Voluntary Regulation: Evidence from Medicare Bundled Payments

Online Appendix

Liran Einav Amy Finkelstein Yunan Ji Neale Mahoney

Contents

A Imputing Bundle Prices **2**

B Variance Decomposition **3**

C Description of the Gibbs Sampler **4**

D Mapping from Estimates to Results **10**

E Targeting **12**

List of Tables

A.1 Cream Skimming 17

A.2 Experimental Estimates During Mandatory Participation Period: Alternative Specifications 18

A.3 Experimental Estimates: Effects Over Time 19

A.4 Correlates of Levels and Slopes with Hospital-Specific Trends 20

A.5 Correlates of Cross-Hospital Heterogeneity 21

A.6 Selection on Slopes with Hospital-Specific Trends 22

A.7 Parameter Estimates, Dropping 2017 Data 23

A.8 Posterior Distributions, Dropping 2017 Data 24

A.9 Counterfactuals, Dropping 2017 Data 25

A.10 Counterfactuals Incorporating Choice Shifter 26

List of Figures

A.1 Observed and Predicted Bundle Prices 15
A.2 Gibbs Sample Posteriors 16

A. Imputing Bundle Prices

In each performance year, CMS assigns a bundle price to each participating hospital based on the hospital's quality and the case-mix of the episode. In period 3, CMS stopped reporting bundle prices for voluntary treatment hospitals that selected out. We therefore construct a predicted period 3 bundle price for all treatment hospitals based on a linear regression of period 3 bundle prices on period 2 bundle prices among all mandatory treatment and voluntary select-in hospitals where both prices are observed.

CMS generates a bundle price for each of the four case-mix groups: DRG 470 without hip fracture, DRG 470 with hip fracture, DRG 469 without hip fracture, and DRG 469 with hip fracture.[1] Let h index hospitals and c index case-mix groups. For each case mix group, we estimate the regression

$$b_{3,hc} = \beta_1 b_{2,hc} + \beta_{2,r(h)} + \epsilon_{hc}, \tag{11}$$

where $b_{3,hc}$ is the bundle price from period 3, $b_{2,hc}$ is the bundle price from period 2, and $\beta_{2,r(h)}$ is the region fixed effect. We control for region fixed effects because bundle prices are set based on both the hospital's historical spending and that of the region's. The adjusted R^2 from the regressions are 0.96, 0.96, 0.95 and 0.95 for the case mix groups.

To obtain one bundle price per hospital for our model estimation, we take a weighted average of the predicted values for each case-mix group for each hospital, weighting case-mix groups by the number of episodes for that case-mix group at that hospital in period 3. Appendix Figure A.1 shows a scatter plot of observed and predicted hospital-level bundle prices, for hospitals where we observe both.

[1] Technically, hospitals can receive a higher bundle price by obtaining a higher composite quality score (CQS). However, because of how CQS is constructed, we expect little change in quality over time. The largest component of CQS is the THA/TKA (total hip arthroplasty/total knee arthroplasty) complication measure, which is a three-year rolling average of standardized 90-day complication rate. The other main component is the HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) survey measure, which is a hospital-level measure of patient experience that covers all patients and not limited to LEJR. Consistent with this, Finkelstein et al. (2018) show that hospital quality scores are unaffected by bundled payments. We ignore this aspect of pricing for this exercise.

B. Variance Decomposition

In this section, we describe the variance decomposition exercise we use to quantify the power of patient and hospital characteristics in explaining the heterogeneity in levels and slopes across hospitals.

As we discuss in Section III, we use period 1 outcomes to proxy for hospital levels. Letting i denote episodes and h denote hospitals, we first residualize period 1 outcomes with episode-level regression of the form

$$outcome_{1i} = \beta_0 + \beta x_i + \gamma z_h + \epsilon_i, \tag{12}$$

where x_i are patient characteristics and z_h are hospital characteristics. We next take the residuals from these regressions ϵ_i and regress them on hospital fixed effects. We then compare the standard deviation of these estimated hospital fixed effects, based on the residualized outcomes, to the standard deviation of fixed effects of hospital fixed effects, where we do not first net out controls. Standard deviations are calculated weighting each hospital by the number of episodes in period 1.

To analyze sources of the heterogeneity in slopes, we conduct a similar exercise. Letting h index hospitals and s index strata, we estimate hospital-level regressions of the form

$$outcome_{h2} = \beta_0 + \beta_{1z} z_h BP_h + \beta_2 outcome_{h,2014} + \beta_3 outcome_{h,2013} + \delta_{s(h)} + \epsilon_{h2}, \tag{13}$$

where z_h are hospital characteristics (including, in one case, average characteristics of the hospital's patient population), and $\delta_{s(h)}$ are strata fixed effects. This is the analogue to Equation 2 except that we allow for the treatment effects to vary with hospital characteristics.

We then re-estimate Equation 2 using the estimated residual ϵ_{h2} from estimating the above equation as the outcome. As before, we then compare the standard deviation of the slopes based on the residualized outcomes, to the standard deviation of the slopes without residualizing, weighting each hospital by the number of episodes in period 1 as before.

Appendix Table A.5 shows the results from this analysis for the three main outcomes. The top row shows the standard deviations where we do not control for any hospital or patient characteristics, the subsequent rows show the impact of separately controlling for hospital characteristics, and the final rows show the effect of controlling for all the previously listed hospital characteristics and of controlling for the "kitchen

sink" of hospital and patient characteristics. The bottom line from this analysis is that observables explain only a modest share of the cross-hospital variation in levels and slopes. Controlling for all the hospital characteristics, along with strata and MSA fixed effects, explains a quarter of the variation in levels and about one-sixth of the variation in slopes. The kitchen sink specifications that additionally control for all available patient characteristics still leave at least half of the variation unexplained.

C. Description of the Gibbs Sampler

In this appendix, we describe the Gibbs sampler that we use to estimate the model. A key advantages of the Gibbs sampler is that it allows for data augmentation of latent variables (Tanner and Wong, 1987). In our context, we augment the hospital-specific level, slope, and choice shifter $\{\lambda_h, \omega_h, \nu_h\}_{h=1}^{H}$ as additional parameters of the model.

Let $H_C = \{h : BP_{h2} = 0\}$ be the set of control group hospitals and let $H_T = \{h : BP_{h2} = 1\}$ be the set of treatment group hospitals. Among the treatment group hospitals, let H_V be the set of hospitals that were given the decision whether to voluntarily select into the bundled payment program in period 3 and let $H_M := H_T \setminus H_V$ be the set of treatment group hospitals who were mandated to remain in the program.

We weight hospitals according to the average number of CJR episodes at that hospital so that our estimates are representative of the average episode in our sample. Let w_h denote the normalized number of episodes at hospital h, such that $\sum_{h=1}^{H} w_h = H$.

We can write the full model as follows:

$$
\begin{aligned}
\ln \lambda_{ht} &= \ln \lambda_h + \gamma_t + \epsilon_{ht}, \quad t = 1,2,3, \\
y_{ht} &= \lambda_{ht} - BP_{ht}\omega_h, \quad t = 1,2,3, \\
BP_{h3} = 1 &\iff b_{h3} - \lambda_h \exp(\gamma_3) + \frac{\omega_h}{2} + \nu_h > 0, \\
\begin{pmatrix} \ln \lambda_h \\ \ln \omega_h \end{pmatrix} &\sim N\left(\begin{pmatrix} x_h'\beta^{\lambda} \\ x_h'\beta^{\omega} \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_\omega \\ \rho\sigma_\lambda\sigma_\omega & \sigma_\omega^2 \end{pmatrix} \right), \\
\epsilon_{ht} &\sim N(0, \sigma_\epsilon^2), \; t = 1,2,3, \\
\nu_h &\sim N(x_h'\beta^{\nu}, \sigma_\nu^2),
\end{aligned}
$$

where x_h is a vector of hospital characteristics, including strata indicators, and we normalize $\gamma_2 = 0$. The selection equation in the third line only applies to hospitals in the voluntary treatment group H_V. Hospitals in the mandatory treatment group H_M have $BP_{h3} = 1$ automatically.

The data we observe are $\{y_{h1}, y_{h2}, y_{h3}, BP_{h1}, BP_{h2}, BP_{h3}, b_{h2}, b_{h3}, x_h\}_{h=1}^H$. The set of parameters (and pseudo parameters) we want to estimate is given by

$$\boldsymbol{\theta} = \left\{ \beta^\lambda, \beta^\omega, \beta^\nu, \Sigma = \begin{pmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_\omega \\ \rho\sigma_\lambda\sigma_\omega & \sigma_\omega^2 \end{pmatrix}, \sigma_\nu, \sigma_\epsilon, \gamma_1, \gamma_3, \boldsymbol{\lambda} = \{\lambda_h\}_{h=1}^H, \boldsymbol{\omega} = \{\omega_h\}_{h=1}^H, \boldsymbol{\nu} = \{\nu_h\}_{h\in H_V} \right\}.$$

For notational convenience, for any parameter $\delta \in \boldsymbol{\theta}$, we will use $\delta|\boldsymbol{\theta}^-$ to denote the parameter δ conditional on all other parameters in $\boldsymbol{\theta}$ and all data we can observe.

To facilitate the estimation, we make two transformations of the parameters. First, we denote $\tau_\epsilon = \sigma_\epsilon^{-2}$ as the precision of a normal distribution. Second, we re-write the participation equation as $\sigma_\nu^{-1}(b_{h3} - \lambda_h \exp(\gamma_3) + \frac{\omega_h}{2}) + \tilde{\nu}_h > 0$, where $\tilde{\nu}_h = \nu_h\sigma_\nu^{-1} \sim N(x_h'\beta^\nu\sigma_\nu^{-1}, 1)$. This transformation allows us to treat the participation equation as a Probit equation with linear parameters $\tilde{\beta}^\nu = \{\sigma_\nu^{-1}, \beta^\nu\sigma_\nu^{-1}\}$.

We now specify the prior distribution for each parameter in $\boldsymbol{\theta}$ and derive the expression for its posterior distribution conditional on all other parameters and the data we can observe. For $\beta \in \{\beta^\lambda, \beta^\omega, \tilde{\beta}^\nu\}$, the prior specifies that $\beta \sim N(\mu_0^\beta, \boldsymbol{\Sigma}_0^\beta)$. We use a conventional diffuse prior so that $\mu_0^\beta = 0, (\boldsymbol{\Sigma}_0^\beta)^{-1} = \mathbf{0}$. Let

$$\boldsymbol{X} = \begin{pmatrix} x_1' & 0 \\ ... & ... \\ x_H' & 0 \\ 0 & x_1' \\ ... & ... \\ 0 & x_H' \end{pmatrix}, \ \boldsymbol{y} = \begin{pmatrix} \ln\lambda_1 \\ ... \\ \ln\lambda_H \\ \ln\omega_1 \\ ... \\ \ln\omega_H \end{pmatrix}, \ \Omega = \text{Var}(\boldsymbol{y})\Big|\boldsymbol{\theta}^- = \Sigma \otimes \boldsymbol{I}_H.$$

Let $\boldsymbol{W}$ be a $2H-$dimensional diagonal matrix where the hth and $(H+h)$th elements are w_h. Then

$$(\beta^\lambda, \beta^\omega)|\boldsymbol{\theta}^- \sim N((\boldsymbol{X}'\Omega^{-1}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\Omega^{-1}\boldsymbol{W}\boldsymbol{y}, (\boldsymbol{X}'\Omega^{-1}\boldsymbol{W}\boldsymbol{X})^{-1}).$$

Similarly, let $\tilde{x}_h = (b_{h3} - \lambda_h \exp(\gamma_3) + \frac{\omega_h}{2}, x_h')'$, $z_h = \tilde{x}_h'\tilde{\beta}^\nu + \nu_h$, $\tilde{\boldsymbol{x}} = (\tilde{x}_h)_{h\in H_V}$, $\boldsymbol{z} = (z_h)_{h\in H_V}$,

and $\boldsymbol{w}$ be an $|H_V|-$dimensional diagonal matrix with elements $\{\frac{w_h}{\sum_{h\in H_V} w_h}\}_{h\in H_V}$ on the diagonal. Then

$$\tilde{\beta}^{\nu}|\boldsymbol{\theta}^{-} \sim N((\tilde{\boldsymbol{x}}'\boldsymbol{w}\tilde{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{x}}'\boldsymbol{w}\boldsymbol{z}, (\tilde{\boldsymbol{x}}'\boldsymbol{w}\tilde{\boldsymbol{x}})^{-1}).$$

We restrict to hospitals that were given the decision whether to voluntarily select into the bundled payments program, as nothing in the data tells us about ν_h for hospitals who did not face this decision.

For Σ, the prior specifies $\Sigma^{-1} \sim Wishart(\nu_0, S_0^{-1})$. We use a diffuse prior with $\nu_0 = 0, S_0 = 0$. Let $u_h = (\ln\lambda_h - x_h\beta^{\lambda}, \ln\omega_h - x_h\beta^{\omega})'$. Then

$$\Sigma^{-1}|\boldsymbol{\theta}^{-} \sim Wishart(H, (\sum_{h=1}^{H} w_h u_h u_h')^{-1}).$$

For τ_ϵ, the prior specifies $\tau_\epsilon \sim \Gamma(a_0, b_0)$. We use a diffuse prior with $a_0 = 0, b_0 = 0$. Then

$$\tau_\epsilon|\boldsymbol{\theta}^{-} \sim \Gamma(\frac{3}{2}H, b_\epsilon),$$

where

$$\begin{aligned} b_\epsilon =& \frac{1}{2}\sum_{h=1}^{H} w_h(\ln y_{h1} - \ln\lambda_h - \gamma_1)^2 + \frac{1}{2}\sum_{h=1}^{H} w_h(\ln(y_{h2} + BP_{h2}\omega_h) - \ln\lambda_h)^2 \\ &+ \frac{1}{2}\sum_{h=1}^{H} w_h(\ln(y_{h3} + BP_{h3}\omega_h) - \ln\lambda_h - \gamma_3)^2. \end{aligned}$$

For $\gamma \in \{\gamma_1, \gamma_3\}$, the prior specifies that $\gamma \sim N(\mu_0^{\gamma}, \sigma_0^{\gamma})$. We use a diffuse prior that $\mu_0^{\gamma} = 0, (\sigma_0^{\gamma})^{-1} = 0$. Then

$$\gamma_1|\boldsymbol{\theta}^{-} \sim N\left(\frac{1}{H}\sum_{h=1}^{H} w_h(\ln y_{h1} - \ln\lambda_h), \frac{1}{H}\tau_\epsilon^{-1}\right),$$

$$\gamma_3|\boldsymbol{\theta}^{-} \sim N\left(\frac{1}{H}\sum_{h=1}^{H} w_h(\ln(y_{h3} + BP_{h3}\omega_h) - \ln\lambda_h), \frac{1}{H}\tau_\epsilon^{-1}\right).$$

So far, once we condition on $\{\lambda_h\}_{h=1}^{H}$, $\{\omega_h\}_{h=1}^{H}$, and $\{\mu_h\}_{h=1}^{H}$, the model is pretty standard, with a system of three normally-distributed equations, and all parameters sampled from common posterior distributions, including normal, Gamma, and Wishart. The part of the Gibbs sampler that is less standard involves sampling from the conditional distribution of the augmented hospital-specific parameters: $\boldsymbol{\lambda} = \{\lambda_h\}_{h=1}^{H}$,

$\boldsymbol{\omega} = \{\omega_h\}_{h=1}^{H}$, and $\boldsymbol{\nu} = \{\nu_h\}_{h\in H_V}$. As each hospital is independent of the others, conditional on the other parameters, it does not depend on other hospitals' augmented parameters. Thus, all we need to describe is the conditional probability of λ_h, ω_h, and ν_h. Let $\phi(x,\mu,\sigma) = \exp(-1/2((x-\mu)/\sigma)^2)$ be proportional to the normal probability density function.

We start from ν_h. If we ignore the participation decision, ν_h follows the distribution $N(\tilde{x}_h'\tilde{\beta}^{\nu}, 1)$. When we incorporate the participation decision, for $h \in H_V$, ν_h follows a truncated normal distribution (truncated from below if $BP_{h3} = 1$ or from above if $BP_{h3} = 0$):

$$\Pr(\nu_h|\boldsymbol{\theta}^{-}) \propto \begin{cases} \phi(\nu_h, \tilde{x}_h'\tilde{\beta}^{\nu}, 1), & \text{if } BP_{h3} = \mathcal{I}(\nu_h > -(b_{h3} - \lambda_h \exp(\gamma_3) + \frac{\omega_h}{2})) \\ 0, & \text{if } BP_{h3} \neq \mathcal{I}(\nu_h > -(b_{h3} - \lambda_h \exp(\gamma_3) + \frac{\omega_h}{2})) \end{cases}.$$

Then λ_h. If we ignore the participation decision, the joint distribution of $(\ln\lambda_h, \ln y_{h1}, \ln(y_{h2} + BP_{h2}\omega_h), \ln(y_{h3} + BP_{h3}\omega_h))$ conditional on other parameters is

$$\begin{pmatrix} \ln\lambda_h \\ \ln y_{h1} \\ \ln(y_{h2} + BP_{h2}\omega_h) \\ \ln(y_{h3} + BP_{h3}\omega_h) \end{pmatrix}\Big|_{\boldsymbol{\theta}^{-}} \sim N\left(\begin{pmatrix} \mu_\lambda(\omega_h) \\ \mu_\lambda(\omega_h) + \gamma_1 \\ \mu_\lambda(\omega_h) \\ \mu_\lambda(\omega_h) + \gamma_3 \end{pmatrix}, \begin{pmatrix} \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 \\ \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 + \sigma_\epsilon^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 \\ \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 + \sigma_\epsilon^2 & \tilde{\sigma}_\lambda^2 \\ \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 + \sigma_\epsilon^2 \end{pmatrix}\right),$$

where

$$\mu_\lambda(\omega_h) = x_h'\beta^{\lambda} + \rho\frac{\sigma_\lambda}{\sigma_\omega}(\ln\omega_h - x_h'\beta^{\omega}),\ \tilde{\sigma}_\lambda^2 = (1-\rho^2)\sigma_\lambda^2.$$

The conditional marginal density of λ_h is proportional to $\phi(\ln\lambda_h, \tilde{\mu}_h^{\lambda}, \tilde{\sigma}_h^{\lambda})$, where

$$\tilde{\mu}_h^{\lambda} = \mu_\lambda(\omega_h) + \Sigma_{12}\Sigma_{22}^{-1}\begin{pmatrix} \ln y_{h1} - \mu_\lambda(\omega_h) - \gamma_1 \\ \ln(y_{h2} + BP_{h2}\omega_h) - \mu_\lambda(\omega_h) \\ \ln(y_{h3} + BP_{h3}\omega_h) - \mu_\lambda(\omega_h) - \gamma_3 \end{pmatrix}, \tilde{\sigma}_h^{\lambda} = \sqrt{\tilde{\sigma}_\lambda^2 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}},$$

$$\Sigma_{12} = \Sigma_{21}' = (\tilde{\sigma}_\lambda^2, \tilde{\sigma}_\lambda^2, \tilde{\sigma}_\lambda^2),\ \Sigma_{22} = \begin{pmatrix} \tilde{\sigma}_\lambda^2 + \sigma_\epsilon^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 \\ \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 + \sigma_\epsilon^2 & \tilde{\sigma}_\lambda^2 \\ \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 & \tilde{\sigma}_\lambda^2 + \sigma_\epsilon^2 \end{pmatrix}.$$

When we incorporate the participation decision, for control group hospitals ($h \in H_C$) and treatment group

hospitals mandated to remain in the program ($h \in H_M$), the density of λ_h remains the same. For treatment group hospitals, which were given the decision whether to voluntarily select into the bundled payments program ($h \in H_V$), $\ln \lambda_h$ follows a truncated normal distribution (truncated from above if $BP_{h3} = 1$ or from below if $BP_{h3} = 0$):

$$\Pr(\lambda_h|\boldsymbol{\theta}^-) \propto \begin{cases} \phi(\ln \lambda_h, \tilde{\mu}_h^\lambda, \tilde{\sigma}_h^\lambda), & h \in H_C \cup H_M \\ \phi(\ln \lambda_h, \tilde{\mu}_h^\lambda, \tilde{\sigma}_h^\lambda), & \text{if } BP_{h3} = \mathcal{I}(\lambda_h < \exp(-\gamma_3)(b_{h3} + \frac{\omega_h}{2} + \nu_h)), h \in H_V \\ 0, & \text{if } BP_{h3} \neq \mathcal{I}(\lambda_h < \exp(-\gamma_3)(b_{h3} + \frac{\omega_h}{2} + \nu_h)), h \in H_V \end{cases}.$$

Last ω_h. If we ignore the participation decision, the joint distribution of $(\ln \omega_h, \ln y_{h1}, \ln(y_{h2} + BP_{h2}\omega_h), \ln(y_{h3} + BP_{h3}\omega_h))$ conditional on other parameters is

$$\begin{pmatrix} \ln \omega_h \\ \ln y_{h1} \\ \ln(y_{h2} + BP_{h2}\omega_h) \\ \ln(y_{h3} + BP_{h3}\omega_h) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\omega(\lambda_h) \\ \ln \lambda_h + \gamma_1 \\ \ln \lambda_h \\ \ln \lambda_h + \gamma_3 \end{pmatrix}, \begin{pmatrix} \tilde{\sigma}_\omega^2 & 0 & 0 & 0 \\ 0 & \sigma_\epsilon^2 & 0 & 0 \\ 0 & 0 & \sigma_\epsilon^2 & 0 \\ 0 & 0 & 0 & \sigma_\epsilon^2 \end{pmatrix} \right),$$

where

$$\mu_\omega(\lambda_h) = x_h'\beta^\omega + \rho\frac{\sigma_\omega}{\sigma_\lambda}(\ln \lambda_h - x_h'\beta_\lambda),\ \tilde{\sigma}_\omega^2 = (1-\rho^2)\sigma_\omega^2.$$

Note that conditional on other parameters, y_{h1} and ω_h are independent. Using transformations of random variables, we can write down the conditional joint density of ω_h, y_{h2}, and y_{h3} as

$$\begin{aligned} f(\omega_h, y_{h2}, y_{h3}|\boldsymbol{\theta}^-) = & \frac{1}{\sqrt{1-\rho_{\lambda\omega}^2}\sigma_\omega\sqrt{2\pi}\omega_h} \exp\left(-\frac{(\ln \omega_h - \mu_\omega(\lambda_h))^2}{2\tilde{\sigma}_\omega^2}\right) \\ & \times \frac{1}{\sigma_\epsilon\sqrt{2\pi}(y_{h2} + BP_{h2}\omega_h)} \exp\left(-\frac{(\ln(y_{h2} + BP_{h2}\omega_h) - \ln \lambda_h)^2}{2\sigma_\epsilon^2}\right) \\ & \times \frac{1}{\sigma_\epsilon\sqrt{2\pi}(y_{h3} + BP_{h3}\omega_h)} \exp\left(-\frac{(\ln(y_{h3} + BP_{h3}\omega_h) - \ln \lambda_h - \gamma_3)^2}{2\sigma_\epsilon^2}\right). \end{aligned}$$

For control group hospitals ($h \in H_C$), it simplifies to $\ln \omega_h \sim N(\mu_\omega(\lambda_h), (1-\rho_{\lambda\omega}^2)\sigma_\omega^2)$. For treatment

hospitals ($h \in H_T$), the conditional marginal density of ω_h, denoted by $g(\cdot)$, is

$$g(\omega_h|\boldsymbol{\theta}^-) = \frac{f(\omega_h, y_{h2}, y_{h3}|\boldsymbol{\theta}^-)}{f(y_{h2}, y_{h3}|\boldsymbol{\theta}^-)} = \frac{f(\omega_h, y_{h2}, y_{h3}|\boldsymbol{\theta}^-)}{\int f(\omega_h, y_{h2}, y_{h3}|\boldsymbol{\theta}^-)\mathrm{d}\omega_h} \propto f(\omega_h, y_{h2}, y_{h3}|\boldsymbol{\theta}^-).$$

When we incorporate the participation decision, for control group hospitals ($h \in H_C$) and treatment group hospitals mandated to remain in the program ($h \in H_M$), the density of λ_h remains the same. For treatment group hospitals who were given the decision whether to voluntarily select into the BP program ($h \in H_V$), ω_h follows a truncated distribution (from below if $BP_{h3} = 1$ or from above if $BP_{h3} = 0$):

$$\Pr(\omega_h|\boldsymbol{\theta}) \propto \begin{cases} \phi(\ln\omega_h, \mu_\omega(\lambda_h), (1-\rho^2)\sigma_\omega^2), & \text{if } h \in H_C \\ g(\omega_h|\boldsymbol{\theta}^-), & \text{if } h \in H_M \\ g(\omega_h|\boldsymbol{\theta}^-), & \text{if } BP_{h3} = \mathcal{I}(\omega_h > 2(\lambda_h \exp(\gamma_3) - t_h - \nu_h)), h \in H_V \\ 0, & \text{if } BP_{h3} \neq \mathcal{I}(\omega_h > 2(\lambda_h \exp(\gamma_3) - t_h - \nu_h)), h \in H_V \end{cases}.$$

We restrict ω_h from being too large by truncating it from above at $0.71\lambda_h$ and from belong too small by truncating it from below at 1.[2] There is no closed-form expression for the density $g(\omega_h|\boldsymbol{\theta}^-)$, so we adopt the "invert CDF" sampling method (Devroye, 2006)[3] and use numerical integration to construct the CDF of ω_h for $h \notin H_C$.

With these posterior distributions, we can draw each parameter at a time, conditional on all other parameters and the data, and do this iteratively. We tested the algorithm on simulated data, where we know the true parameters, and it performs well. The posteriors cover the true parameters in most simulations.

We initialize the algorithm with the following initial values. In principle, we can start from any reasonable values and the algorithm will converge eventually. We make informed guesses to expedite the convergence. Let $W_C = \sum_{h \in H_C} w_h$ denote the sum of hospital weights in the control group H_C. Then

$$\gamma_1^1 = \frac{1}{W_C}\sum_{h \in H_C} w_h(\ln y_{h1} - \ln y_{h2}),\ \gamma_3^1 = \frac{1}{W_C}\sum_{h \in H_C} w_h(\ln y_{h3} - \ln y_{h2}),$$

[2] The upper bound of $0.71\lambda_h$ is imposed for economic reasons discussed in Section V.A. The lower bound of 1 is imposed for the computational reason of avoiding occasionally large negative values of $\log(\omega_h)$ that lead to unreasonable negative draws of σ_ω and create issues in the numerical integration.

[3] Let $F(x)$ be the cumulative distribution function of a random variable x. The "invert CDF" sampling method draws from this distribution by drawing r from a uniform distribution on [0,1] and computing $F^{-1}(r)$.

$$\sigma_\epsilon^1 = \sqrt{\frac{1}{2W_C}\sum_{h\in H_C} w_h(\ln y_{h1} - \ln y_{h2} - \gamma_1^1)^2 + \frac{1}{2W_C}\sum_{h\in H_C} w_h(\ln y_{h3} - \ln y_{h2} - \gamma_3^1)^2},$$

$$\ln\lambda_h^1 = \ln y_{h1} - \gamma_1^1,\ \beta^{\lambda^1} = (X'X)^{-1}(X'\ln\lambda^1),\ \text{where } X = (x_h)_{h=1}^H, \ln\lambda^1 = (\ln\lambda_h^1)_{h=1}^H,$$

$$d_h = \exp(\ln y_{h1} - \gamma_1^1 + \sigma_\epsilon^{1^2}) - y_{h2},\ \beta^{\omega 1} = (X'X)^{-1}(X'\ln d),\ \text{where } X = (x_h)_{h\in H_T, d_h>0}, d = (d_h)_{h\in H_T, d_h>0},$$

$$\sigma_\lambda^1 = \sqrt{\frac{1}{H}\sum_{h=1}^H w_h(\ln y_{h1} - \frac{1}{H}\sum_{h=1}^H w_h \ln y_{h1})^2 - \sigma_\epsilon^{1^2}},$$

$$\sigma_\omega^1 = \sqrt{\frac{1}{\sum_{h\in H_T, d_h>0} w_h}\sum_{h\in H_T, d_h>0} w_h(\ln d_h - \frac{1}{\sum_{h\in H_T, d_h>0} w_h}\sum_{h\in H_T, d_h>0} w_h d_h)^2},$$

$$\rho^1 = Corr(\ln\lambda_h^1, \ln d_h)_{h\in H_T, d_h>0},\ \ln\omega_h^1 \sim N(x_h'\beta^{\omega 1} + \rho^1\frac{\sigma_\omega^1}{\sigma_\lambda^1}(\ln\lambda_h^1 - x_h'\beta^{\lambda^1}), (1-\rho^{1^2})\sigma_\omega^{1^2}),$$

$$\beta^{\nu 1} = 0,\ \sigma_\nu^1 = 5,000,\ \nu_h^1 \sim N(x_h'\beta^{\nu 1}, \sigma_\nu^{1^2}) \text{ such that } BP_{h3} = \mathcal{I}(\nu_h^1 > -(b_{h3} - \lambda_h^1\exp(\gamma_3^1) + \frac{\omega_h^1}{2})).$$

We run the Gibbs sampler for 100,000 iterations. Each iteration takes a few seconds and in total it takes 5-6 days on a standard server. Appendix Figure A.2 shows the evolution of all parameters. It converges to the stable posterior distribution of the parameters fairly quickly, after several thousand iterations. We burn the first 10,000 iterations and construct our results using the remaining 90,000 iterations.

D. Mapping from Estimates to Results

In this appendix, we describe how we map from the Gibbs sampler to the estimates reported in Table IV, V, and Figure IV and the counterfactual results reported in Table VI and Figure V.

As we describe in Appendix C, we run the Gibbs sampler for 100,000 iterations. The posterior distributions of the parameters stabilize after several thousand iterations, so we burn the first 10,000 iterations and construct our results using the remaining 90,000 iterations.

Let $k = 1, 2, ..., K$ index iterations with $K = 90,000$. For each iteration, we have a random draw from the posterior distribution for each parameter, denoted θ^k. In Table IV, we report the posterior mean and

posterior standard deviation for each parameter calculated as

$$m_\theta = \frac{1}{K}\sum_{k=1}^{K}\theta^k, \quad s_\theta = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(\theta^k - m_\theta)^2}.$$

In Table V, we present summary statistics of the distribution of key economic objects across hospitals. Specifically, we compute the mean, standard deviation, and different percentiles across hospitals using draws for a given iteration k, weighting by the number of episodes at each hospital. Then we average across all iterations to construct the values we report in the table. For example, we construct

$$\mathbb{E}(\ln \lambda_h) = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{H}\sum_{h=1}^{H} w_h \ln \lambda_h^k\right)$$

where, as before, w_h is the normalized number of episodes at each hospital and $H = \sum_{h=1}^{H} w_h$ is the sum of these hospital weights.

We use a (slightly) modified approach to calculate statistics for λ_{h3}. Recall that λ_h is an augmented parameter in the Gibbs sampler, so each iteration produces random draws for it, while λ_{h3} is not. Therefore, for λ_{h3}, we first draw the stochastic component ϵ_{h3}^k for each hospital in each iteration from the normal distribution $N(0, \left(\sigma_\epsilon^k\right)^2)$. We next calculate $\lambda_{h3}^k = \lambda_h^k \exp(\gamma_3^k + \epsilon_{h3}^k)$. We then average over hospitals and iterations as before.

In Panel A of Table VI and Panel A of Appendix Table A.10 we present the counterfactual estimates, restricting our analysis to hospitals in the voluntary treatment group H_V where we can observe bundle prices. We use posterior draws from the Gibbs sampler to compute the following objects of interest:

(1) Bundled payment indicator: BP_{h3}^k. The first row of Panel A reports the counterfactual with no bundled payment program where all hospitals are paid under FFS, so $BP_{h3}^k = 0, \forall h, k$. The second row considers a counterfactual in which all hospitals were mandated to enroll in bundled payments in period 3, so $BP_{h3}^k = 1, \forall h, k$. The third row considers the voluntary selection scenario that actually took place, in which case BP_{h3}^k is the same as the observed participation decision. Given the nature of the Gibbs sampler, which conditions on the observed data, this BP_{h3}^k indicator is consistent with the draws in each iteration, so that $BP_{h3}^k = 1 \iff b_{h3}^k - \lambda_h^k \exp(\gamma_3^k) + \omega_h^k/2 + \nu_h^k > 0$.

(2) Government spending: $BP_{3h}^k b_{h3} + (1 - BP_{3h}^k)\lambda_{h3}^k$

(3) Relative social costs: $(1+0.15)(BP^k_{3h}b_{h3} + (1-BP^k_{3h})\lambda^k_{h3} - \lambda^k_{h3}) = (1+0.15)BP^k_{3h}(b_{h3} - \lambda^k_{h3})$

(4) Relative hospital profit (without choice shifter): $BP^k_{3h}(b_{h3} - \lambda^k_{h3} + \omega^k_h/2)$

(5) Relative social surplus (without choice shifter): (4) - (3) = $BP^k_{3h}(\omega^k_h/2 - 0.15(b_{h3} - \lambda^k_{h3}))$

(6) Relative hospital profit (with choice shifter): $BP^k_{3h}(b_{h3} - \lambda^k_{h3} + \omega^k_h/2 + \nu^k_h)$

(7) Relative social surplus (with choice shifter): (6) - (3) = $BP^k_{3h}(\omega^k_h/2 + \nu^k_h - 0.15(b_{h3} - \lambda^k_{h3}))$

As before, for a given object of interest we first take the weighted average across draws for a single iteration k and then we average over all iterations. The posterior values of λ^k_{h3} and ω^k_h are the same across counterfactuals.

E. Targeting

In order to explore the impact of better targeting in a systematic fashion, we approximate the observed bundle prices using a parametric distribution, and then change the parameters that govern the covariance of bundle prices with hospital costs. Specifically, we model bundle prices as lognormally distributed, such that they are correlated with λ_h but only correlated with ω_h via the correlation between λ_h and ω_h. That is, we assume b_{h3} and λ_{h3} follow the following joint lognormal distribution:

$$\begin{pmatrix} \ln b_{h3} \\ \ln \lambda_{h3} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_b \\ \mu_{\lambda_3} \end{pmatrix}, \begin{pmatrix} \sigma_b^2 & \rho_{\lambda_3 b}\sigma_b\sigma_{\lambda_3} \\ \rho_{\lambda_3 b}\sigma_b\sigma_{\lambda_3} & \sigma_{\lambda_3}^2 \end{pmatrix} \right).$$

We focus on voluntary treatment group hospitals (H_V) in period 3, as we do in the other counterfactuals. For these hospitals, the maximum likelihood estimators for the mean and standard deviation of the log bundled price distribution are given by:

$$\widehat{\mu_b} = \frac{1}{W_V}\sum_{h\in H_V} w_h \ln b_{h3}, \quad \widehat{\sigma_b} = \sqrt{\frac{1}{W_V}\sum_{h\in H_V} w_h \left(\ln b_{h3} - \widehat{\mu_b}\right)^2},$$

where w_h is, as before, the (normalized) number of episodes at hospital h and $W_V = \sum_{h\in H_V} w_h$ is the sum of these weights for the voluntary treatment group.

To recover $\rho_{\lambda_3 b}$, we estimate the covariance between $\ln \lambda_h$ and $\ln b_{h3}$. Since $Cov(\ln y_{h1}, \ln b_{h3}) = Cov(\ln \lambda_h - \gamma_1 + \epsilon_{h1}, \ln b_{h3}) = Cov(\ln \lambda_h, \ln b_{h3})$, due to the independence of ϵ_{h1}. It follows that the

maximum likelihood estimator of the covariance is

$$\widehat{Cov(\ln \lambda_h, \ln b_{h3})} = \frac{1}{W_V} \sum_{h \in H_V} w_h (\ln y_{h1} - \widehat{\mu_{\lambda_V}})(\ln b_{h3} - \widehat{\mu_b}),$$

where $\mu_{\lambda_V} = \frac{1}{W_V} \sum_{h \in H_V} w_h \ln y_{h1}$.

The counterfactuals reported in Panel B of Table VI and Panel B of Appendix Table A.10 are based on bundle prices drawn from alternative distributions. For a given counterfactual, defined by the parameters $\{\mu_b^k, \sigma_b^k, \rho_{\lambda_3 b}^k\}$, we draw bundle prices for each hospital h and each iteration k conditional on the simulated value of λ_{h3}^k :

$$\ln b_{h3}^k \sim N\left(\mu_b^k + \rho_{\lambda_3 b}^k \frac{\sigma_b^k}{s_{\lambda_3}^k}\left(\ln \lambda_{h3}^k - m_{\lambda_3}^k\right), (1 - {\rho_{\lambda_3 b}^k}^2)\sigma_b^{k2}\right),$$

where $m_{\lambda_3}^k$ and $s_{\lambda_3}^k$ are the weighted mean and standard deviation of $\ln \lambda_{h3}^k$ for that iteration:

$$m_{\lambda_3}^k = \frac{1}{W_V} \sum_{h \in H_V} w_h \ln \lambda_{h3}^k, \quad s_{\lambda_3}^k = \sqrt{\frac{1}{W_V} \sum_{h \in H_V} w_h (\ln \lambda_{h3}^k - m_{\lambda_3}^k)^2}.$$

Given the simulated b_{h3}^k, we then derive hospitals' participation decisions BP_{h3}^k using the selection equation and calculate the other quantities of interest following the approach described in Appendix D.

In the first row of Panel B, we consider the counterfactual of "perfect targeting," which sets bundle prices to equal to realized claims under FFS, except that we maintain the mean log bundle price at its observed level. This is equivalent to setting $\mu_b^k = \widehat{\mu_b}$, $\sigma_b^k = s_{\lambda_3}^k$, and $\rho_{\lambda_3 b}^k = 1$.

The second row examines the case of "feasible targeting," which sets the bundle price based on a hospital's underlying type and the time trend, but does not account for the stochastic component ϵ_{h3}^k. Specifically, we set $\mu_b^k = \widehat{\mu_b}$, $\sigma_b^k = s_\lambda^k$, and $\rho_{\lambda_3 b}^k = \frac{s_\lambda^k}{s_{\lambda_3}^k}$, where

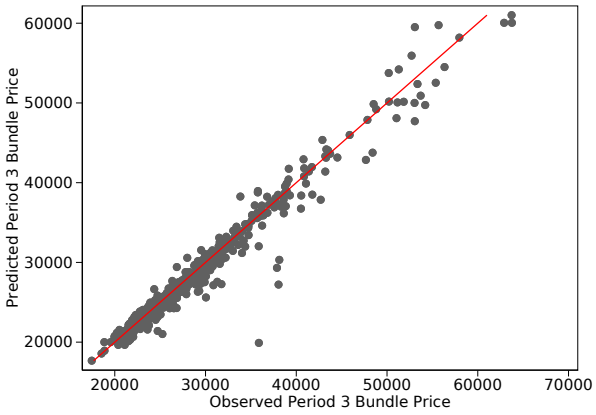
$$s_\lambda^k = \sqrt{\frac{1}{W_V} \sum_{h \in H_V} w_h \left(\ln \lambda_h^k - \frac{1}{W_V} \sum_{h \in H_V} w_h \ln \lambda_h^k\right)^2}.$$

The third row reports results under "observed targeting," which sets $\mu_b^k = \widehat{\mu_b}$, $\sigma_b^k = \widehat{\sigma_b}$, and $\rho_{\lambda_3 b}^k = \frac{\widehat{Cov(\ln \lambda_h, \ln b_{h3})}}{\widehat{\sigma_b} s_{\lambda_3}^k}$.

The fourth row considers a case in which bundle prices are uniform across hospitals and equal to the observed level of bundle prices. We set $\mu_b^k = \widehat{\mu_b}$ and $\sigma_b^k = 0$.

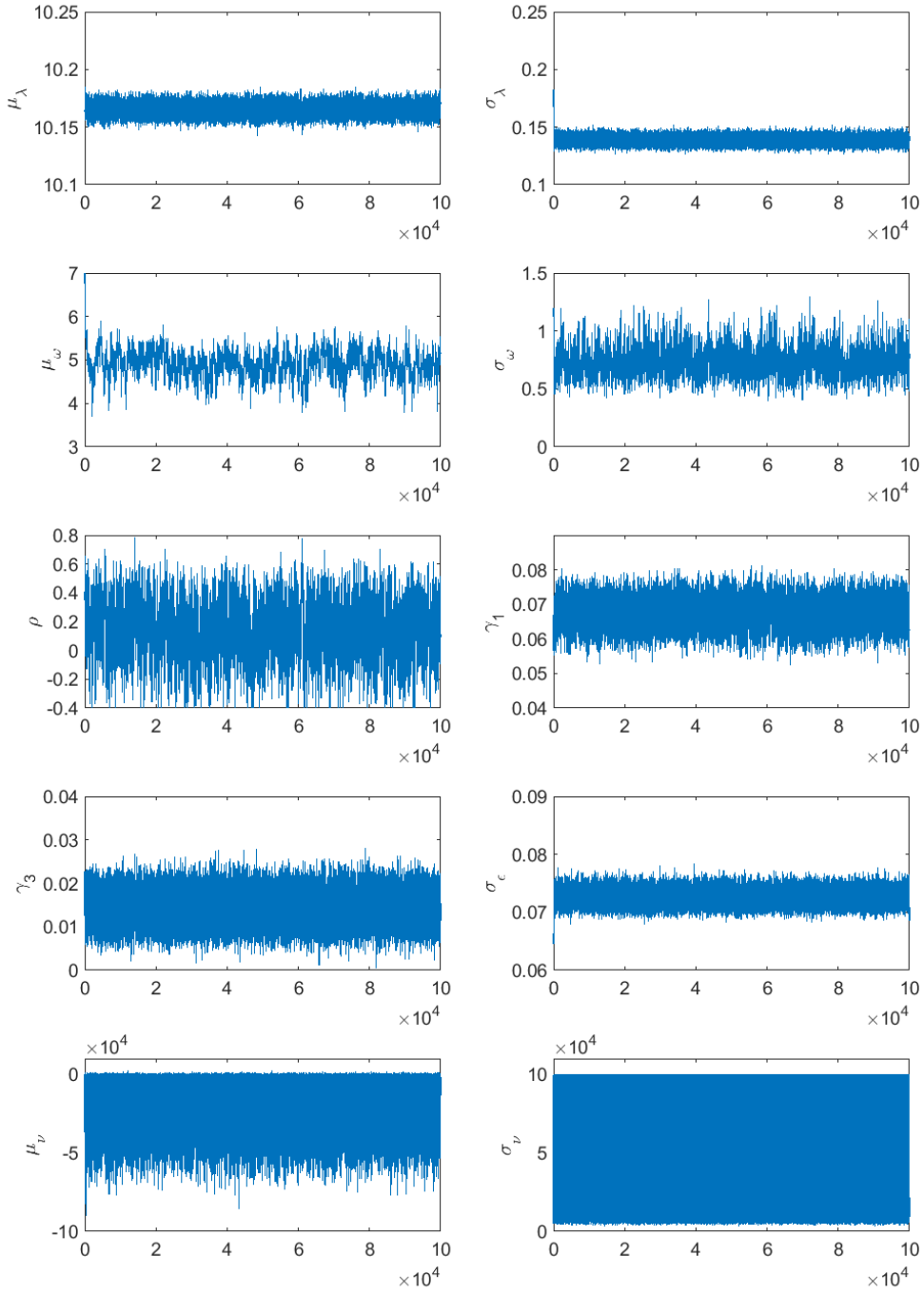
The last row considers "narrow bundling, no targeting," in which we remove hospital claims from the bundle and assume that we cannot target prices for other claims. Within our framework, this is equivalent to setting the bundle price as the sum of ex post, realized hospital claims and a predetermined, fixed payment for other costs: $b_h = f_h^{HOSP} + \mathbb{E}_h[f_h^{OTH}]$. We estimate the parameters for this bundle price using the empirical distributions of $\ln b_h = \ln(f_h^{HOSP} + \mathbb{E}_h[f_h^{OTH}])$ and $\ln \lambda_{h3}$. The resulting parameters used for the counterfactual are $\mu_b^k = \widehat{\mu_b}$, $\sigma_b^k = 0.09$, and $\rho_{\lambda_3 b}^k = 0.76$.

FIGURE A.1
Observed and Predicted Bundle Prices



*N*otes: Figure shows a scatterplot of the observed bundle price and the predicted bundle price, for hospitals for which we observe both. We obtained one bundle price per hospital by taking a weighted average across the four bundle prices based on the observed case-mix of episodes at each hospital in period 3.

FIGURE A.2
Gibbs Sample Posteriors



*N*otes: Figure shows posterior draws for each of the 100,000 iterations of the Gibbs sampler. See discussion in Appendix C for additional details.

TABLE A.1
Cream Skimming

	Select-in vs. Select-out Difference	SE	95% CI		P-value
Panel A: Average Elixhauser at hospital in period 3					
Mean among select-out (SD) 2.5 (0.7)					
Control for Period 1 Lag	-0.06	0.10	-0.25	0.14	0.56
Control for Period 2 Lag	-0.03	0.08	-0.19	0.13	0.75
Control for Period 1& 2 Lags	-0.02	0.08	-0.19	0.14	0.79
Panel B: Number of episodes at hospital in period 3 (Apr 1 - Sept 15)					
Mean among select-out (SD) 68.1 (67.2)					
Control for Period 1 Lag	-0.42	5.29	-11.19	10.35	0.94
Control for Period 2 Lag	-4.46	4.46	-13.55	4.63	0.32
Control for Period 1& 2 Lags	-4.66	4.28	-13.38	4.07	0.29
Panel C: Log number of episodes at hospital in period 3 (Apr 1 - Sept 15)					
Mean among select-out (SD) 3.6 (1.3)					
Control for Period 1 Lag	0.02	0.08	-0.13	0.18	0.77
Control for Period 2 Lag	-0.04	0.05	-0.14	0.06	0.40
Control for Period 1& 2 Lags	-0.04	0.05	-0.14	0.06	0.39

*N*otes: Table reports estimates from a hospital-level regression of the period 3 outcome an indicator for select-in, controlling for the lagged outcome from period 1 and/or period 2, and MSA fixed effects. The sample is restricted to hospitals in the voluntary treatment group that faced a select-in decision in period 3. Standard errors are clustered at the MSA level. The Elixhauser index of co-morbidities in Panel A is a measure of patient health, with a higher number indicating worse health.

TABLE A.2
Experimental Estimates During Mandatory Participation Period:
Alternative Specifications

	MSA-Level, Unweighted		MSA-Level, Weighted		Hospital-Level, Weighted	
	Control Mean (SD)	Average Treatment Effect (SE) *p*-value	Control Mean (SD)	Average Treatment Effect (SE) *p*-value	Control Mean (SD)	Average Treatment Effect (SE) *p*-value
Panel A: Healthcare Claims and Utilization						
Claims	25,294	-790	25,887	-588	25,899	-407
	(3,603)	(204)	(3,340)	(212)	(5,142)	(286)
		0.001		*0.006*		*0.157*
Claims for Index Admission	13,542	-169	13,750	-113	13,754	-47
	(2,389)	(89)	(2,080)	(90)	(2,670)	(133)
		0.06		*0.21*		*0.72*
Claims for Institutional PAC	4,119	-499	4,111	-334	4,117	-323
	(1,378)	(128)	(1,214)	(119)	(2,411)	(123)
		0.001		*0.006*		*0.009*
Claims for Home Health	1,800	-89	1,996	18	1,997	26
	(918)	(59)	(872)	(87)	(996)	(89)
		0.13		*0.84*		*0.77*
Other Claims	5,832	28	6,029	11	6,032	89
	(532)	(55)	(537)	(44)	(1049)	(86)
		0.61		*0.80*		*0.30*
Utilization Measures						
Number of Days in Index Admission	2.6	-0.1	2.6	0.0	2.6	0.0
	(0.4)	(0.04)	(0.3)	(0.04)	(0.6)	(0.04)
		0.22		*0.75*		*0.86*
Number of Days in Institutional PAC	7.7	-0.6	7.6	-0.3	7.6	-0.4
	(2.3)	(0.23)	(1.9)	(0.20)	(4.2)	(0.21)
		0.01		*0.13*		*0.08*
Discharge Destination						
Institutional Post Acute Care	0.313	-0.034	0.320	-0.021	0.320	-0.021
	(0.104)	(0.009)	(0.093)	(0.009)	(0.151)	(0.009)
		0.001		*0.015*		*0.020*
Home Health Agency	0.339	0.004	0.378	0.025	0.378	0.026
	(0.196)	(0.018)	(0.173)	(0.019)	(0.222)	(0.018)
		0.81		*0.19*		*0.16*
Home (w/o Home Health Agency)	0.329	0.042	0.289	0.013	0.289	0.005
	(0.232)	(0.018)	(0.212)	(0.018)	(0.265)	(0.020)
		0.02		*0.45*		*0.81*
Other	0.019	-0.004	0.013	-0.004	0.013	-0.004
	(0.032)	(0.002)	(0.023)	(0.002)	(0.030)	(0.002)
		0.05		*0.07*		*0.05*
Panel B: Quality Measures						
Complication Rate	0.011	0.001	0.011	0.001	0.011	0.002
	(0.005)	(0.001)	(0.004)	(0.000)	(0.010)	(0.001)
		0.26		*0.01*		*0.003*
ER Visit During Episode	0.198	0.003	0.194	0.002	0.194	0.002
	(0.027)	(0.003)	(0.021)	(0.002)	(0.047)	(0.003)
		0.40		*0.41*		*0.50*
90-day All Cause Readmission Rate	0.102	-0.001	0.103	-0.002	0.103	-0.001
	(0.015)	(0.002)	(0.014)	(0.002)	(0.031)	(0.002)
		0.73		*0.22*		*0.39*
Panel C: Admissions and Patient Composition						
LEJR Admissions (per 1,000 enrollees)	29.9	-0.8	27.3	-0.5	7.9	-0.5
	(15.8)	(0.5)	(13.6)	(0.3)	(12.4)	(0.2)
		0.10		*0.16*		*0.01*
CJR-eligible Admissions (per 1,000 enrollees)	23.6	0.1	21.1	0.1	6.7	-0.4
	(11.3)	(0.5)	(9.7)	(0.5)	(9.5)	(0.2)
		0.89		*0.87*		*0.02*
Elixhauser Comorbidity Score	2.4	0.0001	2.3	-0.004	2.3	-0.0001
	(0.3)	(0.029)	(0.2)	(0.027)	(0.4)	(0.027)
		0.998		*0.873*		*0.997*

*N*otes: Table compares regression estimates in Table I with estimates from two alternative specifications: an MSA-level specification, where observations are weighted based on number of episodes in the MSA in period 2, and a hospital-level specification where observations are weighted based on the number of episodes at the hospital in period 2. See Table I notes for more details.

TABLE A.3
Experimental Estimates: Effects Over Time

	Period	Control Mean	SD	Change with Bundled Payment	SE	P-Value
Panel A: All MSAs						
Total Episode Claims	2016 Q2	25,337	3,642	-591	232	0.02
	2016 Q3Q4	25,552	3,710	-739	223	0.01
	2017 Q1Q2	24,996	3,624	-684	224	0.01
	2017 Q3Q4	25,427	3,744	-900	253	0.01
Claims for Institutional PAC	2016 Q2	4,296	1,499	-443	156	0.01
	2016 Q3Q4	4,246	1,445	-406	140	0.01
	2017 Q1Q2	3,905	1,390	-502	139	0.01
	2017 Q3Q4	4,070	1,601	-635	161	0.01
Share Discharged to Institutional PAC	2016 Q2	0.34	0.12	-0.040	0.010	0.01
	2016 Q3Q4	0.32	0.11	-0.027	0.009	0.01
	2017 Q1Q2	0.30	0.10	-0.034	0.009	0.01
	2017 Q3Q4	0.29	0.10	-0.031	0.009	0.01
Panel B: Voluntary MSAs						
Total Episode Claims	2016 Q2	25,222	4,455	-573	353	0.11
	2016 Q3Q4	25,335	4,575	-629	344	0.08
	2017 Q1Q2	24,993	4,376	-719	335	0.04
	2017 Q3Q4	25,305	4,614	-1,001	362	0.01
Claims for Institutional PAC	2016 Q2	3,883	1,537	-328	209	0.13
	2016 Q3Q4	3,763	1,489	-255	190	0.19
	2017 Q1Q2	3,557	1,393	-448	174	0.02
	2017 Q3Q4	3,585	1,708	-442	213	0.05
Share Discharged to Institutional PAC	2016 Q2	0.29	0.10	-0.020	0.014	0.16
	2016 Q3Q4	0.27	0.08	-0.013	0.012	0.28
	2017 Q1Q2	0.25	0.08	-0.024	0.011	0.04
	2017 Q3Q4	0.24	0.08	-0.017	0.011	0.13
Panel C: Mandatory MSAs						
Total Episode Claims	2016 Q2	25,452	2,631	-483	347	0.17
	2016 Q3Q4	25,770	2,601	-671	313	0.04
	2017 Q1Q2	24,998	2,715	-616	327	0.07
	2017 Q3Q4	25,549	2,644	-713	391	0.08
	2018 Q1Q2	25,857	2,589	-959	406	0.03
	2018 Q3	26,214	2,783	-1,282	396	0.01
Claims for Institutional PAC	2016 Q2	4,709	1,352	-486	236	0.05
	2016 Q3Q4	4,730	1,235	-435	208	0.05
	2017 Q1Q2	4,254	1,310	-481	219	0.04
	2017 Q3Q4	4,555	1,334	-752	244	0.01
	2018 Q1Q2	4,591	1,165	-644	254	0.02
	2018 Q3	4,907	1,418	-901	279	0.01
Share Discharged to Institutional PAC	2016 Q2	0.40	0.11	-0.053	0.016	0.01
	2016 Q3Q4	0.37	0.11	-0.033	0.014	0.03
	2017 Q1Q2	0.34	0.10	-0.039	0.014	0.01
	2017 Q3Q4	0.34	0.09	-0.048	0.014	0.01
	2018 Q1Q2	0.35	0.07	-0.059	0.014	0.01
	2018 Q3	0.35	0.09	-0.054	0.017	0.01

*N*otes: Table replicates the analysis shown in Table I separately by time period for the three key outcome variables. Period 2 is 2016 and 2017. Period 3 is 2018. See Table I notes for more details.

TABLE A.4
Correlates of Levels and Slopes with Hospital-Specific Trends

	Panel A: Heterogeneity in Levels						Panel B: Heterogeneity in Slopes					
	Claims		Claims for Institutional PAC		Probability of Discharge to PAC		Claims		Claims for Institutional PAC		Probability of Discharge to PAC	
Mean (S.D.)	28,357	(5,998)	5,814	(3,021)	0.455	(0.187)	-1,001	(6,205)	-813	(3,253)	-0.033	(0.168)
Coefficient (S.E.) from Bivariate Regression												
Number of CJR Episodes	-5.31	(1.56)	-3.39	(0.69)	-0.0001	(0.0001)	1.04	(3.35)	1.71	(0.95)	0.00008	(0.00006)
Quality	-441	(41)	-197	(18)	-0.010	(0.001)	-26	(79)	36	(31)	-0.001	(0.002)
Number of Beds	3.90	(1.21)	0.62	(0.32)	0.0001	(0.00003)	1.45	(0.68)	0.36	(0.33)	0.000026	(0.00002)
Teaching	4,528	(599)	561	(258)	0.049	(0.021)	502	(689)	-387	(467)	-0.033	(0.023)
For-Profit	-3,030	(660)	-387	(304)	-0.064	(0.025)	-2,636	(1280)	-780	(425)	-0.056	(0.030)
Non-Profit	-219	(596)	369	(264)	0.008	(0.023)	-1,048	(507)	-596	(301)	-0.019	(0.024)

*N*otes: Table replicates Table II except that the hospital-specific slopes in Panel B are based on estimating an augmented version of Equation 2, which additionally controls for hospital-specific linear time trends. Specifically, we report estimates of $\beta_{4,h}$ from the regression $y_h = \beta_0 + \beta_{1,h} + \beta_{2,t} + \beta_{3,h} \times t + \beta_{4,h} \times BP_h + \epsilon_{h,t}$, where $\beta_{1,h}$ are hospital fixed effects, $\beta_{2,t}$ are calendar year fixed effects, $\beta_{3,h} \times t$ are hospital fixed effects interacted with a linear time trend, and $\beta_{4,h} \times BP_h$ are hospital fixed effects interacted with an indicator for treatment hospitals in period 2 (when bundled payment was in effect); we estimate this equation using data from 2010-2014, 2016 and 2017. In both panels, the coefficients on for-profit and non-profit are obtained from the same regression, where government-owned is the omitted category. All regressions are weighted by the number of episodes in period 2. Robust standard errors are shown in parentheses.

TABLE A.5
Correlates of Cross-Hospital Heterogeneity

	Panel A: Heterogeneity in Levels			Panel B: Heterogeneity in Slopes		
	Total Episode Claims	Institutional PAC Claims	Probability of Discharge to PAC	Total Episode Claimis	Institutional PAC Claims	Probability of Discharge to PAC
Unconditional S.D. of Hospital Fixed Effects	5,998	3,021	0.187	3,054	1,809	0.105
S.D. of Hospital Fixed Effects with additional controls:						
Number of CJR Episodes	5,843 (97.4%)	2,894 (95.8%)	0.183 (98.1%)	3,055 (100.0%)	1,799 (99.4%)	0.105 (99.9%)
Quality	5,604 (93.4%)	2,867 (94.9%)	0.180 (96.2%)	3,026 (99.1%)	1,778 (98.3%)	0.103 (98.0%)
Number of Beds	5,864 (97.8%)	3,017 (99.9%)	0.186 (99.7%)	3,048 (99.8%)	1,804 (99.7%)	0.104 (99.7%)
Teaching	5,751 (95.9%)	3,017 (99.9%)	0.186 (99.6%)	3,041 (99.6%)	1,802 (99.6%)	0.104 (99.4%)
Ownership (For-Profit, Non-Profit, Government)	5,896 (98.3%)	3,006 (99.5%)	0.185 (98.9%)	3,041 (99.6%)	1,809 (100.0%)	0.105 (99.9%)
Strata Fixed Effects	5,718 (95.3%)	2,882 (95.4%)	0.173 (92.8%)	3,054 (100.0%)	1,809 (100.0%)	0.105 (100.0%
MSA Fixed Effects	4,532 (75.6%)	2,557 (84.6%)	0.139 (74.5%)	2,665 (87.2%)	1,685 (93.2%)	0.092 (87.9%)
All of the above	3,537 (59.0%)	2,236 (74.0%)	0.125 (66.8%)	2,637 (86.4%)	1,638 (90.5%)	0.090 (85.9%)
All of the above, as well as all observed patient characteristics	2,830 (47.2%)	1,687 (55.8%)	0.107 (57.4%)	2,169 (71.0%)	1,236 (68.3%)	0.077 (73.1%)

*N*otes: Panels A and B report the standard deviation of hospital-specific levels and slopes, respectively. In parentheses, we also report the ratio of the conditional standard deviation to the unconditional standard deviation for that outcome. In Panel A, we first obtain residuals from an episode-level regression of the period 1 outcome (shown in the columns) on the row variable. We then regress the residuals on hospital fixed effects. Table reports the standard deviation of the hospital fixed effects, weighted by the number of episodes at each hospital in period 2. In Panel B, we first obtain residuals from a hospital-level regression of the period 2 outcome (shown in the columns) on an indicator for treatment, the row variable, lagged outcomes from 2013 and 2014, and strata fixed effects. We weight the regression by the number of episodes at each hospital in period 2. We then regress the residuals on the interaction between hospital fixed effects and treatment indicator, controlling for lagged outcomes from 2013 and 2014, and strata fixed effects (i.e. Equation 2). Table reports the standard deviation of the hospital-specific specific treatment effects (measured by the coefficients on the interaction terms), weighted by the number of episodes at each hospital in period 2. The set of patient characteristics include age, race, sex, disability, hip fracture, DRG, and number of Elixhauser comorbidities (in Panel B the hospital-level averages are being used).

TABLE A.6
Selection on Slopes with Hospital-Specific Trends

	Baseline			With Hospital-Specific Trends		
	Voluntary Select-In	Voluntary Select-Out	P-Value of Select-In vs. Select-Out Difference	Voluntary Select-In	Voluntary Select-Out	P-Value of Select-In vs. Select-Out Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Impact on episode claims	-791 (1,931)	-665 (2,826)	0.73	-534 (4,966)	-522 (5,487)	0.98
Impact on Institutional PAC claims	-518 (973)	-176 (1,474)	0.05	-388 (2,297)	-280 (2,793)	0.75
Impact on Share Discharged to Institutional PAC	-3.3% (7.8%)	-1.2% (9.2%)	0.122	-2.8% (14.4%)	-2.3% (15.0%)	0.81

*N*otes: The panel titled "Baseline" replicates Table III Panel B from the paper, and reports the average (and standard deviation) over different hospitals of $\beta_{1,h}$ from estimating Equation 2 by OLS. The panel titled "With Hospital-Specific Trends" reports the analogous results from estimating and augmented version of Equation 2, which additionally controls for hospital-specific linear time trends. Specifically, we report estimates of $\beta_{4,h}$ from the regression $y_h = \beta_0 + \beta_{1,h} + \beta_{2,t} + \beta_{3,h} \times t + \beta_{4,h} \times BP_h + \epsilon_{h,t}$, where $\beta_{1,h}$ are hospital fixed effects, $\beta_{2,t}$ are calendar year fixed effects, $\beta_{3,h} \times t$ are hospital fixed effects interacted with a linear time trend, and $\beta_{4,h} \times BP_h$ are hospital fixed effects interacted with an indicator for treatment hospitals in period 2 (when bundled payment was in effect); we estimate this equation using data from 2010-2014, 2016 and 2017. All estimates are weighted by the number of episodes in the hospital in period 2. The *p*-values in columns (3) and (6) are based on robust standard errors.

TABLE A.7
Parameter Estimates, Dropping 2017 Data

Panel A: Equation-Specific Parameters

	ln(λ) equation		ln(ω) equation		ν equation	
	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
Constant*	10.175	0.005	4.866	0.311	-8,635	12,943
ln(CJR Episodes)	-0.067	0.005	-0.560	0.174	5,774	9,359
ln(Beds)	0.049	0.006	0.439	0.283	1,078	5,396
Quality score	-0.170	0.005	4.605	0.272	48,028	11,762
Teaching	0.017	0.005	-0.033	0.272	-2,512	10,872
For-profit	-0.008	0.005	0.070	0.188	4,114	11,523
Government-owned	-0.002	0.004	-0.097	0.206	-712	4,849
Non-profit	omitted category		omitted category		omitted category	
Strata fixed effects	yes		yes		yes	
σ	0.139	0.003	0.696	0.113	27,044	28,666

Panel B: Additional Model Parameters

	Mean	Std. Err.
γ_1	0.060	0.004
γ_2	normalized to 0	
γ_3	0.008	0.003
σ_ε	0.076	0.001
ρ	0.081	0.234

*N*otes: Table reports posterior means and posterior standard deviations of the model parameters (parameter estimates) dropping 2017 data from period 2. See Table IV notes for more details.

* Constant is episode-weighted average of strata specific constants.

TABLE A.8
Posterior Distributions, Dropping 2017 Data

	E(x)	SD(x)	P5	P25	P50	P75
B.1. All hospitals						
$Ln(\lambda_h)$	10.17	0.19	9.90	10.05	10.15	10.28
λ_{h2}	26,668	5,844	19,359	22,663	25,579	29,601
λ_{h3}	26,989	6,021	19,392	23,047	25,880	29,671
$Ln(\omega_h)$	4.866	1.723	1.851	3.704	4.989	6.130
ω_h	460	1,096	7	44	156	479
B.2. Hospitals in the Voluntary Treatment Group Only						
λ_h	25,490	4,595	19,662	22,382	24,625	27,672
λ_{h3}	25,721	5,389	18,839	22,067	24,716	28,263
b_h	23,717	3,787	19,547	21,284	22,814	25,156
$b_h - \lambda_h - \gamma_3$	-1,987	2,752	-6,852	-3,302	-1,716	-344
ω_h	242	606	5	24	72	215
$(b_h - \lambda_h - \gamma_3) + \omega_h/2$	-1,866	2,738	-6,726	-3,157	-1,594	-244
v	-8,275	34,190	-65,597	-31,058	-7,741	14,962

*N*otes: Table presents summary statistics on the distribution of economic objects, dropping 2017 data from period 2. See Table V notes for more details.

TABLE A.9
Counterfactuals, Dropping 2017 Data

	Share Selecting In	Government Spending	Relative Social Costs	Ignoring Choice Shifter	
				Relative Hospital Profit	Relative Social Surplus
	(1)	(2)	(3)	(4)	(5)
Panel A: Mandatory vs. Voluntary					
Mandatory FFS (Benchmark)	0.00%	25,721	0	0	0
Mandatory Bundled Payment	100.00%	23,717	-2,304	-1,883	421
Voluntary Bundled Payment	39.20%	25,244	-548	-421	127
Panel B: Alternative Voluntary Regimes with Different Bundle Prices					
Perfect targeting	43.25%	25,721	0	61	61
Feasible targeting	42.72%	25,712	-11	53	64
Observed targeting	42.85%	25,823	118	165	47
No targeting	42.89%	26,218	572	558	-14
Narrow bundle, no targeting	42.82%	25,856	155	196	41

*N*otes: Table reports counterfactual estimates for the 259 hospitals in the voluntary treatment group, based on estimated that drop 2017 data from period 2. See Table VI notes for more details.

TABLE A.10
Counterfactuals Incorporating Choice Shifter

	Share Selecting In	Government Spending	Relative Social Costs	Incorporating Choice Shifter: Relative Hospital Profit	Incorporating Choice Shifter: Relative Social Surplus
	(1)	(2)	(3)	(4)	(5)
Panel A: Mandatory vs. Voluntary					
Mandatory FFS (Benchmark)	0.0%	25,517	0	0	0
Mandatory Bundled Payment	100.0%	23,659	-2,137	-9,455	-7,318
Voluntary Bundled Payment	38.8%	25,055	-532	8,393	8,925
Panel B: Alternative Voluntary Regimes with Different Bundle Prices					
Perfect targeting	38.7%	24,870	-745	8,225	8,970
Feasible targeting	38.5%	24,908	-700	8,232	8,932
Observed targeting	38.7%	25,018	-574	8,302	8,876
No targeting	39.1%	25,302	-248	8,526	8,773
Narrow bundle, no targeting	38.5%	25,045	-543	8,330	8,873

*N*otes: Table is identical to Table VI except that that columns (4) and (5) report hospital profits and social surplus, relative to the FFS counterfactual, under the assumption that ν_h is welfare relevant.