compute the effect of pay-for-performance as  $\mu(\alpha_t) - \mu(0)$  or  $\mu(\alpha_t) - \mu(\alpha_{t-1})$  at different quintiles of the lagged  $\text{TPS}_{it}$ , where quintile 1 denotes the lowest quality and quintile 5 denotes the highest. We investigate whether the effect is positive for  $\mu(\alpha_t) - \mu(0)$  in 2013–2019 (and for  $\mu(\alpha_t) - \mu(\alpha_{t-1})$  in 2013–2017) and whether the effect increases by quintile.

Testing *H3a* and *H3b* involves computing the predicted values of  $\text{TPS}_{it}$  at the mean value of each covariate for different quintiles of the lagged  $\text{TPS}_{it}$  and examining whether in 2013–2019 they change from positive in the lowest quintiles to negative or insignificant in the highest quintiles. Average difference between predicted TPS and lagged TPS shows the expected change in quality in consecutive years (the net total effect) which is the sum of the effect of pay-for-performance and the impact of mean reversion.

### 3.2.3 Estimation of the multivariate effect due to policy reform and mean reversion

Evaluation of *H3c* involves calculating annual values of the net total effect for quintiles of the lagged TPS and examining whether the negative values of the net total effect become less frequent across highest quintiles during the period 2013–2017 and stays constant in 2017–2019.

## 4 Data

### 4.1 Data sources and variables

The analysis uses data for Medicare hospitals in 2011–2019 from several sources. We use *Hospital Compare* data archives (January 2021 update) for quality measures, hospital ownership, and geographic location. The medical school affiliation of a hospital, the number of hospital beds, nurses, and physicians come from *Provider of Service* files. Other hospital control variables are taken from the Final Rules, which are Medicare's annual documents on reimbursement rates in the inpatient prospective payment system. Specifically, we use information from the *Impact Files*, which accompany the Final Rules and estimate the impact of the reimbursement mechanism on hospital characteristics. The variables taken from the *Impact Files* are the share of Medicare's discharges, ownership, and urban location.

Patient characteristics are also taken from the *Impact Files*. The casemix variable reflects the relative weight of each DRG in financial terms and is adjusted for transfers of patients between hospitals.<sup>16</sup> Casemix makes it possible to control for the composition of patient cases taking account of the objective link between severity of illness and hospital resources. The disproportionate-share index accounts for the share of low-income patients and makes it possible to proxy a patient's income.

To account for other major channels of quality improvement by Medicare hospitals over the observed time period, we use the data for two programs run by the Centers for Medicare and Medicaid Services. One of them is the HRRP, which applies to Medicare hospitals since fiscal year 2013 and penalizes them for excess readmissions. Specifically, the payment reduction which may equal from 0 to 3% is applied to hospital's Medicare remuneration, higher values of the percentage for the penalty represent more excess readmissions at the hospital. Using the *HRRP Supplemental data files*, which accompany annual Final Rules on acute inpatient PPS (June 2020 update), we find the HRRP penalty for 2013–2019 and use it as one of the control variables in the empirical analysis.

We also consider the EHR Incentive Program, which was in force since 2011. The program establishes hospital attestation on the use of EHR. The adoption of quality-improving information technology requires substantial fixed cost, so the binary variable for hospital attestation within EHR makes it possible to control for the fixed cost in the empirical analysis. The EHR promotion program consists of three stages (sequentially introduced in 2011, 2014, and 2017). Using data from *The Eligible Hospitals Public Use Files* on the EHR incentive program (February 2020 update), we set the EHR attestation dummy equal to one if the hospital passed its attestation for the given year at any stage. Owing to non-availability of data on the third stage of the program, we extend the second stage data from year 2016 to years 2017–2019. Use of an attestation dummy lets us control for the fact of incurring the fixed cost of quality-improvement efforts. Owing to the small size of the non-EHR group (only 8–10% of the sample), we do not analyze whether quality goes up faster in the group of the hospitals (for instance, we do not interact the attestation dummy with  $\alpha$ ).

### 4.2 Sample

The non-anonymous character of the data sources allows us to merge them by year and hospital name. Our analysis focuses on acute-care Medicare's hospitals, as the pay-for-performance incentive contract applies exclusively to this group. We restrict the sample by considering only hospitals with share of Medicare cases greater than 5%.

<sup>16</sup> If a patient was transferred to/from a hospital, then the transfer-adjustment factor is the lesser of one and the value of the patient's length of stay relative to the geometric mean of the national length of stay for this DRG. See Federal Register 2011, 42 CFR, Part 412.

**Table 1:** Descriptive statistics for Medicare's acute-care hospitals in 2013–2019

Variable	Definition	Obs	Mean	St.Dev	Min	Max
<b>Hospital performance</b>						
TPS	Hospital TPS	18,545	37.265	11.468	2.727	98.182
<b>Patient characteristics</b>						
Casemix	Transfer-adjusted casemix index	18,545	1.599	0.298	0.834	3.972
Dsh	Disproportionate share index, reflecting the prevalence of low-income patients	18,545	0.307	0.165	0	1.232
<b>Hospital characteristics</b>						
Nurses/beds	Nurse-to-bed ratio	18,545	1.312	3.849	0	170.479
Physicians/beds	Physician-to-bed ratio	18,545	0.099	0.947	0	70.992
Beds	Number of beds	18,545	272.158	241.538	3	2,891
log(beds)	Number of beds (in logs)	18,545	5.283	0.819	1.099	7.969
Medicare share	Share of Medicare cases	18,545	0.378	0.118	0.050	0.983
HRRP penalty	Percentage reduction of the Medicare payments under HRRP	18,545	0.498	0.590	0	3.000
MUEHR	=1 if passed attestation for meaningful usage of EHR	18,545	0.924	0.265	0	1
Urban	=1 if an urban hospital	18,545	0.711	0.453	0	1
Public	=1 if managed by federal, state or local government, or hospital district or authority	18,545	0.147	0.354	0	1
Teaching	=1 if hospital has medical school affiliation	18,545	0.364	0.481	0	1
<b>Hospital location</b>						
New England	=1 if located in Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, or Vermont	18,545	0.046	0.210	0	1
Mid-Atlantic	=1 if located in New Jersey, New York, or Pennsylvania	18,545	0.123	0.328	0	1
East North Central	=1 if located in Illinois, Indiana, Michigan, Ohio, or Wisconsin	18,545	0.168	0.374	0	1
West North Central	=1 if located in Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, or South Dakota	18,545	0.081	0.272	0	1
South Atlantic	=1 if located in Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, or West Virginia	18,545	0.177	0.381	0	1
East South Central	=1 if located in Alabama, Kentucky, Mississippi, or Tennessee	18,545	0.087	0.282	0	1
West South Central	=1 if located in Arkansas, Louisiana, Oklahoma, or Texas	18,545	0.129	0.335	0	1
Mountain	=1 if located in Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, or Wyoming	18,545	0.069	0.253	0	1
Pacific	=1 if located in California, Oregon, Washington, Alaska, or Hawaii	18,545	0.115	0.319	0	1

Note: Section 401 hospitals are treated as rural hospitals.

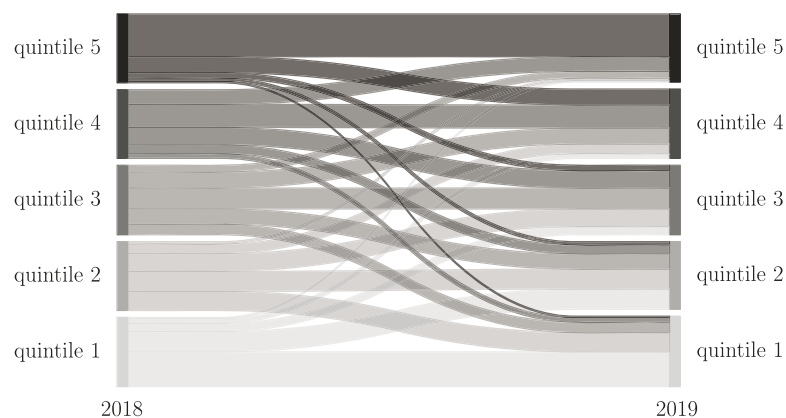
The specification with second-order lag enables estimation of the fitted values of TPS, and the values of  $\mu_t$  only starting 2013. Accordingly, we can evaluate the impact of the incentive contract on quality improvement in 2013–2019. There are 2,984 hospitals in our sample for 2013–2019, which make TPSN observations (Table 1).

### 4.3 Flow of quality and evidence of mean reversion

Descriptive analysis of the values of TPS offers suggestive evidence in support of some of the main hypotheses generated by the model. Specifically, we focus on the flow of hospitals between quintiles of TPS in different years. The Sankey diagrams in Figures 1 and 2 use the width of arrows as the intensity of flow rates



**Figure 1:** Flow of hospitals between quintiles of TPS in 2012–2013.



**Figure 2:** Flow of hospitals between quintiles of TPS in 2018–2019.

and demonstrates how hospitals change their position in quintiles of the composite quality measure after the introduction of pay-for-performance (e.g., from 2012 to 2013).

As can be inferred from Figure 1, there is considerable movement of hospitals between quintiles. For instance, consider hospitals which in 2012 belonged to the fifth quintile of TPS (quintile with the highest performance). Fewer than half of these hospitals remained in the fifth quintile of TPS in 2013, and the rest saw a decline of their position relative to other hospitals by moving to quintiles one through four. Similar tendencies are observed for hospitals in any other given quintile of TPS in 2012: only a small share of hospitals continue to belong to the same quintile in the subsequent year. This can be viewed as graphic support for the phenomenon of mean reversion since hospitals would rarely change their quintile from year to year in the absence of mean reversion.

It is plausible to assume that mean reversion becomes weaker when there is an increase of  $\alpha$ . Figure 2 supports this prediction. It shows the flow of hospitals between quintiles of TPS from 2018 to 2019, when the value of  $\alpha$  was 0.02. Compared to Figure 1 with  $\alpha$  equal to 0.01 in 2013, the flows in 2018–2019 are much weaker than the flows in 2012–2013, so hospitals change their position in quintiles less often.

## 5 Empirical results

The first set of our results is reported in Table 2 and concerns the mean effect of pay-for-performance at Medicare hospitals.

**Table 2:** Effect of pay-for-performance on the mean quality

	2013	2014	2015	2016	2017	2018	2019
$\alpha_t$	1.00	1.25	1.50	1.75	2.00	2.00	2.00
$\mu(\alpha_t)$	30.546*** (0.973)	31.879*** (0.678)	34.823*** (0.385)	36.226*** (0.453)	38.407*** (0.970)	38.684*** (0.961)	38.841*** (0.956)
$\mu(\alpha_t) - \mu(0)$	2.255*** (0.737)	3.179*** (1.009)	4.762*** (1.349)	6.487*** (1.800)	8.461*** (2.355)	8.530*** (2.309)	8.495*** (2.266)
$\mu(\alpha_t) - \mu(\alpha_{t-1})$	2.255*** (0.737)	1.333*** (0.375)	2.944** (0.474)	1.403*** (0.542)	2.181** (0.634)	0.277 (0.240)	0.157 (0.237)

Notes: Standard errors calculated using the delta-method are in parentheses. \*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

Measured as  $\mu(\alpha_t) - \mu(0)$ , the mean effect of pay-for-performance is positive in 2013–2019. The value of the effect  $\mu(\alpha_t)$  increases in  $\alpha_t$  in 2013–2017. However, the increase in 2017–2019 is negligible and is in line with the fact that  $\alpha$  has remained flat since 2017. Similarly, the change in the effect of pay-for-performance in consecutive years, defined as  $\mu(\alpha_t) - \mu(\alpha_{t-1})$ , is positive for 2013–2017 but is extremely small and statistically insignificant in 2018–2019 in comparison with the previous years. This finding corresponds to our hypothesis *H1a* of improvement in mean quality owing to the introduction of pay-for-performance (i.e., the increase of  $\alpha$  from 0 to 1) and of expected rise of mean quality due to the linearly increasing reward function ( $\alpha$  gradually goes up from 1 to 2 in 2013–2017).

Note that the mean value of  $\mu(\alpha_t)$  increases in  $\alpha_t$ , which supports our supposition that hospital managers take account of future benefits from improving current values of hospital quality. Table 3 shows the second set of results for heterogeneity of hospital response to pay-for-performance. The parameter  $\lambda$  is estimated as the inverse of the smaller root of AR(2) or as ACF(1). The values are significant and less than one under both approaches. This points to mean reversion, so quality decreases toward the mean at high-quality hospitals and goes up toward the mean at hospitals with low quality. The values of  $\lambda$  rise with an increase in the size of incentives  $\alpha$ , which implies that the persistence of the dynamic process increases, and

**Table 3:** Effect of pay-for-performance on mean reversion

	2013	2014	2015	2016	2017	2018	2019
$\alpha_t$	1.00	1.25	1.50	1.75	2.00	2.00	2.00
$\lambda(\alpha_t)$	0.286*** (0.112)	0.435*** (0.032)	0.531*** (0.020)	0.598*** (0.017)	0.659*** (0.016)	0.651*** (0.016)	0.642*** (0.016)
$\lambda(\alpha_t) - \lambda(0)$	-0.169 (0.151)	-0.020 (0.069)	0.076 (0.055)	0.144*** (0.048)	0.204*** (0.044)	0.196*** (0.044)	0.187*** (0.045)
$\lambda(\alpha_t) - \lambda(\alpha_{t-1})$	-0.169 (0.151)	0.149* (0.082)	0.096*** (0.016)	0.068*** (0.009)	0.061*** (0.007)	-0.008* (0.005)	-0.009* (0.005)
$\lambda(\alpha_t)$ (alternative)	0.408*** (0.018)	0.442*** (0.015)	0.482*** (0.013)	0.519*** (0.013)	0.561*** (0.015)	0.554*** (0.014)	0.548*** (0.014)
$\lambda(\alpha_t) - \lambda(0)$ (alternative)	0.132*** (0.019)	0.166*** (0.024)	0.206*** (0.029)	0.244*** (0.034)	0.285*** (0.039)	0.278*** (0.038)	0.272*** (0.038)
$\lambda(\alpha_t) - \lambda(\alpha_{t-1})$ (alternative)	0.132*** (0.019)	0.034*** (0.005)	0.040*** (0.006)	0.037*** (0.006)	0.041*** (0.006)	-0.006* (0.003)	-0.006* (0.003)

Notes: Standard errors calculated using the delta-method are in parentheses. \*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

The persistence parameter  $\lambda(\alpha_t)$  is estimated as the inverse of the smaller root of AR(2) or as ACF(1), the latter is denoted as “alternative.”

Since the values of  $\lambda(\alpha_t)$  are well below 1, we can conclude that the estimated AR(2) processes are indeed stationary for each  $\alpha_t$ .



**Table 4:** Effect of pay-for-performance as  $\mu(\alpha_t) - \mu(0)$  for quintiles of  $\text{TPS}_{t-1}$ 

	2013	2014	2015	2016	2017	2018	2019
Quintile 1	1.828*** (0.741)	2.910*** (1.007)	4.149*** (1.334)	5.798*** (1.727)	7.673*** (2.264)	7.518*** (2.162)	7.683*** (2.202)
Quintile 2	2.150*** (0.745)	2.881*** (1.013)	4.396*** (1.358)	5.953*** (1.803)	8.293*** (2.375)	8.063*** (2.293)	8.360*** (2.250)
Quintile 2 minus Quintile 1	0.321** (0.138)	-0.029 (0.124)	0.247 (0.174)	0.155 (0.233)	0.620* (0.328)	0.545* (0.327)	0.677** (0.313)
Quintile 3	2.187*** (0.738)	3.072*** (1.023)	4.279*** (1.378)	6.368*** (1.849)	8.093*** (2.346)	8.602*** (2.381)	8.301*** (2.278)
Quintile 3 minus Quintile 2	0.037 (0.078)	0.191* (0.114)	-0.117 (0.155)	0.416* (0.232)	-0.200 (0.321)	0.539 (0.356)	-0.059 (0.355)
Quintile 4	2.298*** (0.739)	3.218*** (1.027)	4.742*** (1.392)	6.534*** (1.850)	8.287*** (2.391)	8.260*** (2.338)	8.087*** (2.279)
Quintile 4 minus Quintile 3	0.110 (0.074)	0.146 (0.111)	0.463** (0.230)	0.165 (0.235)	0.194 (0.368)	-0.342 (0.419)	-0.214 (0.349)
Quintile 5	2.483*** (0.741)	3.261*** (1.030)	5.135*** (1.418)	6.529*** (1.849)	8.381*** (2.506)	8.282*** (2.476)	8.501*** (2.421)
Quintile 5 minus Quintile 4	0.186* (0.101)	0.043 (0.251)	0.393 (0.503)	-0.004 (0.433)	0.094 (0.568)	0.022 (0.503)	0.414 (0.497)

Notes: quintile 1 denotes the lowest quality and quintile 5 – the highest. The table reports the effect at each corresponding quintile and the differences in the effects at consecutive quintiles.

\*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

Standard errors (calculated using the delta-method for the difference of the reform effects across the corresponding two categories of each time-invariant hospital characteristic) are in parentheses.

There are two sources of errors in the estimates shown in the table: the error of the regression coefficient and the error of the mean values of covariates. The first part of the error does not vary across all result tables, while the second part of the error depends on the group size and is approximately  $\sqrt{5}$  times larger than its counterpart in Table 2. However, the errors of the regression coefficient are considerably bigger than those of mean values of covariates, so the increase in the standard errors in this table and two subsequent tables relative to the standard error in Table 2 is only minor.

hence the effect of mean reversion becomes weaker. Similarly, the values of  $\lambda(\alpha_t) - \lambda(0)$  are positive and increase in  $\alpha_t$ . The time change in the convergence parameter:  $\lambda(\alpha_t) - \lambda(\alpha_{t-1})$  is positive for 2013–2017. The results support hypothesis *H1b* of weakening of quality convergence to the mean value with a rise in  $\alpha$ . (The value of  $\lambda(\alpha_t) - \lambda(\alpha_{t-1})$  approaches zero in 2018–2019, when parameter  $\alpha$  remains flat.)

The heterogeneous changes in hospital quality owing to pay-for-performance are given in Tables 4, 5, 6 where hospitals are divided into quintiles according to the values of their TPS. Note that the change in hospital quality is a function of the regression coefficient and the mean values of covariates. So its standard error consists of two parts: the error of the estimated regression coefficient and the error of the mean values of covariates. Only the second part of this error depends on sample size and should go up approximately  $\sqrt{5}$  times due to analysis by quintiles. However, the weight of this second part proves to be relatively small in case of our data, so the standard errors in Tables 4–6 are only slightly larger than standard errors in Table 2.

The estimates of the effect of pay-for-performance in terms of  $\mu(\alpha_t) - \mu(0)$  or in terms of  $\mu(\alpha_t) - \mu(\alpha_{t-1})$  show that the higher the quintile of the quality distribution in the previous year, the larger is the impact of the reform (Tables 4 and 5). Statistically significant differences in the effect of pay-for-performance across consecutive quintiles of lagged TPS are observed in many years, for instance, in 4 years out of 7 for quintiles 1–2 in case of  $\mu(\alpha_t) - \mu(0)$  and for quintiles 4–5 in case of  $\mu(\alpha_t) - \mu(\alpha_{t-1})$ . The change of the effect of pay-for-performance over time  $\mu(\alpha_t) - \mu(\alpha_{t-1})$  increases with a rise of the quality incentive  $\alpha$  but almost stops increasing in 2018–2019 when  $\alpha$  becomes constant, as shown in Table 5. So pay-for-performance stimulates quality increase in all groups of Medicare's hospitals, and the impact of pay-for-performance is greater at higher-quality hospitals.

**Table 5:** Effect of pay-for-performance as  $\mu(\alpha_t) - \mu(\alpha_{t-1})$  by quintiles of  $TPS_{t-1}$ 

	2013	2014	2015	2016	2017	2018	2019
Quintile 1	1.828*** (0.741)	-0.257 (0.517)	1.032** (0.497)	2.442*** (0.582)	1.738*** (0.656)	0.536 (0.440)	-0.215 (0.434)
Quintile 2	2.150*** (0.745)	0.678 (0.489)	2.131*** (0.546)	1.908*** (0.626)	2.405*** (0.736)	-0.125 (0.465)	0.655 (0.505)
Quintile 2 minus Quintile 1	0.321** (0.138)	0.935* (0.522)	1.099** (0.501)	-0.534 (0.545)	0.667 (0.585)	-0.661 (0.629)	0.869 (0.676)
Quintile 3	2.187*** (0.738)	0.933** (0.458)	1.561*** (0.530)	3.086*** (0.669)	1.940*** (0.723)	0.743 (0.501)	0.104 (0.531)
Quintile 3 minus Quintile 2	0.037 (0.078)	0.255 (0.490)	-0.570 (0.514)	1.178** (0.545)	-0.465 (0.609)	0.868 (0.677)	-0.551 (0.758)
Quintile 4	2.298*** (0.739)	1.656** (0.464)	3.222*** (0.626)	0.929 (0.702)	2.397*** (0.789)	0.505 (0.552)	-0.045 (0.510)
Quintile 4 minus Quintile 3	0.110 (0.074)	0.722 (0.500)	1.662*** (0.565)	-2.157*** (0.630)	0.458 (0.654)	-0.238 (0.745)	-0.149 (0.748)
Quintile 5	2.483*** (0.741)	3.445*** (0.620)	6.247*** (0.864)	-1.518* (0.875)	2.105*** (0.892)	-0.631 (0.583)	0.666 (0.567)
Quintile 5 minus Quintile 4	0.186* (0.101)	1.789*** (0.599)	3.024*** (0.829)	-2.447*** (0.876)	-0.292 (0.780)	-1.136 (0.803)	0.711 (0.762)

Notes: quintile 1 denotes the lowest quality and quintile 5 – the highest. The table reports the effect at each corresponding quintile and the differences in the effects at consecutive quintiles.

\*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

Standard errors calculated using the delta-method are in parentheses.

**Table 6:** Net total effect by quintiles of  $TPS_{t-1}$ : Predicted TPS minus lagged TPS

	2013	2014	2015	2016	2017	2018	2019
Quintile 1	2.686*** (0.458)	4.989*** (0.329)	1.153*** (0.298)	7.715*** (0.271)	5.647*** (0.263)	5.351*** (0.273)	0.226 (0.266)
Quintile 2	-3.796*** (0.288)	0.708*** (0.249)	-2.275*** (0.236)	4.388*** (0.223)	2.586*** (0.212)	2.145*** (0.224)	-2.572*** (0.238)
Quintile 2 minus Quintile 1	-6.482*** (0.367)	-4.281*** (0.285)	-3.428*** (0.254)	-3.327*** (0.215)	-3.061*** (0.207)	-3.205*** (0.221)	-2.799*** (0.228)
Quintile 3	-7.249*** (0.282)	-1.938*** (0.236)	-5.054*** (0.237)	2.405*** (0.218)	0.646*** (0.216)	0.383* (0.228)	-4.523*** (0.235)
Quintile 3 minus Quintile 2	-3.453*** (0.276)	-2.646*** (0.243)	-2.778*** (0.212)	-1.982*** (0.201)	-1.940*** (0.193)	-1.762*** (0.214)	-1.950*** (0.231)
Quintile 4	-10.783*** (0.321)	-4.199*** (0.273)	-6.789*** (0.280)	-0.182 (0.250)	-1.337*** (0.256)	-1.861*** (0.272)	-6.893*** (0.270)
Quintile 4 minus Quintile 3	-3.534*** (0.289)	-2.261*** (0.243)	-1.736*** (0.247)	-2.587*** (0.230)	-1.983*** (0.228)	-2.244*** (0.267)	-2.370*** (0.229)
Quintile 5	-15.403*** (0.439)	-8.318*** (0.402)	-10.637*** (0.476)	-3.815*** (0.388)	-4.806*** (0.381)	-5.654*** (0.381)	-10.791*** (0.396)
Quintile 5 minus Quintile 4	-4.620*** (0.353)	-4.119*** (0.330)	-3.847*** (0.406)	-3.634*** (0.336)	-3.469*** (0.330)	-3.794*** (0.323)	-3.898*** (0.330)

Notes: quintile 1 denotes the lowest quality and quintile 5 – the highest. The table reports the effect at each corresponding quintile and the differences in the effects at consecutive quintiles.

\*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

Standard errors calculated using the delta-method are in parentheses.

**Table 7:** Effect of pay-for-performance as  $\mu(\alpha_t) - \mu(0)$  by hospital ownership, teaching status, and urban location

	2013	2014	2015	2016	2017	2018	2019
Public	2.099*** (0.723)	2.945*** (0.984)	4.226*** (1.304)	5.869*** (1.739)	7.413*** (2.254)	7.365*** (2.203)	7.475*** (2.148)
Private	2.274*** (0.740)	3.214*** (1.015)	4.858*** (1.360)	6.595*** (1.814)	8.637*** (2.377)	8.726*** (2.333)	8.656*** (2.289)
Private minus Public	0.174 (0.118)	0.269 (0.168)	0.632*** (0.256)	0.725** (0.323)	1.224*** (0.432)	1.361*** (0.437)	1.181*** (0.414)
Urban	2.383*** (0.739)	3.291*** (1.003)	4.735*** (1.314)	6.253*** (1.721)	8.059*** (2.203)	8.037*** (2.146)	7.929*** (2.104)
Rural	2.048*** (0.741)	3.069*** (1.070)	4.353*** (1.494)	6.649*** (2.054)	9.195*** (2.802)	9.275*** (2.706)	9.321*** (2.578)
Rural minus Urban	-0.335 (0.227)	-0.222 (0.333)	-0.382 (0.493)	0.396 (0.599)	1.136 (0.832)	1.239* (0.749)	1.392** (0.615)
Teaching	2.363*** (0.755)	3.294*** (1.013)	4.667*** (1.334)	6.327*** (1.735)	8.177*** (2.224)	8.284*** (2.198)	8.309*** (2.156)
Non-teaching	2.238*** (0.732)	3.177*** (1.018)	4.807*** (1.373)	6.546*** (1.854)	8.664*** (2.467)	8.688*** (2.407)	8.610*** (2.362)
Non-teaching minus Teaching	-0.125 (0.179)	-0.118 (0.236)	0.140 (0.366)	0.219 (0.430)	0.487 (0.609)	0.404 (0.578)	0.301 (0.560)

Notes: Standard errors (calculated using the delta-method for the difference of the reform effects across the corresponding two categories of each time-invariant hospital characteristic) are in parentheses.

\*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

Table 6 gives estimates of the net total effect, i.e., the expected change in hospital quality over time, measured as the difference between the predicted TPS and the lagged TPS. The net total effect is the sum of the impact of mean reversion and the effect of pay-for-performance.

Note that the estimation of the fitted value of TPS includes time effects which account both for time trend and for important changes in the incentive mechanism not captured by variation in  $\alpha$ . An example of such change occurred in 2015 and temporarily decreased the value of TPS for each hospital.<sup>17</sup> Accordingly, Table 6 shows that the values of predicted TPS minus lagged TPS go down in 2015 for each quintile.

The values of net total effect reveal an increase of quality in the groups of low-quality hospitals, while quality deteriorates in high-quality groups. Negative total effect is less prevalent or is smaller in absolute terms at high-quality hospitals in 2016–2017. The result can be attributed to the weakening of mean reversion with increase in  $\alpha$ . Yet, when  $\alpha$  becomes constant in 2018–2019, the prevalence of negative total effect and the absolute value of the negative effect returns to that of 2015.

Finally, we focus on the effect of pay-for-performance for groups of Medicare hospitals according to their ownership, teaching status, urban location, and geographic region. The mean effect increases in  $\alpha$  for public and private hospitals, for urban and rural hospitals, for teaching and non-teaching hospitals, and for hospitals in each geographic region (Tables 7 and 8).

The effect of pay-for-performance is greater for private hospitals than for public hospitals, which corresponds to findings in [13] and [78]. The result can be explained by a greater emphasis on financial incentives at these healthcare institutions. These profit constraints, combined with the altruistic character of healthcare services, induce more effective quality competition at non-public hospitals [16]. The difference in the effect for private and public hospitals is statistically significant in most years.

As for teaching status, quality improvement owing to the incentive scheme is often higher at non-teaching hospitals, which may be because they can devote all of their labor resources to patient treatment,

<sup>17</sup> Specifically, the pneumonia cohort was expanded and it caused a rise in pneumonia readmission rates in 2015. Additionally, the safety domain with relatively low scores in comparison to measures of other domains was added to the list of measures which constitute TPS.

**Table 8:** Effect of pay-for-performance as  $\mu(\alpha_t) - \mu(0)$  for hospitals at different geographic regions

	2013	2014	2015	2016	2017	2018	2019
New England	1.864** (0.815)	2.080 (1.264)	3.605** (1.752)	6.858*** (2.105)	10.211*** (3.078)	10.071*** (2.976)	9.934*** (2.888)
Mid-Atlantic	1.800*** (0.737)	2.550*** (0.999)	3.617*** (1.321)	5.189*** (1.739)	7.013*** (2.253)	7.194*** (2.219)	7.030*** (2.146)
Mid-Atlantic minus New England	-0.064 (0.328)	0.470 (0.702)	0.012 (0.978)	-1.670*** (0.574)	-3.198*** (1.082)	-2.877*** (0.997)	-2.904*** (0.981)
East North Central	2.078*** (0.751)	3.084*** (1.036)	4.773*** (1.404)	6.552*** (1.865)	8.142*** (2.393)	8.205*** (2.387)	8.246*** (2.347)
East North Central minus New England	0.215 (0.316)	1.004 (0.694)	1.168 (0.955)	-0.307 (0.451)	-2.069** (0.905)	-1.866*** (0.801)	-1.688** (0.775)
West North Central	2.318*** (0.741)	3.329*** (1.052)	5.154*** (1.431)	7.342*** (2.025)	10.138*** (2.842)	10.140*** (2.797)	10.172*** (2.765)
West North Central minus New England	0.455 (0.332)	1.249* (0.709)	1.549 (1.001)	0.483 (0.520)	-0.073 (0.808)	0.069 (0.780)	0.238 (0.757)
South Atlantic	2.248*** (0.761)	3.210*** (1.040)	4.946*** (1.422)	6.803*** (1.887)	8.886*** (2.481)	8.863*** (2.407)	8.744*** (2.374)
South Atlantic minus New England	0.384 (0.326)	1.131 (0.699)	1.341 (0.949)	-0.056 (0.434)	-1.325 (0.826)	-1.208 (0.788)	-1.190 (0.760)
East South Central	2.295*** (0.780)	3.118*** (1.065)	4.962*** (1.492)	6.913*** (2.042)	9.422*** (2.803)	8.821*** (2.597)	8.643*** (2.460)
East South Central minus New England	0.432 (0.356)	1.038 (0.720)	1.357 (0.989)	0.054 (0.517)	-0.789 (0.719)	-1.250* (0.705)	-1.291* (0.726)
West South Central	2.276*** (0.735)	3.225*** (0.990)	5.193*** (1.325)	6.324*** (1.755)	7.926*** (2.241)	8.380*** (2.227)	8.274*** (2.175)
West South Central minus New England	0.413 (0.347)	1.146 (0.723)	1.588 (1.022)	-0.535 (0.583)	-2.286* (1.061)	-1.691* (0.969)	-1.659* (0.933)
Mountain	1.795*** (0.647)	2.809*** (0.863)	4.303*** (1.119)	5.537*** (1.415)	7.291*** (1.809)	7.686*** (1.819)	7.941*** (1.861)
Mountain minus New England	-0.069 (0.371)	0.729 (0.756)	0.698 (1.073)	-1.322 (0.869)	-2.920** (1.468)	-2.385* (1.345)	-1.993 (1.242)
Pacific	2.524*** (0.716)	3.324*** (0.957)	4.276*** (1.238)	5.957*** (1.613)	7.923*** (2.101)	7.910*** (2.067)	8.190*** (2.066)
Pacific minus New England	0.661* (0.388)	1.245 (0.765)	0.671 (1.072)	-0.901 (0.827)	-2.288* (1.315)	-2.161* (1.245)	-1.744 (1.189)

Notes: Standard errors (calculated using the delta-method for the difference of the reform effects across New England hospitals and hospitals in each corresponding geographic region) are in parentheses.

\*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

while teaching hospitals lose some efficiency due to their educational activities [64]. Also, teaching hospitals may be treating more difficult cases. This complexity could not be fully captured by the casemix variable in our analysis and may cause a downward bias of the estimated effect at teaching hospitals, explaining the lower value of the effect at teaching than at non-teaching hospitals. Yet, the difference in the values at teaching and non-teaching hospitals is statistically insignificant in each year.

Statistically significant differences in the effect of pay-for-performance for urban and rural hospitals are observed only in the last 2 years: the effect is larger at rural hospitals.

As regards geographic location, there is practically no variation in the effect across groups of hospitals in the early years of pay-for-performance. The differences are present mainly in the later few years: for instance, the mean effect of pay-for-performance is greater in New England than in Mid-Atlantic in 2016–2019 and than in East North Central and West South Central regions in 2017–2019.

## 6 Discussion

In this article, we focused on exclusion of mean reversion in evaluating the response of TPS at Medicare hospitals to an incentive contract. Since TPS under this contract becomes an autoregressive process, our analysis deals with dynamic panels.

It should be noted that dynamic panel data models are prevalent in various fields of economics. Examples in macroeconomics include the analysis of a country's growth [11,50] or its current account [81]. Application in corporate finance deals with the study of such firm-level variables as size [33,61], profit [54], leverage [32,36], and such proxies of firm performance as return on asset and Tobin's Q [49,65]. In the banking sphere dynamic panels are applied to ROE and profitability [35,48] while in finance they are used for housing prices [31] and fuel prices [71]. Papers in the economics of labor, health, and welfare employ dynamic panel data models to analyze physician labor supply [4], hospital staffing intensity [82], wealth of households and health status of individuals [57], and quality and efficiency of hospitals (e.g., mortality ratio in [56], and average length of stay in [10]).

The approach used in our study estimates the unconditional mean of the dependent variable in the dynamic panel data model and employs it for policy evaluation. Specifically, the comparison of the fitted values of the unconditional mean at different values of policy intensity offers a measure of the effect of reform. The advantages of the approach are twofold. First, it excludes the impact of mean reversion in groupwise estimations (e.g., in lower and in higher quantiles of hospitals according to their TPS). Second, the approach may also be used in the analysis of the mean effect of the reform if we focus on effects in the long run. Indeed, the unconditional mean in dynamic panel data analysis is sometimes called the long-term mean as it reflects the mean value in the long run. It should be noted that an alternative approach that uses the estimated coefficient for the policy variable as a measure of the mean effect of reform does not suffer from the problem of mean reversion. But in dynamic panel data models it evaluates only the short-term impact of policy.

As regards exclusion of mean reversion in dynamic panels, we note a limitation on the character of mean reversion, imposed by the nature of the dynamic panel model where the unconditional mean is the long-term mean. Mean reversion is not instantaneous: if a deviation from the mean is observed in period  $t$ , the return to the mean occurs not in period  $t + 1$  but only in later periods.

It may be noted that our approach is similar to difference-in-difference estimations. The long-run effect of reform under our approach is the difference in the fitted value of the long-term mean under the value  $\alpha_t$  and under counterfactual value of zero (similar to [48]). Alternatively, we can take the difference in the fitted values of the long-term means under the value of  $\alpha_t$  and  $\alpha_{t-1}$ . To summarize: in focusing on the long-run impact of the reform in dynamic panel data models, the estimation of either the mean effect or of the groupwise effects requires an unconditional mean. The approach also excludes mean reversion which contaminates policy evaluation in case of groupwise estimations.

As regards policy evaluation based on panel data fixed effects methodology, our approach of computing the unconditional mean as a function of the policy variable  $\alpha$  produces the conventional linear prediction of the dependent variable. The mean effect of reform in the static panel is either the coefficient for the reform variable or the difference in the fitted value of  $y$  under  $\alpha_t$  and the fitted value of  $y$  under 0 (counterfactual).

Finally, we note the prerequisites for identification of the unconditional mean which are similar to the assumptions in difference-in-difference estimations. Two requirements apply both to the static and dynamic panels. First, time variation in the policy variable is required for identification of the coefficient for the policy variable in the unconditional mean function. Second, if there is only time variation in the policy variable  $\alpha$  (and no cross-section variation in  $\alpha_t$  at a given value of  $t$ , i.e., no control group), the reform effect cannot be distinguished from other time effects. So cross-section variation in another variable, which is correlated with the policy variable, is required. In our case, this variable is the Medicare share: the higher the share of Medicare patients in the hospital, the stronger the impact of  $\alpha$  (the share of hospital funds at risk under the Medicare program becomes more important for total revenues of the hospital). The use of

dynamic panel data models requires a third assumption: the unconditional mean must be defined, and for this reason the process  $y$  has to be stationary.

## 7 Conclusion

Studies of incentive contracts usually focus on the mean tendency and give scant attention to potentially heterogeneous response to the policy of interest by agents at different percentiles of the distribution of the dependent variable. But insufficient analysis of such heterogeneity may lead to speculation on ceiling effects and belief among agents with better values of the variable of interest that there are no ways of making further financial gains by further improvements.

This article highlights the fact that there is a multivariate dependence of the variable of interest in such incentive contracts. Specifically, a part of intertemporal dependence can be attributed to the policy reform and a part to mean reversion. So the article proposes a method to help model such multivariate dependence by excluding the impact of mean reversion. As mean reversion contaminates judgment regarding the time profile of the dependent variable, and this contamination is different for agents in lower and higher percentiles of the variable of interest, clearing out the reform effect of mean reversion makes the method suitable for assessing heterogeneity of incentive schemes.

In an application to the longitudinal data for Medicare's acute-care hospitals taking part in the nationwide quality incentive mechanism ("value-based purchasing"), we find that the higher the quintile of quality in the prior period, the larger the increase in the composite quality measure owing to the reform. Quality improvement in each quintile increases with the increase in size of the quality incentive.

Our results reveal that increase in the quality measure owing to pay-for-performance is greater at hospitals with higher levels of quality. The finding suggests stronger emphasis on quality activities at high-quality hospitals, and this is indeed discovered in a number of works. For instance, top-performing hospitals in the US pilot program paid more attention to quality enhancement than bottom-performing hospitals [77]. Under the proportional pay-for-performance mechanism in California, high-quality physicians similarly placed more emphasis on an organizational culture of quality and demonstrated stronger dedication to addressing quality issues than low-quality physicians [21]. The desire of high-quality hospitals, which have reached top deciles of hospital performance, to pursue quality improvement by means additional to those proposed by the policy regulator is further evidence in support of our research [37].

Directions for future work in health economics applications may include analysis of heterogeneous hospital response to quality incentives by considering different dimensions of the composite quality measure. A related field of research is the study of potential sacrifice of quality of non-incentivized measures in favor of measures incentivized by pay-for-performance. This has been analyzed at the mean level [27,47] and may be expanded to account for different behavior by high-quality and low-quality hospitals.

**Acknowledgements:** This article was prepared in the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE). The authors are grateful to the editor and two anonymous referees for helpful comments.

**Conflict of interest:** The authors state no conflict of interest.

## Appendix

### A Estimation with the dynamic panel

**Table A1:** Explaining the TPS in Medicare's acute care hospitals

	TPS	
$L(\text{TPS})$	0.333***	(0.045)
$L^2(\text{TPS})$	-0.207***	(0.040)
$L(\text{TPS}) \cdot \text{medicare share} \cdot \alpha$	0.229***	(0.070)
$L^2(\text{TPS}) \cdot \text{medicare share} \cdot \alpha$	0.418***	(0.065)
Medicare share	-5.201	(5.530)
Medicare share $\cdot \alpha$	-39.873***	(4.904)
Casemix	9.307***	(1.522)
Physicians/beds	1.092***	(0.187)
physicians/beds $\cdot \text{medicare share} \cdot \alpha$	-1.577***	(0.220)
Nurses/beds	0.007	(0.030)
Nurses/beds $\cdot \text{medicare share} \cdot \alpha$	-0.099*	(0.052)
Dsh	4.210	(3.695)
$\log(\text{beds})$	-7.537***	(0.653)
$\log(\text{beds}) \cdot \text{medicare share} \cdot \alpha$	4.794***	(0.854)
HRRP penalty	0.417*	(0.227)
MUEHR	1.955*	(1.061)
year = 2013	3.327***	(0.749)
year = 2014	2.827***	(0.416)
year = 2015	-2.174***	(0.182)
year = 2016	0.996***	(0.201)
year = 2017	0.274	(0.372)
year = 2018	-0.185	(0.379)
year = 2019	-5.065***	(0.390)
Constant	49.349***	(5.317)
Observations	18,545	
Hospitals	2,984	
Arellano–Bond test statistic	-0.881	
$R^2$ (without individual effects)	0.494	
$R^2$ (with individual effects)	0.781	

Notes: The sum of coefficients for annual dummies is normalized to zero.

Robust standard errors are in parentheses.

\*, \*\*, and \*\*\* show significance at levels of 0.1, 0.05, and 0.01, respectively.

The Sargan statistic is not applicable to the specification with robust standard errors.

## B Data sources

Total performance scores and other Hospital Compare data were downloaded from <https://data.medicare.gov/data/hospital-compare> (Table A2).

TPSs and other Hospital Compare data were downloaded from <https://data.medicare.gov/data/hospital-compare> (Table A3).

Provider of Service data come from <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services> (Table A4).

Impact Files data are taken from <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS> (Table A5).

**Table A2:** List of variables and databases

Variable	Source
TPS	Hospital compare
Casemix	Impact files
Dsh	Impact files
Medicare share	Impact files
Urban	Impact files
Public	Hospital compare
Physicians	Provider of service files
Nurses	Provider of service files
Teaching	Provider of service files
Beds	Provider of service files
Regional dummies	Hospital compare
HRRP penalty	HRRP supplemental files to acute inpatient PPS final rules
MUEHR	EHR incentive program eligible hospitals public use files

**Table A3:** Download links for the TPS data

Year	Date	Link
2011	10/01/2013	<a href="https://medicare.gov/download/HospitalCompare/2013/October/HOSArchive_Revised_Flatfiles_20131001.zip">https://medicare.gov/download/HospitalCompare/2013/October/HOSArchive_Revised_Flatfiles_20131001.zip</a>
2012	10/23/2014	<a href="https://medicare.gov/download/HospitalCompare/2014/October/HOSArchive_Revised_Flatfiles_20141023.zip">https://medicare.gov/download/HospitalCompare/2014/October/HOSArchive_Revised_Flatfiles_20141023.zip</a>
2013	10/08/2015	<a href="https://medicare.gov/download/HospitalCompare/2015/October/HOSArchive_Revised_FlatFiles_20151008.zip">https://medicare.gov/download/HospitalCompare/2015/October/HOSArchive_Revised_FlatFiles_20151008.zip</a>
2014	11/10/2016	<a href="https://medicare.gov/download/HospitalCompare/2016/October/Hospital_Revised_FlatFiles_20161110.zip">https://medicare.gov/download/HospitalCompare/2016/October/Hospital_Revised_FlatFiles_20161110.zip</a>
2015	10/24/2017	<a href="https://medicare.gov/download/HospitalCompare/2017/October/HOSArchive_Revised_FlatFiles_20171024.zip">https://medicare.gov/download/HospitalCompare/2017/October/HOSArchive_Revised_FlatFiles_20171024.zip</a>
2016	10/31/2018	<a href="https://medicare.gov/download/HospitalCompare/2018/October/HOSArchive_Revised_FlatFiles_20181031.zip">https://medicare.gov/download/HospitalCompare/2018/October/HOSArchive_Revised_FlatFiles_20181031.zip</a>
2017	10/30/2019	<a href="https://medicare.gov/download/HospitalCompare/2019/October/HOSArchive_Revised_Flatfiles_20191030.zip">https://medicare.gov/download/HospitalCompare/2019/October/HOSArchive_Revised_Flatfiles_20191030.zip</a>
2018	01/29/2020	<a href="https://medicare.gov/download/HospitalCompare/2020/January/HOSArchive_Revised_Flatfiles_20200129.zip">https://medicare.gov/download/HospitalCompare/2020/January/HOSArchive_Revised_Flatfiles_20200129.zip</a>
2019	01/31/2021	<a href="https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2021/hospitals_01_2021.zip">https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2021/hospitals_01_2021.zip</a>



**Table A4:** Download links for the Provider of Services data

Year	Link
2011	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2011">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2011</a>
2012	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2012">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2012</a>
2013	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2013">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2013</a>
2014	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2014">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2014</a>
2015	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2015">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2015</a>
2016	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2016">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2016</a>
2017	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2017">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2017</a>
2018	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2018">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2018</a>
2019	<a href="https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2019">https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2019</a>

**Table A5:** Download links for the Impact Files data

Year	Link
2011	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_13_FR_Impact_File.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_13_FR_Impact_File.zip</a>
2012	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_14_FR_Impact_File.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_14_FR_Impact_File.zip</a>
2013	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2015-FR-Impact-File.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2015-FR-Impact-File.zip</a>
2014	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2016-CMS-1632-FR-Impact.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2016-CMS-1632-FR-Impact.zip</a>
2015	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2017-CMS-1655-FR-Impact.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2017-CMS-1655-FR-Impact.zip</a>
2016	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2018-CMS-1677-FR-Impact.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2018-CMS-1677-FR-Impact.zip</a>
2017	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2019-CMS-1694-FR-Impact.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2019-CMS-1694-FR-Impact.zip</a>
2018	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2020-FR-Impact-File.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2020-FR-Impact-File.zip</a>
2019	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2021-FR-Impact-File.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2021-FR-Impact-File.zip</a>

The Hospitals Readmissions Reduction Program data were downloaded from <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Archived-Supplemental-Data-Files> (Table A6).

The EHR (Electronic Hospital Records) Incentive Program (renamed to the Promoting Interoperability (PI) Program) data are taken from <https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/PUF> (Table A7).

**Table A6:** Download links for the HRRP data

Year	Link
2013	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_2013_FR_Readmissions_File.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_2013_FR_Readmissions_File.zip</a>
2014	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/Readmissions-Supplemental-Data-PUF.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/Readmissions-Supplemental-Data-PUF.zip</a>
2015	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2015-FR-Readmit-Supp-Data-File.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2015-FR-Readmit-Supp-Data-File.zip</a>
2016	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2016-CMS-1632-FR-Readmissions.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2016-CMS-1632-FR-Readmissions.zip</a>
2017	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2017-CMS-1655-FR-Hospital-Readmissions.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2017-CMS-1655-FR-Hospital-Readmissions.zip</a>
2018	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2018-FR-Hospital-Readmissions.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2018-FR-Hospital-Readmissions.zip</a>
2019	<a href="https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2019-CMS-1694-FR-Hospital-Readmissions.zip">https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2019-CMS-1694-FR-Hospital-Readmissions.zip</a>

**Table A7:** Download links for the EHR Incentive Program data

Stage	Link
1	<a href="https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EH_PUF_Q32018Stage1.zip">https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EH_PUF_Q32018Stage1.zip</a>
2	<a href="https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EH_PUF_Q32018Stage2.zip">https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EH_PUF_Q32018Stage2.zip</a>

## References

- [1] Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton: Princeton University Press.
- [2] Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58, 277–297.
- [3] Armstrong, C. S., Blouin, J. L., Jagolinzer, A. D., & Larcker, D. F. (2015). Corporate governance, incentives, and tax avoidance. *Journal of Accounting and Economics*, 60(1), 1–17.
- [4] Baltagi, B. H., Bratberg, E., & Holmås, T. H. (2005). A panel data study of physicians' labor supply: The case of Norway. *Health Economics*, 14(10), 1035–1045.
- [5] Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2004). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215–220.
- [6] Barro, R. J. (2013). Inflation and economic growth. *Annals of Economics and Finance*, 14(1), 85–109.
- [7] Bazzi, S., & Clemens, M. A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics*, 5(2), 152–86.
- [8] Beaulieu, N. D., & Horrigan, D. R. (2005). Putting smart money to work for quality improvement. *Health Services Research*, 40(5p1), 1318–1334.
- [9] Besstremyannaya, G. (2011). Managerial performance and cost efficiency of Japanese local public hospitals: A latent class stochastic frontier model. *Health Economics*, 20(S1), 19–34.
- [10] Besstremyannaya, G. (2016). Differential effects of declining rates in a per diem payment system. *Health Economics*, 25(12), 1599–1618.
- [11] Bleaney, M., Gemmell, N., & Kneller, R. (2001). Testing the endogenous growth model: Public expenditure, taxation, and growth over the long run. *Canadian Journal of Economics*, 34(1), 36–57.
- [12] Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.
- [13] Borah, B. J., Rock, M. G., Wood, D. L., Roellinger, D. L., Johnson, M. G., & Naessens, J. M. (2012). Association between value-based purchasing score and hospital characteristics. *BMC Health Services Research*, 12(1), 464.

- [14] Bousquet, F., Bisiaux, R., & Chi, Y.-L. (2014). France: Payment for public health services. In C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, & S. Thomson (Eds.), *Paying for performance in health care* (pp. 141–156). Berkshire, England: Open University Press.
- [15] de Brantes, F. S., & d'Andrea, B. G. (2009). Physicians respond to pay-for-performance incentives: Larger incentives yield greater participation. *American Journal of Managed Care*, 15(5), 305–10.
- [16] Brekke, K. R., Siciliani, L., & Straume, O. R. (2012). Quality competition with profit constraints. *Journal of Economic Behavior and Organization*, 84(2), 642–659.
- [17] Centers for Medicare and Medicaid Services. (2007). *Medicare hospital value-based purchasing plan development. 1st public listening session, January 17, 2007. Issues paper*. U.S. Department of Health and Human Services.
- [18] Christianson, J. B., Leatherman, S., & Sutherland, K. (2008). Lessons from evaluations of purchaser pay-for-performance programs. *Medical Care Research and Review*, 65(6 Suppl), 5S–35S.
- [19] Coleman, K., Reiter, K. L., & Fulwiler, D. (2007). The impact of pay-for-performance on diabetes care in a large network of community health centers. *Journal of Health Care for the Poor and Underserved*, 18(5), 966–983.
- [20] Dafny, L., & Dranove, D. (2008). Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. *The Rand Journal of Economics*, 39(3), 790–821.
- [21] Damberg, C. L., Raube, K., Teleki, S. S., & delaCruz, E. (2009). Taking stock of pay-for-performance: A candid assessment from the front lines. *Health Affairs*, 28(2), 517–525.
- [22] Damberg, C. L., Raube, K., Williams, T., & Shortell, S. M. (2005). Paying for performance: Implementing a statewide project in California. *Quality Management in Healthcare*, 14(2), 66–79.
- [23] Damberg, C. L., Sorbero, M. E., Lovejoy, S. L., Martsolf, G. R., Raaen, L., & Mandel, D. (2014). *Measuring success in health care value-based purchasing programs: Findings from an environmental scan, literature review, and expert panel discussions. Research report*. Santa Monica, CA: RAND Corporation.
- [24] Davis, C. E. (1976). The effect of regression to the mean in epidemiologic and clinical studies. *American Journal of Epidemiology*, 104(5), 493–498.
- [25] Dias, D. A., & Marques, C. R. (2010). Using mean reversion as a measure of persistence. *Economic Modelling*, 27(1), 262–273.
- [26] Doran, T., Fullwood, C., Kontopantelis, E., & Reeves, D. (2008). Effect of financial incentives on inequalities in the delivery of primary clinical care in England: Analysis of clinical activity indicators for the quality and outcomes framework. *The Lancet*, 372(9640), 728–736.
- [27] Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics*, 24(1), 211–223.
- [28] Eijkenaar, F., Emmert, M., Scheppach, M., & Schöffski, O. (2013). Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy*, 110(2), 115–130.
- [29] Friedman, M. (1992). Do old fallacies ever die? *Journal of Economic Literature*, 30(4), 2129–2132.
- [30] Galton, F., & Dickson, H. (1886). Family likeness in stature. *Proceedings of the Royal Society of London*, 40, 42–73.
- [31] Gao, A., Lin, Z., & Na, C. F. (2009). Housing market dynamics: Evidence of mean reversion and downward rigidity. *Journal of Housing Economics*, 18(3), 256–266.
- [32] Gaud, P., Jani, E., Hoesli, M., & Bender, A. (2005). The capital structure of Swiss companies: An empirical analysis using dynamic panel data. *European Financial Management*, 11(1), 51–69.
- [33] Geroski, P. A., Machin, S. J., & Walters, C. F. (1997). Corporate growth and profitability. *The Journal of Industrial Economics*, 45(2), 171–189.
- [34] Glickman, S. W., Ou, F.-S., DeLong, E. R., Roe, M. T., Lytle, B. L., Mulgund, J. ..., Peterson, E. D. (2007). Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*, 297(21), 2373–2380.
- [35] Goddard, J., Molyneux, P., & Wilson, J. O. (2004). The profitability of European banks: A cross-sectional and dynamic panel analysis. *The Manchester School*, 72(3), 363–381.
- [36] González, V. M., & González, F. (2012). Firm size and capital structure: Evidence using dynamic panel data. *Applied Economics*, 44(36), 4745–4754.
- [37] Grossbart, S. R. (2006). What's the return? Assessing the effect of “pay-for-performance” initiatives on the quality of care delivery. *Medical Care Research and Review*, 63(1 Suppl), 29S–48S.
- [38] Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, 114(1), 83–116.
- [39] Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton University Press.
- [40] Hannan, E. L., Kumar, D., Racz, M., Siu, A. L., & Chassin, M. R. (1994). New York State's cardiac surgery reporting system: Four years later. *The Annals of Thoracic Surgery*, 58(6), 1852–1857.
- [41] Hibbard, J. H., Stockard, J., & Tusler, M. (2003). Does publicizing hospital performance stimulate quality improvement efforts? *Health Affairs*, 22(2), 84–94.
- [42] Hibbard, J. H., Stockard, J., & Tusler, M. (2005). Hospital performance reports: Impact on quality, market share, and reputation. *Health Affairs*, 24(4), 1150–1160.
- [43] Higuchi, S. (2010). Chuusho jichitai byouin-no genjo-to kadai (current situation and tasks for small and medium local public hospitals). *Journal of Japan Hospital Association*, 5, 95–101.

- [44] Hillman, A. L., Pauly, M. V., Kerman, K., & Martinek, C. R. (1991). HMO managers' views on financial incentives and quality. *Health Affairs*, 10(4), 207–219.
- [45] Hisamichi, S. (2010). Byouin keiei koto hajime. Byouin jigyou kanri-no tachiba kara (Starting hospital management. Point of view of a hospital manager). *Journal of Japan Hospital Association*, 2, 98–119.
- [46] Jones, N. B. (2014). *Health care executives participating in value-based purchasing: A qualitative phenomenological study*. (Ph.D. thesis). Phoenix: University of Phoenix.
- [47] Kaarbøe, O. M., & Siciliani, L. (2011). Multi-tasking, quality and pay for performance. *Health Economics*, 20(2), 225–238.
- [48] Knapp, M., Gart, A., & Chaudhry, M. (2006). The impact of mean reversion of bank profitability on post-merger performance in the banking industry. *Journal of Banking & Finance*, 30(12), 3503–3517.
- [49] Kyereboah-Coleman, A. (2008). Corporate governance and firm performance in Africa: A dynamic panel data analysis. *Studies in Economics and Econometrics*, 32(2), 1–24.
- [50] Laeven, L., Levine, R., & Michalopoulos, S. (2015). Financial innovation and endogenous growth. *Journal of Financial Intermediation*, 24(1), 1–24.
- [51] Li, J., Hurley, J., DeCicca, P., & Buckley, G. (2014). Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, 23(8), 962–978.
- [52] Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., & Bratzler, D. W. (2007). Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine*, 356(5), 486–496.
- [53] Ma, C.-t.A., & Mak, H. Y. (2015). Information disclosure and the equivalence of prospective payment and cost reimbursement. *Journal of Economic Behavior and Organization*, 117, 439–452.
- [54] Machin, S., & Van Reenen, J. (1993). Profit margins and the business cycle: Evidence from UK manufacturing firms. *The Journal of Industrial Economics*, 41, 29–50.
- [55] Manary, M., Staelin, R., Kosel, K., Schulman, K. A., & Glickman, S. W. (2015). Organizational characteristics and patient experiences with hospital care: A survey study of hospital chief patient experience officers. *American Journal of Medical Quality*, 30(5), 432–440.
- [56] Mark, B. A., Harless, D. W., McCue, M., & Xu, Y. (2004). A longitudinal examination of hospital registered nurse staffing and quality of care. *Health Services Research*, 39(2), 279–300.
- [57] Michaud, P.-C., & van Soest, A. (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics*, 27(5), 1312–1325.
- [58] Morton, V., & Torgerson, D. J. (2003). Effect of regression to the mean on decision making in health care. *British Medical Journal*, 326(7398), 1083–1084.
- [59] Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20(4), 111–132.
- [60] Ogundeji, Y. K., Bland, J. M., & Sheldon, T. A. (2016). The effectiveness of payment for performance in health care: A meta-analysis and exploration of variation in outcomes. *Health Policy*, 120(10), 1141–1150.
- [61] Oliveira, B., & Fortunato, A. (2006). Firm growth and liquidity constraints: A dynamic analysis. *Small Business Economics*, 27(2), 139–156.
- [62] Oxholm, A. S., Kristensen, S. R., & Sutton, M. (2018). Uncertainty about the effort-performance relationship in threshold-based payment schemes. *Journal of Health Economics*, 62, 69–83.
- [63] Pai, C.-W., Finnegan, G. K., & Satwicz, M. J. (2002). The combined effect of public profiling and quality improvement efforts on heart failure management. *The Joint Commission Journal on Quality Improvement*, 28(11), 614–624.
- [64] Pauly, M. (1980). *Doctors and their workshops: Economic models of physician behavior. A National Bureau of Economic Research Monograph*. Chicago and London: The University of Chicago Press.
- [65] Pérez-Calero, L., delMarVillegas, M., & Barroso, C. (2016). A framework for board capital. *Corporate Governance*, 16, 452–475.
- [66] Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158.
- [67] Rosenthal, M. B., Fernandopulle, R., Song, H. R., & Landon, B. (2004). Paying for quality: Providers' incentives for quality improvement. *Health Affairs*, 23(2), 127–141.
- [68] Ryan, A. M. (2009). Effects of the premier hospital quality incentive demonstration on medicare patient mortality and cost. *Health Services Research*, 44(3), 821–842.
- [69] Ryan, A. M., Blustein, J., Michelow, M. D., & Casalino, L. P. (2012). The effect of phase 2 of the premier hospital quality incentive demonstration on incentive payments to hospitals caring for disadvantaged patients. *Health Services Research*, 47(4), 1418–1436.
- [70] Ryan, A. M., Krinsky, S., Maurer, K. A., & Dimick, J. B. (2017). Changes in hospital quality associated with hospital value-based purchasing. *New England Journal of Medicine*, 376(24), 2358–2366.
- [71] Santos, G. F. (2013). Fuel demand in Brazil in a dynamic panel data approach. *Energy Economics*, 36, 229–240.
- [72] Siciliani, L., Straume, O. R., & Cellini, R. (2013). Quality competition with motivated providers and sluggish demand. *Journal of Economic Dynamics and Control*, 37(10), 2041–2061.

- [73] Smith, L. R. (2017). *A case study: The executive leadership response at a community hospital to the value-based purchasing requirements on the patient protection and affordable care act.* (Ph.D. thesis). Jacksonville: University of North Florida, College of Education and Human Services.
- [74] Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in us macroeconomic time series. *Journal of Monetary Economics*, 28(3), 435–459.
- [75] Suwabe, A. (2004). Our efforts on DPC in Iwate Medical University hospital. *Rinsho Byouri*, 52, 1011–1014.
- [76] Vaghela, P., Ashworth, M., Schofield, P., & Gulliford, M. C. (2009). Population intermediate outcomes of diabetes under pay-for-performance incentives in England from 2004 to 2008. *Diabetes Care*, 32(3), 427–429.
- [77] Vina, E. R., Rhew, D. C., Weingarten, S. R., Weingarten, J. B., & Chang, J. T. (2009). Relationship between organizational factors and performance among pay-for-performance hospitals. *Journal of General Internal Medicine*, 24(7), 833.
- [78] Werner, R. M., Kolstad, J. T., Stuart, E. A., & Polsky, D. (2011). The effect of pay-for-performance in hospitals: Lessons for quality improvement. *Health Affairs*, 30(4), 690–698.
- [79] Wilcox, M. A., Chang, A. M., & Johnson, I. R. (1996). The effects of parity on birthweight using successive pregnancies. *Acta Obstetrica et Gynecologica Scandinavica*, 75(5), 459–463.
- [80] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126, 25–51.
- [81] Wu, J.-L. (2000). Mean reversion of the current account: Evidence from the panel data unit-root test. *Economics Letters*, 66(2), 215–222.
- [82] Zhao, M., Bazzoli, G. J., Clement, J. P., Lindrooth, R. C., Nolin, J. M., & Chukmaitov, A. S. (2008). Hospital staffing decisions: Does financial performance matter? *Inquiry*, 45(3), 293–307.