# MEASURING QUALITY EFFECTS IN EQUILIBRIUM

Seth Richards-Shubik
Mark S. Roberts
Julie M. Donohue

Measuring Quality Effects in Equilibrium
Seth Richards-Shubik, Mark S. Roberts, and Julie M. Donohue
NBER Working Paper No. 28817
May 2021
JEL No. C31,I11,L15

## ABSTRACT

Unlike demand studies in other industries, models of provider demand in health care often must omit a price or any other factor that equilibrates the market. Estimates of the consumer response to quality may consequently be attenuated, if the limited capacity of individual providers prevents some consumers from obtaining higher quality. We propose a tractable method to address this problem by adding a congestion effect to standard discrete-choice models. We show analytically how this improves forecasts of the consumer response to quality. We then apply this method to the market for heart surgery, and find that the attenuation bias in estimated quality effects can be important empirically.

Seth Richards-Shubik
Department of Economics
College of Business and Economics
Lehigh University
Rausch Business Center, Room 465
621 Taylor St
Bethlehem, PA 18015
and NBER
sethrs@lehigh.edu

Julie M. Donohue
Department of Health Policy & Management
Graduate School of Public Health
University of Pittsburgh
130 Desoto Street
Pittsburgh, PA 15261
jdonohue@pitt.edu

Mark S. Roberts
Department of Health Policy & Management
Graduate School of Public Health
University of Pittsburgh
130 Desoto Street
Pittsburgh, PA 15261
mroberts@pitt.edu

How demand responds to quality is a major concern in healthcare markets. The economics literature often points to low responsiveness to quality as a source of inefficiency, and a variety of public quality measures have been adopted in attempts to improve market allocations.[1] However, healthcare markets are unusual in that consumers often do not face the full prices of the products or services they acquire, and this has important implications for demand estimation. Without a price or other equilibrating factor in consumer utility, estimates of the effect of quality on demand may be biased toward zero and demand forecasts may similarly be attenuated, when standard discrete-choice models are used. This is particularly relevant for studies on the demand for individual providers, such as in the literature on the effects of public "report cards" for heart surgeons. Because of the usual institutional features and data limitations, several important studies in that literature have estimated discrete-choice models without equilibrating factors such as prices or wait times, and so their conclusions may be partially affected by this attenuation bias. As a consequence, the value of investments in provider quality may be greater than indicated by these and similar results, because the market response to quality improvements may be greater than their models would predict.

The intuition behind the potential attenuation bias is simple: if more patients want to see providers of higher quality, but the providers have limits on the number of patients they can treat—e.g., some form of capacity constraints—then higher quality providers may be harder to see. Hence there may be a *congestion effect* among patients, which does not appear in standard models. A market-clearing price or wait times would capture this effect, but the price may not exist and good data on wait times are often not available. Absent an equilibrating factor such as these in the demand model, the congestion effect could generate a downward bias in the estimate of the effect of a quality measure on demand (or the effect any desirable characteristic, for that matter). In theory, this bias could be present any time the supply from individual providers is not perfectly elastic, because the negative spillover from the congestion effect would be larger for higher quality providers. It is natural to expect the supply from individual producers of skilled services would not be perfectly elastic, because of fixed capacity or time constraints, or any other

---

[1] Surveys by Kolstad and Chernew (2009), Dranove (2011), and Skinner (2011) discuss the limited response to quality in healthcare markets, although recent work by Chandra et al. (2016) provides important evidence of a positive relationship between hospital quality and market shares. Public "report cards" have been issued on physicians, hospitals, and insurance plans since the 1990s (e.g., surgeon report cards in Pennsylvania and New York, Medicare's "Hospital Compare," and National Committee for Quality Assurance ratings).

source of increasing marginal costs. Hence this issue would be particularly relevant for studies about the effect of quality on demand for individual specialists.

Without an equilibrating factor, the standard models of provider choice cannot be interpreted structurally as representing consumer demand (unless supply is perfectly elastic).[2] Instead, these models fit the equilibrium probabilities that patients with certain characteristics receive services from physicians with certain characteristics. Fitting and forecasting these probabilities is nevertheless important, for example to measure the consumer responsiveness to quality and to predict the effect of quality improvements on patient outcomes. However, as we show, without accounting for the congestion effect among patients, forecasts of changes in equilibrium outcomes given changes in the quality or number of physicians in a market are also systematically biased—not just the parameter estimates themselves. For some intuition, consider a simple example where a multinomial logit is used for an entry simulation. Suppose the model includes an observable quality measure (e.g., the 30-day mortality rate of patients in the prior year) but no equilibrating factor. Then, if each provider were "cloned" (i.e., the number of alternatives were doubled, with identical characteristics) the model would predict that the market share for each original physician would be cut in half, and the clones would have the same market shares as their originals.[3] There would be no aggregate shift toward higher quality physicians, even though there is a large increase in supply relative to demand. This is clearly contrary to economic intuition. Our goal in this paper is to understand and address this problem.

We propose a relatively simple and tractable way to incorporate a congestion effect among consumers into discrete-choice models of provider choice, and thereby to obtain more accurate estimates of the effects of changes in provider quality on equilibrium outcomes. The approach is to include the number or proportion of patients treated by each provider in a market as an explanatory variable in the model. In a similar spirit as Ackerberg and Rysman (2005), we develop a stylized equilibrium model that would yield this specification—a queuing model where wait times are linear in the number of patients—but our emphasis is on the improvement in forecasts of changes in market shares given changes in quality. We argue that these counterfactual forecasts can be useful even if the structural interpretation is not maintained, and

---

[2] Consequently, for example, such models should not be used to compute consumer welfare.
[3] This is not simply a consequence of the IIA assumption; we find the same result in simulations with a mixed logit model in Section 3.3.

in the application we show that the improvements are empirically relevant. Naturally, this approach is not needed if complete data on wait times are available, but typical sources (e.g., claims or discharge records) do not show a patient's prospective wait times at every provider in the market.

Our econometric specification and estimation procedure is based on Bayer and Timmins (2007), which develops an equilibrium discrete-choice model for location choices. Their framework is similar to the demand model from Berry, Levinsohn, and Pakes (1995), but it replaces the price of each alternative with a peer effect that reflects various possible spillovers (positive or negative) among household or firms at the same location. In our case, this spillover is the negative congestion effect among patients seeking services from the same provider. As with prices in the Berry et al. (1995) framework, the peer effect (e.g., the market share of each alternative) is endogenous and requires an instrument. In models of provider choice, the distances between patients and providers in a market offer a natural instrument. While the distance between a patient and a given provider directly affects the propensity to see that provider, the distances between *other* patients and that provider only affect the propensity through that provider's patient volume. Our empirical application demonstrates the performance of this instrument. Additionally, we show that the two-step estimation procedure from Bayer and Timmins (2007) can be implemented using standard commands in econometric software such as Stata, making this approach readily available for applied research.

We then use this method to estimate quality effects in the market for heart surgery, a context where there appears to be meaningful variation in provider quality and where highly relevant quality measures (surgeon-level mortality rates) have been collected and publicized in several states.[4] There is a substantial literature on this topic, and several important studies have applied discrete-choice models to estimate the effect of publicly-reported quality measures on the demand for heart surgeons (e.g., Dranove and Sfekas 2008, Epstein 2010, Wang et al. 2011, Kolstad 2013, Gaynor et al. 2016).[5] The results in this literature are consistent with a possible congestion effect limiting patients' ability to see higher quality surgeons. In particular, Dranove and Sfekas (2008) and Wang et al. (2011) find that low quality scores have negative effects on

---

[4] To be clear, our paper is about the market response to quality, not the effect of public reporting *per se*.
[5] Gaynor et al. (2016) include a measure of wait times. Other important studies such as Dranove et al. (2003) do not involve demand estimation and forecasts, so they would not be impacted by the congestion effect considered here.

market share, but high quality scores do not have positive effects.  This suggests that capacity constraints may be limiting the ability of high-quality surgeons to increase their supply, but may not be affecting low-quality surgeons because they are not binding.  On the other hand, Chandra et al. (2016) estimate a stronger relationship between hospital quality and market share in cardiac care.  This would be consistent with a smaller congestion effect for hospitals, because capacity constraints are less likely to be binding for organizations than for individual physicians.

Using data from Pennsylvania, our analysis finds a large congestion effect for heart surgeons, and it shows that the attenuation bias in estimates of the market response to quality can be economically important.  In one regional market, the estimated quality parameter is biased downward by 40% to 80%, and a standard model without congestion underpredicts the aggregate consumer response to surgeon quality improvements by 25% or more.  This suggests that the returns to investments in provider quality may be greater than what is indicated by models that do not account for a potential congestion effect among patients.  In two other regional markets, we do not find a significant response to provider quality overall (although we do find a modest response by patients with more comorbidities).  In those two markets, interestingly, we see evidence that referrals for heart surgery are more likely to be kept within health systems, which may limit patients' ability to see higher quality surgeons (see, e.g., Baker, Bundorf, and Kessler 2016, and Gaynor, Propper, and Seiler 2016).

To be clear, we are not the first to recognize the potential importance of supply constraints in this context.  Mukamel, Weimer, and Mushlin (2007) and Epstein (2010) also note that typical models of demand for heart surgery do not account for supply constraints and may therefore produce biased estimates.  However they do not propose a method to address the issue. Ching, Hayashi, and Wang (2015) develop a novel demand model based on the Berry et al. (1995) framework that allows for fixed capacity constraints, and they apply it to nursing homes—another context where variation in quality is thought to be important.  For identification their model requires that there is a known market segment where the constraints do not bind and the prices are observed (in their case, private-pay patients), which may be restrictive in other applications.  In contexts that are substantially different from ours, Conlon and Mortimer (2013) and Geyer and Seig (2013) develop other methods to estimate discrete-choice models where consumers face unobserved limitations in supply (e.g., stock outs and wait lists).[6]  Last, although

---

[6] Carlton (1989) provides a summary of earlier work on non-price market clearing mechanisms.

not on demand estimation, Cutler, Huckman, and Kolstad (2010) consider the impact of the supply of heart surgeons on hospital entry into the heart surgery market, and they find evidence that the relatively inelastic supply of surgeons has affected entry.

In what follows, there are three main sections. The first develops our approach: it extends a standard discrete-choice model of provider choice to include a congestion effect, then presents an equilibrium model that would yield this specification, and finally describes the estimation procedure. The second section analyzes the potential attenuation bias in forecasts from standard models and provides a simple numerical example. The third section presents the empirical analysis, which uses data on heart surgeries in Pennsylvania in 2010 and 2011.

# 1. Model and Estimation

The market in our application consists of three sets of individuals: patients, referring physicians (cardiologists), and specialists (heart surgeons). The specialists are the providers being chosen. Their supply decisions are not modeled, but we assume that they have increasing marginal costs, so their supply is not perfectly elastic and hence a congestion effect may be present. Also, in the background we have in mind a matching process, where the equilibrium probability that a patient sees a particular specialist depends on the payoffs on both sides, and so we refer to outcomes in terms of matches and match probabilities.

To develop the model we start with a standard discrete-choice framework. The utility that patient $i$ would obtain from specialist $j$ is a function of a quality measure, $Q_j$, other specialist characteristics, $Z_j$, and observable attributes of the patient-specialist pair, $X_{ij}$, such as distance, along with an unobservable term, $\epsilon_{ij}$. We also incorporate information about the referring physician $r$ by including attributes of the pair of physicians, $X_{rj}$, such as the distance between their offices and whether they attended the same medical school.[7] Last, we allow for a common unobserved factor affecting the demand for specialist $j$, denoted $\xi_j$ (e.g., quality factors that are unobserved to the analyst). With the standard linear specification, utility is thus

$$U_{irj} = \beta_1 Q_j + Z_j' \beta_2 + X_{ij}' \beta_3 + X_{rj}' \beta_4 + \xi_j + \epsilon_{irj}. \qquad (1)$$

---

[7] Under an assumption that referring physicians are perfect agents for their patients, these pairwise attributes might capture the effects of informational frictions, for example.

This can be interpreted as the utility that the referring physician $r$ expects patient $i$ to receive from specialist $j$.

Specification (1) yields a probabilistic choice model, given some distribution for $\epsilon$. Assuming independent extreme value shocks yields the common multinomial logit model:[8]

$$\Pr(j|\{Q_k, Z_k, X_{ik}, X_{rk}, \xi_k\}_{k \in M}) = \frac{\exp(\beta_1 Q_j + Z_j'\beta_2 + X_{ij}'\beta_3 + X_{rj}'\beta_4 + \xi_j)}{\sum_k \exp(\beta_1 Q_k + Z_k'\beta_2 + X_{ik}'\beta_3 + X_{rk}'\beta_4 + \xi_k)}, \quad (2)$$

where the set $M$ collects the specialists available in patient $i$'s market. However, this model only captures demand-side factors. Notably absent, especially in comparison with similar models applied to other industries, is a price or any other equilibrating factor. The insurance payments to heart surgeons, for example, are not highly relevant for consumer choices. Nevertheless, as discussed in the introduction, it is important to account for the role of the supply side in determining equilibrium match probabilities. A fully developed structural model would approach this by specifying the utility that surgeons receive from treating patients and then applying an equilibrium concept to predict market outcomes. We do not model the supply side in this way, but rather we extend the probabilistic choice model (2) to include a congestion effect among patients. In the next section we show that this specification could arise from a simple equilibrium queuing model. However we believe the main value of this specification does not come from the possible structural interpretation. Rather, this specification can capture the effects of specialist supply decisions and capacity constraints on equilibrium match probabilities in a simple and flexible manner, and thereby yield more accurate forecasts of changes in market shares given changes in provider quality.

The congestion effect is incorporated by adding a measure of the specialist's patient volume, $Y_j$, to the model. This can be specified as the number or proportion of patients treated in a given time period, or with a more complex measure that accounts for heterogeneity in capacity (e.g., due to academic or administrative responsibilities). In the empirical analysis we use the proportion of patients treated by each surgeon (i.e., the market shares), because it is a common

---

[8] We also use a mixed logit model in the empirical analysis.

and easily interpreted measure. When estimation is done within a single market, as in our analysis, this is essentially equivalent to using the number of patients.[9]

Other factors that influence specialist utility and supply decisions may also be included in this extended model. Particularly, because models of physician utility often include compensation (e.g., Ellis and McGuire 1986, Chandra, Cutler, and Song 2011), it would be natural to include the insurance payment to the specialist. In our empirical application, the sample consists entirely of Medicare and Medicaid beneficiaries, so we can assume that payments are fairly uniform from each payer within each market.[10] Therefore we can reasonably account for the effect of different insurance payment amounts on the match probabilities by including indicator(s) for insurance type—here we use an indicator for Medicaid coverage, $I_i$. The main effect of this variable drops out, but its interactions with other variables are identifiable. Specifically, we include the interaction of the Medicaid indicator with the surgeon's volume, $I_i Y_j$, because we expect surgeons facing higher demand to be less likely to see patients whose insurance offers lower payments.[11]

Given these extensions to (2), the econometric specification is as follows:

$$\Pr(j|i,r,M) = \frac{\exp\big(\beta_1 Q_j + Z_j' \beta_2 + X_{ij}' \beta_3 + X_{rj}' \beta_4 + \gamma_1 Y_j + \gamma_2 I_i Y_j + \xi_j\big)}{\sum_k \exp\big(\beta_1 Q_k + Z_k' \beta_2 + X_{ik}' \beta_3 + X_{rk}' \beta_4 + \gamma_1 Y_k + \gamma_2 I_i Y_k + \xi_k\big)} \qquad (3)$$

(here "$i, r, M$" is shorthand for the characteristics of the patient, cardiologist, and all the surgeons in the market). To be clear, the terms inside the exponents do not necessarily have a structural interpretation as patient utility. Instead, this model is designed to measure and forecast the effect of quality on equilibrium match probabilities more accurately than models without an equilibrating factor. The next section briefly describes a queuing model that would in fact provide a structural interpretation for this specification, while Sections 2 and 3 show how including the congestion effect improves forecasts of the market response to quality changes.

---

[9] Within each market, the proportion equals the number of patients divided by a constant (i.e., the total number of patients in the market), so the parameter is simply rescaled.
[10] We have verified this with our available data on payment amounts.
[11] Medicaid generally pays about 70% as much as Medicare (http://kff.org/medicaid/state-indicator/medicaid-to-medicare-fee-index/).

## 1.1. Queuing model

We develop a simple equilibrium queuing model that yields specification (3), because it is useful to have a clear economic interpretation for the congestion effect (i.e., a negative peer effect) in a model of provider choice.[12] The model has similarities to a spatial equilibrium model with local spillovers, such as in Bayer and Timmins (2007), where individuals choose among a set of alternative locations, and equilibrium obtains when the population sizes that affect utility (e.g., $Y_j$) equal the choice probabilities (e.g., $\Pr(j|i, r, M)$).

In this stylized model, all the patients in the market simultaneously choose a provider, and then they wait to be treated based on how many other patients chose the same provider. Specifically, suppose that the patients choose their surgeons in period 0, and the surgeons then treat one patient per period (e.g., day) from $t = 1$ to $T$. Further suppose that patients are treated in random order, and the choice of surgeon is irrevocable. Hence if the number of patients who choose surgeon $j$ is $Y_j$, the expected wait time for surgeon $j$ is $Y_j/2$ (this assumes $Y_j < T$, for all $j$).

The utility function (1) can be extended to include disutility from the wait time, $W_{ij}$, which equals the period $t$ when the patient is treated. Then the utility that patient $i$ receives from specialist $j$ is

$$U_{irj} = \beta_1 Q_j + Z_j' \beta_2 + X_{ij}' \beta_3 + X_{rj}' \beta_4 + \alpha W_{ij} + \xi_j + \epsilon_{irj} \qquad (4)$$

with $\alpha < 0$. Because patients are treated in random order, they and their referring physicians consider the expected wait time, $Y_j/2$, when choosing a surgeon. A pure-strategy Nash equilibrium among patients then yields the match probabilities specified in (3), where $\gamma_1 = \alpha/2$ (absent the insurance term, $\gamma_2 I_i Y_j$).[13] Moreover this equilibrium is unique because $\alpha < 0$ (see Bayer and Timmins 2005).

While this model is highly stylized, it shows how wait times (rather than prices) can clear a market for services from individual physicians. Models where wait times or delivery lags serve

---

[12] Ackerberg and Rysman (2005) similarly use a stylized structural model to provide interpretation for their method, which also works by adding a variable to a standard discrete-choice model—in their case the number of alternatives.
[13] The insurance term would arise in a model where patients with different insurance types are given different positions in the queue. For example, with two insurance types where one is relatively uncommon and its holders are treated last, the parameter $\gamma_2$ would equal $\alpha$.

as equilibrating mechanisms have been well developed in the theoretical literature since at least Zarnowitz (1962).

## *1.2. Estimation with a congestion effect*

To estimate the model in equation (3), we use the two-step procedure developed by Bayer and Timmins (2007), which is similar to methods for estimating the Berry et al. (1995) demand model with micro data (e.g., Berry et al., 2004). First a multinomial logit model is estimated that includes fixed effects for each provider. Denoting these as $\delta_j$, the model is

$$\Pr(j|i,r,M) = \frac{\exp(\delta_j + X'_{ij}\beta_3 + X'_{rj}\beta_4 + \gamma_2 I_i Y_j)}{\sum_k \exp(\delta_k + X'_{ik}\beta_3 + X'_{rk}\beta_4 + \gamma_2 I_i Y_k)}. \qquad (5)$$

The remaining parameters are then estimated using the identity

$$\delta_j = \beta_1 Q_j + Z'_j \beta_2 + \gamma_1 Y_j + \xi_j. \qquad (6)$$

This defines a linear regression model for the provider fixed effects, which is estimated via two-stage least squares because the measure of surgeon volume ($Y_j$) is endogenous to the unobserved demand factors ($\xi_j$) by construction.[14]

Following Bayer and Timmins (2007), the instrument for $Y_j$ is a predicted market share, generated using a multinomial logit model with only the exogenous variables. In other words, we estimate a logit like (2) without the unobserved factor ($\xi_j$), and use it to generate a predicted market share for each provider (e.g., $\hat{Y}_j$). In our application, the crucial exogenous variation for the instrument comes from the distances between all the patients and providers in the market. While the distance between the reference patient $i$ and provider $j$ is included in the main model (3), the distances between all other patients and provider $j$ are excluded and therefore can provide

---

[14] Note that, if the additive specification of the index within the probability model is correct, the endogeneity of the interaction term $I_p Y_j$ is addressed by the surgeon fixed effects. Hence $\gamma_2$ can be consistently estimated in the first step where there are no instruments. This also follows Bayer and Timmins (2007) and Berry et al. (2004).

identifying variation.[15]  The key identifying assumption is that the locations of patients and providers are exogenous.  This is a standard assumption in models of demand for healthcare providers (see Gaynor and Town 2011), and it may be less problematic than the common approach in other industries of using characteristics of competing products or services as instruments for prices.

We also estimate an alternative specification of (5) as a mixed logit model, where the distance variables have random coefficients.[16]  For both specifications, the estimation of (6) uses weights that account for the variance of the estimated surgeon fixed effects.  These procedures are all readily available in Stata.  To generate predictions and counterfactuals from the estimated model, which requires computing new equilibrium market shares, we wrote a simple program in Stata that iterates on the vector of predicted market shares, $\hat{Y}$, until convergence is achieved.

## 2. Analysis of Forecast Bias

This section demonstrates how discrete-choice models without an equilibrating factor can produce systematically biased forecasts of the market response to quality changes.  We provide an analytic result to show one way the bias can arise, and we use a numerical example to illustrate this and a similar result.  We also discuss more generally when the bias can occur, in regards to the identification approaches (i.e., data and estimation) used to generate the forecasts.

Intuitively, the market shares of providers with substantial congestion (e.g., heart surgeons working at full capacity) should be less sensitive to changes in their competitive environment, because they have a surplus of potential customers.  We formalize this in the context of our model by showing that when a congestion effect is present, the (negative) substitution effect from quality improvements at other providers is smaller for providers with larger market shares, compared to the substitution effect predicted by a model without congestion.  This implies that the market shares of higher quality providers would remain larger than predicted by a standard model, because higher quality providers tend to have larger market shares initially.  As a consequence, forecasts from a standard model may underpredict the

---

[15] These distances between the other patients and the providers in the market represent additional information used by the congestion model over the standard model.  Hence the differences between the forecasts from the two models are driven by this additional information, not simply functional form.

[16] We also evaluated a version of (5) with random coefficients on all variables, but most of the estimated standard deviations for the random coefficients on the other variables were not statistically significant.

increase in the aggregate consumption of quality, given an increase in the quality of one or more individual providers in a market.

To focus this analysis, we use a simplified version of model (3) where the quality measure ($Q$) is the only exogenous variable, and we assume the market shares ($Y$) are known. Accordingly, suppose the data are generated by the following model:

$$Y_j = \frac{\exp(\beta Q_j + \gamma Y_j)}{\sum_k \exp(\beta Q_k + \gamma Y_k)}, \qquad (7)$$

where $\beta > 0$ and $\gamma < 0$, and $Y_j$ is the market share of provider $j$. However a standard model without a congestion effect is fit to the data:

$$Y_j = \frac{\exp(\hat{\beta} Q_j)}{\sum_k \exp(\hat{\beta} Q_k)}. \qquad (8)$$

We will show that the substitution effect from quality improvements at other providers ($\frac{dY_j}{dQ_k}$) is relatively smaller for providers with larger market shares, in the true "congestion model" (7) compared to the predictions from the "standard model" (8). Appendix A provides a similar analysis about the effect of entry, also showing that the substitution away from providers with larger market shares is relatively smaller than predicted by the standard model.

First, a useful closed-form expression for $\hat{\beta}$ can be obtained from a linear regression of the log-ratio of market shares, $\ln(Y_j/Y_1)$, on quality differences, $Q_j - Q_1$. (This would be a consistent estimator if the standard model were correct.) Denote the quality differences as $\Delta Q_j \equiv Q_j - Q_1$, and similarly define $\Delta Y_j \equiv Y_j - Y_1$. Then, in the true congestion model, we have $\ln(Y_j/Y_1) = \beta \Delta Q_j + \gamma \Delta Y_j$. The value of $\hat{\beta}$ recovered using the standard model is therefore

$$\hat{\beta} = \frac{\sum_j \Delta Q_j \cdot \ln(Y_j/Y_1)}{\sum_j (\Delta Q_j)^2} = \frac{\sum_j \Delta Q_j \cdot (\beta \Delta Q_j + \gamma \Delta Y_j)}{\sum_j (\Delta Q_j)^2} = \beta + \gamma \frac{\sum_j \Delta Q_j \Delta Y_j}{\sum_j (\Delta Q_j)^2}. \qquad (9)$$

This is the true value, $\beta$, plus $\gamma$ times the slope of the regression of market shares on quality differences. Hence $\hat{\beta}$ is less than $\beta$, because $\gamma$ is negative while the covariance of $Y$ and $Q$ is positive. In addition, this indicates that $\hat{\beta}$ is not invariant to changes in the distribution of provider quality, so it should not be interpreted as representing consumer preferences.

Rather than the value and interpretation of the quality parameter, however, our main interest is in the model's forecasts of changes in market shares given changes in the quality or number of providers in a market. To analyze this, we consider the marginal substitution effect of competitor quality on own market share. As shown in Appendix A, this is

$$\frac{dY_j}{dQ_k} = -Y_j Y_k \left( \beta + \gamma \frac{dY_k}{dQ_k} \right) \cdot (1 - \gamma Y_j)^{-1}. \qquad (10)$$

By comparison, in the standard model, we have the usual result for a multinomial logit:
$$\frac{dY_j}{dQ_k} = -Y_j Y_k \hat{\beta}.$$

Now we can examine the differences in the predicted substitution effect between models (7) and (8). Using the expression for $\hat{\beta}$, the prediction from the standard model is

$$\frac{dY_j}{dQ_k} = -Y_j Y_k \left( \beta + \gamma \frac{\sum_l \Delta Q_l \Delta Y_l}{\sum_l (\Delta Q_l)^2} \right), \qquad (11)$$

while the actual substitution effect is as expressed in (10). One difference is that the standard model uses the average marginal effect of own quality on market share (i.e., the regression slope, $\frac{\sum_l \Delta Q_l \Delta Y_l}{\sum_l (\Delta Q_l)^2}$) where the congestion model uses the individual marginal effect for provider $k$ ($\frac{dY_k}{dQ_k}$). But the key difference is that, given any provider $k$, the effect on provider $j$ is deflated by $(1 - \gamma Y_j)^{-1}$ in the congestion model. Hence the (negative) substitution effect is smaller in magnitude for providers with larger market shares (because $\gamma < 0$). Therefore, because market shares are larger when quality is higher (because $\beta > 0$), the market share losses are smaller for higher quality providers. This makes it possible for the congestion model to have larger market shares for higher quality providers, compared to the predictions from the standard model, in a counterfactual where quality increases for one provider (or more).

This result, and the result on entry in Appendix A, indicate that a model without an equilibrating factor *could* underpredict the increase in the aggregate consumption of quality, given an increase in the quality or number of providers in a market. The empirical analysis demonstrates that this is indeed the case in a salient application. Below we first provide a simple numerical example to illustrate these results.

Figure 1 shows a simulation of models (7) and (8) using parameter values $\beta = +1$ and $\gamma = -1$. The baseline scenario has three providers with quality levels 0, 1, and 2. Panel A plots the log-ratio of market shares ($\ln(Y_j/Y_1)$) against provider quality differences ($Q_j - Q_1$) from the true congestion model, along with the fitted values from the misspecified standard model. (The log-ratios and quality differences are useful quantities because the standard model assumes they have an invariant linear relationship: $\ln(Y_j/Y_1) = \hat{\beta}(Q_j - Q_1)$, as indicated in the plots.) Panel B then shows a scenario where the quality of provider two increases from 1 to 1.5. The standard model predicts the new log-ratios to be on the same line as in the baseline, because $\hat{\beta}$ is invariant, but the congestion model shows that the true log-ratios for the higher quality providers are above this line. Hence the aggregate consumption of quality would be higher than is predicted by the standard model. Panels C and D show the results of entry by a relatively low (panel C) or high (panel D) quality provider. These add a fourth provider with $Q_4 = 0.5$ or 1.5, respectively. In both cases the true log-ratios for the higher quality providers are above the predictions from the standard model, again showing that the aggregate consumption of quality would be higher than is predicted by the standard model.

We finish this section with a more general discussion of why the forecast bias occurs, which stems from how the model is recovered and then used to generate counterfactual forecasts. First, if data were available from many markets, where many different distributions of provider quality are observed, this issue would be irrelevant. A nonparametric model could be estimated, at least in concept, and the desired marginal effects and counterfactual forecasts could be computed simply by comparing outcomes between markets with the appropriate realizations of the distribution of provider quality. This is not the usual empirical setup, however. The analyst typically has data from one or a limited number of markets, and must rely on a parametric model to compute the quantities of interest. In this setup, quality parameters are recovered from observed differences in quality and market shares across providers *within* a market, which holds fixed the market-wide distribution of quality. But even a basic counterfactual forecast, such as

the marginal effect of increasing the quality of one provider while holding the others constant, typically contemplates a change in the aggregate supply of quality. Without an equilibrating factor, a standard model misses how this can affect all the alternatives. By contrast, a model with prices (e.g., Berry, Levinsohn, and Pakes 1995) would allow for changes in the prices of all alternatives when one alternative is improved. The congestion effect is like a price in that it can incorporate such spillovers across alternatives—in this case, how an increase in quality of one provider reduces congestion at another provider.

# 3. Empirical Application

## 3.1 Data and Measures

The data for our empirical application come from Pennsylvania, where highly relevant quality measures are available in heart surgeon "report cards" published every one or two years. The patient sample consists of Medicare fee-for-service beneficiaries with Part D coverage and Medicaid beneficiaries with any plan type (but not dually eligible for Medicare). We identify those beneficiaries who underwent coronary artery bypass graft (CABG) or heart valve repair surgeries in 2010 and 2011, using the same procedure list as the report card. The surgeon who performed the procedure is identified on the insurance claim. The referring physician is then inferred by searching through the patient's claims records for a specific diagnostic procedure: the cardiologist who provides a left heart catheterization to the patient most recently before the surgery (and within 180 days) is treated as the referring physician. This procedure is required before a CABG or valve repair surgery, in order to determine which vessels are to be repaired.

Using data on patient flows, specifically the zip codes of residence and of the hospital where surgery occurred, we construct three regional markets in Pennsylvania: Central, Southeastern, and Western.[17] These markets are largely self-contained: about 90% of the patients residing in these regions have heart surgery in their home regions. Our main results on the congestion effect come from the Central region, but we present results from all three regions both to show how they are comparable and to examine the lack of response to quality in the other regions.

---

[17] Parts of the state, such as the Scranton HRR, are omitted.

The characteristics of the surgeons, patients, and cardiologists in our sample are summarized in Table 1, separately by region. There are respectively 22, 71, and 49 heart surgeons in the Central, Southeastern, and Western regions. On average, these surgeons treated about 33 patients each from our sample of Medicare and Medicaid beneficiaries in 2010 and 2011. To find the total number of heart surgeries performed by each surgeon (across all payers), we also obtained hospital discharge records from the Pennsylvania Health Care Cost Containment Council (PHC4). Over the two-year period each of these surgeons performed an average of about 175 CABG and/or valve repair surgeries in total.[18]

The key variable of interest in our model is the quality measure, $Q_j$. For this we use a risk-adjusted mortality rate (RAMR) that is derived from the 30-day mortality rates provided in the report card, based on surgeries performed in 2008 and 2009. The summary dataset released with the report card lists the raw mortality rates for CABG and valve repair surgeries separately for each surgeon, along with the expected mortality rates for each surgeon derived from a risk-adjustment model.[19] We construct a combined RAMR for CABG and valve repair by first subtracting the expected mortality rates from the raw rates for each type of surgery, and then taking a weighted average of the two using the number of surgeries of each type. If the mortality rate for one type of surgery is not listed (because a surgeon has too few cases), we simply use the RAMR for the type that is reported. Of the 142 surgeons in our sample, 115 have CABG mortality rates and 91 have valve repair mortality rates listed in the report card, and 119 have at least one of the two. (We also include two indicator variables for surgeons without each of the mortality rates in the report card.) The average of the observed 30-day mortality rate is 3.15 per 100, while the average of the combined RAMR is 0.2 per 100.

The other key variable in our model is the measure of patient volume, $Y_j$. We use the market share, computed from all CABG and/or valve repair surgeries in the hospital discharge data. As noted previously (footnote 9), this is equivalent to using the number of patients because the models are estimated within market. More complex measures that account for an individual surgeon's availability or capacity are possible, for example by using information on academic or

---

[18] Most of the additional patients either have private insurance or are Medicare beneficiaries in managed care (40% of Medicare beneficiaries in PA) and/or without Part D coverage (about 30% nationally). See http://kff.org/medicare/state-indicator/enrollees-as-a-of-total-medicare-population/ and Donohue (2014). Also patients are excluded from our analytic sample if they cannot be matched to a cardiologist or if they have surgery outside their home region.

[19] Available at: http://www.phc4.org/reports/cabg/09/download.htm

administrative duties, or on historical working patterns, but we have not found these to improve the performance of the model in our current application.

There are 4,690 patients in the sample. We use the Elixhauser index (Elixhauser et al. 1998) as a measure of the overall health status of patients. This counts the number of certain comorbidities observed using diagnosis codes in health insurance claims over the 365 days prior to the surgery. Also, each patient is associated with a referring cardiologist as described earlier, and there are 479 unique cardiologists in the sample. The distances between patients and surgeons, and between cardiologists and surgeons, are measured using zip code centroids. We also collect data on the medical schools and current hospital affiliations of the cardiologists and surgeons in our sample, and include three indicators for whether a cardiologist and surgeon attended the same school, work at the same hospital, or work in the same health system.

## 3.2. Model Estimates

Key parameter estimates from the standard model (2) and the congestion model (3) are presented in Table 2. The table shows the coefficients on the RAMR and market share variables estimated for the Central region using both multinomial logit and mixed logit specifications. (The full sets of parameters for all three regions appear in supplemental tables B1-B3). We focus on this region for the congestion effect because in the other regions the predicted market share is too weak as an instrument (F-stat < 10, see column 8 of tables B2 and B3).[20]

In the Southeastern and Western regions, the coefficients on the RAMR are smaller and statistically insignificant (and often positive). The lack of any estimated response to quality is somewhat surprising, and it suggests that other features of these two markets may be affecting patient choices to a greater extent. Because we observe the referring physician, we are able to examine the possible role of health systems by using our indicators for whether the referring cardiologist and the surgeon work in the same system, and whether they work in the same hospital. Controlling for the latter (which may reflect many factors), we see that working in the same health system has a much stronger effect in the Southeastern and Western regions, compared to the Central region (tables B2 and B3, compared to B1). The estimated coefficients in those regions give average partial effects about twice as large as those in the Central region.

---

[20] We evaluated alternative measures patient volume ($Y_j$) that adjusted for the average number of days per month on which the surgeon performed an operation in the prior year, but these did not improve the power of the instrument.

For example, using the mixed logit, two-step estimates of the standard model (column 5 in each table), patients in the Southeastern and Western regions are respectively 1.8 and 1.6 percentage points more likely to see each surgeon who is in the same health system as their cardiologist, compared to 0.9 percentage points in the Central region. This apparently greater propensity to refer patients within a health system may have limited patient choices to some degree, thereby reducing the response to quality. Additionally, it is useful to note that the seven hospitals with heart surgery programs in the Central region were in each in a different health system (or independent) at the time, making it a relatively unconcentrated market in that regard.

Now turning to the congestion effect in the Central region, the coefficient on the surgeon's market share in either specification of the congestion model (table 2, columns 5 and 6) indicates a substantial congestion effect among patients. These coefficients can be interpreted roughly as semi-elasticities, so for example the estimate in column 5 implies that a one percentage-point increase in a surgeon's market share reduces an individual patient's probability of seeing that surgeon by 20% (in relative terms). The congestion effect is even greater for Medicaid beneficiaries, whose insurance offers surgeons a lower payment, which implies that Medicaid beneficiaries are less likely to see high-volume surgeons, all else equal.

Comparing the estimates in the first two columns with the last two columns, we see that the standard model has a substantial attenuation bias in the estimate of $\beta_1$ (the parameter on quality). In columns 1 and 2 the coefficient on the RAMR is five or six times smaller in magnitude than in columns 5 and 6. To examine whether this is a consequence of the different estimation procedures, columns 3 and 4 show the results from estimating the standard model using the two-step procedure (with OLS rather than 2SLS in the second step). The coefficient on the RAMR is still much smaller in magnitude than in columns 5 or 6, with a bias toward zero of over 40%.

Figure 2 plots the observed and predicted market shares for each surgeon in the Central region against their RAMR. This is useful to show the baseline distributions that will be manipulated in the simulations below, and to illustrate the fit of the estimated models. The predicted market shares from both the standard and congestion models are quite close to the observed shares for most surgeons (the predictions have the same horizontal position as the observed share for each surgeon because they use the same RAMR), and their correlations with the observed shares are above 0.85 for both models.

## 3.3 Effects of Quality Improvements

As discussed in the introduction and Section 1, our focus is not on the attenuation bias in the parameter estimates themselves, but rather in the predicted effects of quality changes on equilibrium match probabilities and the aggregate consumption of quality. Here we present two simulation exercises that empirically demonstrate the extent of the bias in such forecasts from the standard model. One simulates entry by duplicating surgeons in the market, the other simulates quality improvements by decreasing the historical RAMRs (i.e., from 2008-09) of the existing surgeons. Each exercise consists of a sequence of simulations, where first one surgeon is duplicated or improved, then two, then three, and so on.

The first exercise starts by duplicating the surgeon with the lowest RAMR, meaning that there is an additional alternative with the same characteristics as this surgeon. New match probabilities are computed, which is immediate for the standard model and involves converging to a new vector of market shares for the congestion model. Then going in sequence of increasing RAMR, the other surgeons are duplicated and the match probabilities are recomputed.[21] The second exercise uses the original set of surgeons, but decreases the historical RAMRs by one standard deviation (1.32) for surgeons in the middle range of quality, whose original RAMRs were within one standard deviation from 0.[22] In relation to Figure 2, this would appear as a leftward shift of the points that are currently between -1.32 and +1.32 on the x-axis (and with new predicted market shares for all surgeons).

Figure 3 shows the forecasted changes in the average quality consumed in these two exercises. Each panel plots the change in average quality (y-axis), measured as the average historical RAMR of the chosen surgeons, against the combined baseline market share of the surgeons who were duplicated or improved at that point in the sequence of simulations (x-axis).[23] For example in the first exercise, the first surgeon duplicated has a market share of 14% at baseline. In Panel A, the values of the plotted lines below that point on the x-axis show that the standard model predicts the RAMR of chosen surgeons decreases by 0.22 on average while the

---

[21] The five surgeons without an observed RAMR are duplicated last.

[22] We do not change the RAMR of surgeons who are already more than one standard deviation below zero because that may be unrealistic, and for symmetry we do not change the RAMR of surgeons who are more than one standard deviation above zero.

[23] The average historical RAMR of the chosen surgeons equals $\sum_i \sum_j \hat{p}_{ij} Q_j / N$, where $\hat{p}_{ij}$ is the predicted match probability and $N$ is the number of patients. The combined baseline market share is the same as the weighted proportion of surgeons who are altered, using their baseline market shares as weights. This is superior to the simple proportion because it reflects differences in surgeon market shares due to location and other factors.

congestion model predicts it decreases by 0.27. With more entry, the difference between these forecasts grows larger, ultimately exceeding 0.15. Also, as discussed in the introduction, when all surgeons are duplicated the standard model predicts the same average quality consumed as at baseline, which entirely fails to reflect the increase in the aggregate supply of quality.

Similar differences between the forecasts are seen in the in exercise that decreases the historical RAMRs (Panel B). For example, when four surgeons (with 24% baseline market share) have their RAMRs improved, the standard model predicts the RAMR of chosen surgeons decreases by 0.38 on average while the congestion model predicts it decreases by 0.47. This reflects the smaller market share losses for higher quality surgeons, discussed in Section 2.

The differences between the forecasts from the standard model and the congestion model in these exercises are roughly 10% of a standard deviation of the quality measure (i.e., 0.10-0.15 vs. 1.32). Although somewhat small relative to the variation across surgeons, these differences nevertheless have substantial economic value. For example, assuming that the current RAMRs are generally close to the previous RAMRs, a difference of 0.1 in the average RAMR of the chosen surgeons means a difference of 1 per 1000 in the 30-day mortality risk. Even with a conservative value of statistical life of $2 million, this is worth $2,000 per patient—or $2 million total in a market with 1,000 patients. The $2,000 difference in value per patient is also not trivial relative to the average cost of a CABG, which is $45,000 (Birkmeyer et al., 2010). Broadly, this suggest that the social value of investments in quality improvement for heart surgery may be larger than what is indicated by standard models without an equilibrating factor, because the market response to quality improvements may be greater than their forecasts would predict.

## 4. Conclusion

We have shown that discrete-choice models of demand are subject to attenuation bias if they omit an equilibrating factor, and we have proposed a straightforward way to address the bias by allowing for a congestion effect among consumers. Empirically we find that the congestion effect is substantial in the market for heart surgery, and the resulting biases in predicted responses to quality improvements can be economically important.

Our approach could be applicable to many healthcare services, because the relevant equilibrating factors are often difficult to define or observe, and because the supply from

individual providers typically is not perfectly elastic. The method requires an instrument that is sufficiently predictive of provider market shares. This can be a limitation in some markets, as we have seen, but the distances between patients and providers offer good candidate instruments and are usually available in research datasets. The method is easy to implement in econometric software such as Stata, and so our hope is that it may be useful for applied researchers studying the market response to quality in healthcare and possibly other industries.

# References

Ackerberg, Daniel A., and Marc Rysman. 2005. "Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects." *The RAND Journal of Economics*, 36: 771-788.

Baker, Laurence C., M. Kate Bundorf, and Daniel P. Kessler. 2016. "The Effect of Hospital/ Physician Integration on Hospital Choice." *Journal of Health Economics*, 50: 1-8.

Bayer, Patrick, and Christopher Timmins. 2005. "On the Equilibrium Properties of Locational Sorting Models." *Journal of Urban Economics*, 57: 462-477.

Bayer, Patrick, and Christopher Timmins. 2007. "Estimating Equilibrium Models of Sorting Across Locations." *The Economic Journal*, 117: 353-374.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63: 841-890.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy*, 112: 68-105.

Birkmeyer, John D., Cathryn Gust, Onur Baser, Justin B. Dimick, Jason M. Sutherland, and Jonathan S. Skinner. 2010. "Medicare Payments for Common Inpatient Procedures: Implications for Episode-Based Payment Bundling." *Health Services Research*, 45: 1783-95.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review*, 106: 2110-44.

Chandra, Amitabh, David Cutler, and Zirui Song. 2011. "Who Ordered That? The Economics of Treatment Choices in Medical Care." In *Handbook of Health Economics*, vol. 2: 397-432.

Ching, Andrew T., Fumiko Hayashi, and Hui Wang. 2015 "Quantifying the Impacts of Limited Supply: The Case of Nursing Homes." *International Economic Review*, 56: 1291-1322.

Conlon, Christopher T., and Julie Holland Mortimer. 2013 "Demand Estimation under Incomplete Product Availability." *American Economic Journal: Microeconomics*, 5: 1-30.

Cutler, David M., Robert S. Huckman, and Jonathan T. Kolstad. 2010. "Input constraints and the efficiency of entry: Lessons from cardiac surgery." *American Economic Journal: Economic Policy*, 2: 51-76.

Donohue, Julie M. 2014. "Impact and Evolution of Medicare Part D." *New England Journal of Medicine*, 371: 693-5.

Carlton, Dennis W. 1989. "The theory and the facts of how markets clear: Is industrial organization valuable for understanding macroeconomics?" In *Handbook of Industrial Organization*, vol. 1: 909-946.

Dranove, David. 2011. "Health care markets, regulators, and certifiers." In *Handbook of Health Economics*, vol. 2: 639-690.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is more information better? The effects of "report cards" on health care providers." *Journal of Political Economy*, 111: 555-588.

Dranove, David, and Andrew Sfekas. 2008. "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards." *Journal of Health Economics*, 27: 1201-07.

Elixhauser, Anne, Claudia Steiner, D. Robert Harris, and Rosanna M. Coffey. 1998. "Comorbidity measures for use with administrative data." *Medical Care*, 36: 8-27.

Ellis, Randall P., and Thomas G. McGuire. 1986. "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply." *Journal of Health Economics*, 5: 129-151.

Epstein, Andrew J. 2010. "Effects of Report Cards on Referral Patterns to Cardiac Surgeons." *Journal of Health Economics*, 29: 717-731.

Gaynor, Martin, Carol Propper, and Stephan Seiler. 2016. "Free to choose? Reform, choice, and consideration sets in the English National Health Service." *American Economic Review*, 106: 3521-57.

Gaynor, Martin, and Robert J. Town. 2011. "Competition in Health Care Markets." In *Handbook of Health Economics*, vol. 2: 499-637.

Geyer, Judy, and Holger Seig. 2013 "Estimating a Model of Excess Demand for Public Housing." *Quantitative Economics*, 4: 483-513.

Kolstad, Jonathan T. 2013 "Information and quality when motivation is intrinsic: Evidence from surgeon report cards." *American Economic Review*, 103: 2875-2910.

Kolstad, Jonathan T., and Michael E. Chernew. 2009. "Quality and Consumer Decision Making in the Market for Health Insurance and Health Care Services." *Medical Care Research and Review*, 66: 28S-52S.

Mukamel, Dana B., David L. Weimer, and Alvin I. Mushlin. 2007. "Interpreting Market Share Changes as Evidence for Effectiveness of Quality Report Cards." *Medical Care*, 45: 1227-32.

Skinner, Jonathan. 2011. "Causes and consequences of regional variations in health care." In *Handbook of Health Economics*, vol. 2: 45-93.

Wang, Justin, Jason Hockenberry, Shin-Yi Chou, and Muzhe Yang. 2011. "Do bad report cards have consequences? Impacts of publicly reported provider quality information on the CABG market in Pennsylvania." *Journal of Health Economics*, 30: 392-407.

Zarnowitz, Victor. 1962. "Unfilled Orders, Price Changes, and Business Fluctuations." *The Review of Economics and Statistics* 44: 367-94.
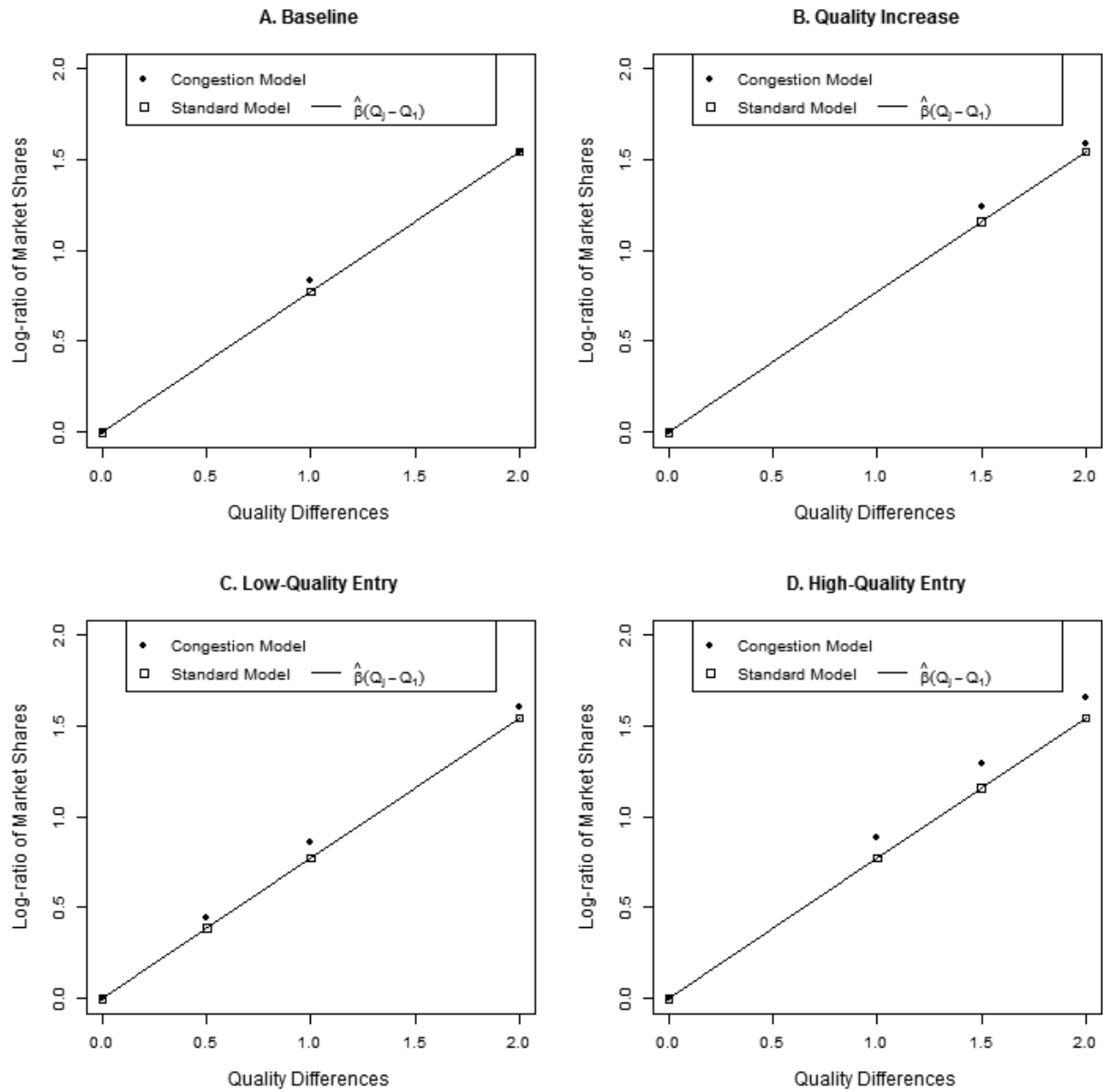
**Figure 1. Numerical Example of Differences in Forecasts**

*Notes:* Figure shows the log-ratio of market shares, $\ln(S_j/S_1)$, from the congestion model (filled circles) and the standard model (open squares) in an example with three providers at baseline. Predictions from the standard model fall on the line $\hat{\beta}(Q_j - Q_1)$, where $\hat{\beta}$ is recovered from the baseline market shares.
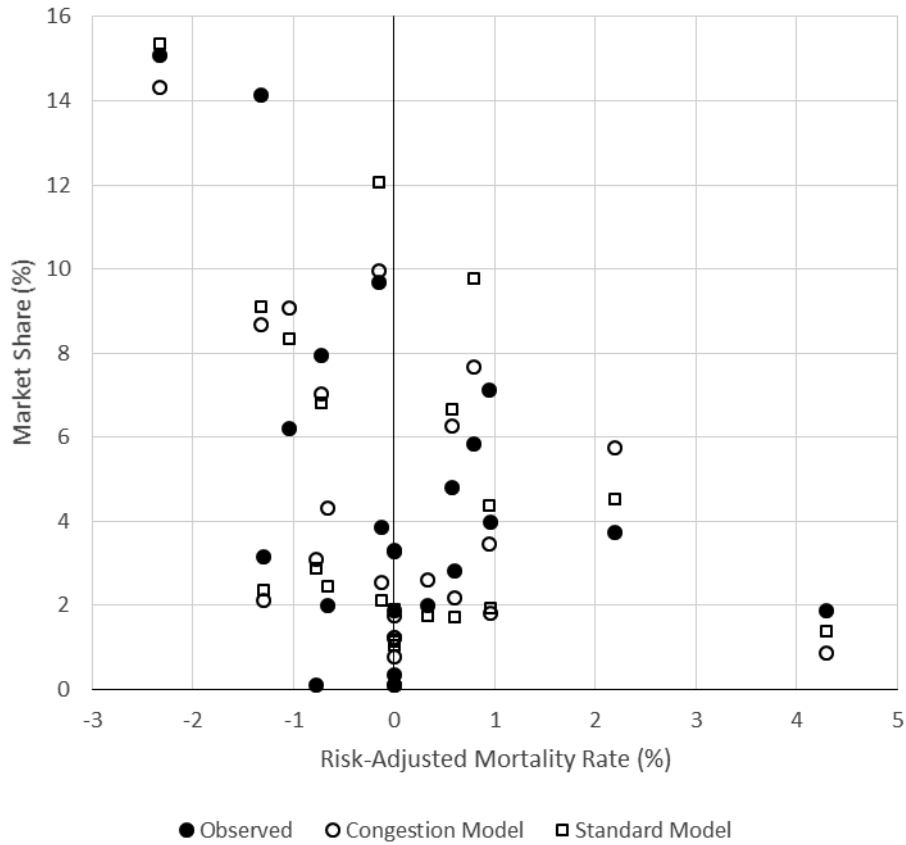
**Figure 2. Surgeon Market Shares and Risk-Adjusted Mortality Rates, Central PA Market**

*Notes:* Observed market shares are computed from the analytic sample of Medicare and Medicaid beneficiaries. Predicted market shares are generated from the mixed logit specifications of the models. The prediction from the congestion model is in equilibrium (i.e., the predicted vector of shares has converged). Five surgeons without an observed RAMR are plotted at a RAMR of 0.

*Panel A. Entry simulation by duplicating surgeons*
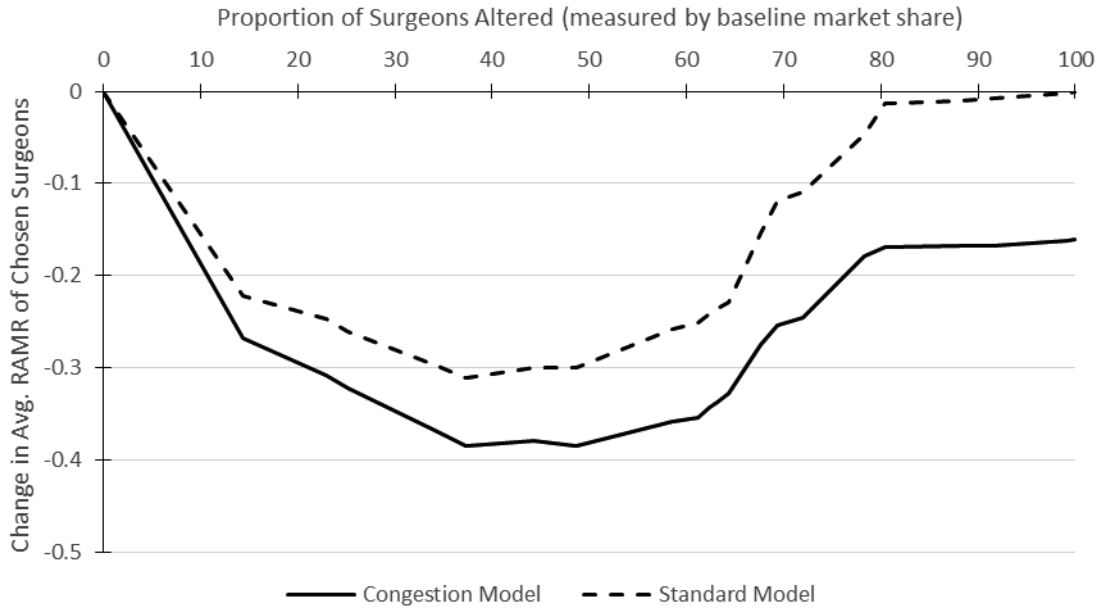


*Panel B. Quality improvement simulation by decreasing historical RAMRs*
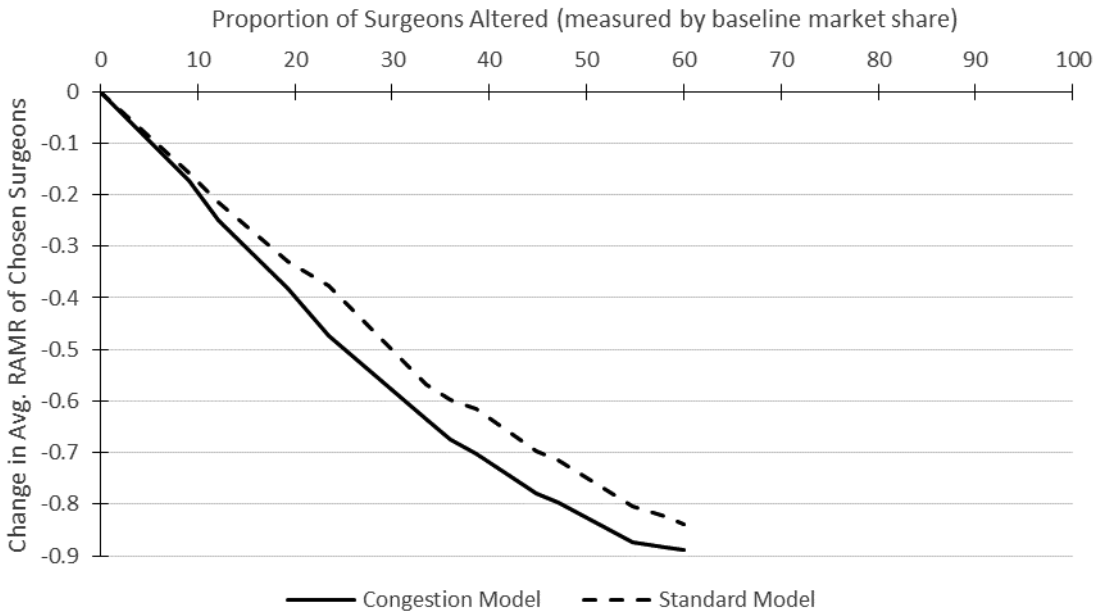


**Figure 3. Forecasts of Quality Improvements**

*Notes:* This figure plots differences in the average quality consumed between simulated markets and the baseline market, in a sequence of simulations. The average quality consumed in each scenario equals $\sum_i \sum_j \hat{p}_{ij} Q_j / N$, where $\hat{p}_{ij}$ is the predicted match probability, $Q_j$ is the 2008-09 RAMR, and $N$ is the number of patients. The predicted quality differences (y-axis) are plotted against proportion of surgeons who were altered in each simulation, weighted by their baseline market shares (x-axis).

**Table 1. Sample Descriptive Statistics**

|  | *Region* | | |
|  | Central PA | Southeastern PA | Western PA |
| --- | --- | --- | --- |
| Surgeons *(N)* | 22 | 71 | 49 |
| Surgeries performed (2010-11): | | | |
|    Analytic sample (Medicare/Medicaid) | 38.9 ± 35.6 | 33.5 ± 25.3 | 29.6 ± 18.3 |
|    Discharge data (all payers) | 219 ± 162 | 162 ± 138 | 182 ± 109 |
| Market share (pct., discharge data) | 4.5 ± 3.4 | 1.4 ± 1.2 | 2.0 ± 1.2 |
| Risk-adj. mortality rate (pct., 2008-09) | 0.1 ± 1.3 | 0.2 ± 2.4 | 0.2 ± 2.3 |
| | | | |
| Cardiologists *(N)* | 74 | 217 | 188 |
| Number of patients in sample | 11.6 ± 10.4 | 11.0 ± 10.1 | 7.7 ± 7.7 |
| Number of surgeons referred to | 2.9 ± 1.8 | 3.3 ± 1.9 | 2.6 ± 1.5 |
| | | | |
| Patients *(N)* | 856 | 2382 | 1452 |
| Elixhauser comorbidity index | 5.2 ± 2.9 | 5.5 ± 2.9 | 5.5 ± 3.0 |
| Age | 69.7 ± 11.7 | 70.0 ± 11.4 | 67.3 ± 12.3 |
| Female (pct.) | 43.9 | 41.3 | 43.9 |
| | | | |
| Patient/Cardiologist – Surgeon Pairs *(N)** | 17,601 | 143,613 | 67,039 |
| Patient-surgeon distance | 55.9 ± 32.7 | 48.1 ± 28.7 | 83.3 ± 50.7 |
| Cardiologist-surgeon distance | 49.6 ± 33.4 | 45.9 ± 27.2 | 77.8 ± 55.3 |
| Cardiologist and surgeon: | | | |
|    Attended same medical school (pct.) | 1.5 | 1.9 | 1.3 |
|    Admit to same hospital (pct.) | 11.2 | 7.9 | 11.2 |
|    Work in same hospital system (pct.) | 15.4 | 13.3 | 23.3 |

*Notes:* Data from Medicare and Medicaid insurance claims, hospital discharge records, Healthcare Organization Services from IMS Health, and the AMA Physician Masterfile. Standard deviations are shown after ± where appropriate.

* Surgeons without any surgeries in the quarter when a patient has surgery are excluded.

**Table 2. Key Parameter Estimates, Central PA Market**

| Variables | Standard Model | | | | Congestion Model | |
| --- | --- | --- | --- | --- | --- | --- |
| | ML / MSL Estimation | | 2-Step Estimation | | 2-Step Estimation | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Risk-adjusted mort. rate | -0.110 | -0.136 | -0.379 | -0.396 | -0.661 | -0.716 |
| | (0.061) | (0.063) | (0.263) | (0.261) | (0.303) | (0.283) |
| Elixhauser x RAMR | -0.037 | -0.039 | -0.038 | -0.039 | -0.035 | -0.037 |
| | (0.010) | (0.010) | (0.010) | (0.011) | (0.010) | (0.011) |
| Surgeon's market share | | | | | -0.198 | -0.238 |
| | | | | | (0.083) | (0.088) |
| Medicaid x mkt. share | | | | | -0.073 | -0.074 |
| | | | | | (0.031) | (0.032) |
| *Specification* | *MNL* | *Mixed* | *MNL* | *Mixed* | *MNL* | *Mixed* |

*Notes:* Models are defined in equations (2) and (3), for multinomial logit (MNL) specifications. Mixed logit specifications include random coefficients on distance variables. All models include: distance between patient and surgeon, distance between cardiologist and surgeon, indicators for cardiologist and surgeon attending same medical school, working at same hospital, and working in same hospital system, and two indicators for surgeons without CABG mortality rates or without valve repair mortality rates listed in the report card. Standard errors shown in parentheses. Full results are presented in Appendix Tables B1-B3.

# Appendix A. Analysis of Forecast Bias

Here we first provide the derivation of one result presented in Section 2 and then develop a related result on entry. The result reported in equation (10) is found as follows:

$$\frac{dY_j}{dQ_k} = \frac{\exp(\beta Q_j + \gamma Y_j)\,\gamma\,\frac{dY_j}{dQ_k}}{\sum_l \exp(\beta Q_l + \gamma Y_l)} - \frac{\exp(\beta Q_j + \gamma Y_j)\exp(\beta Q_k + \gamma Y_k)\left(\beta + \gamma\frac{dY_k}{dQ_k}\right)}{[\sum_l \exp(\beta Q_l + \gamma Y_l)]^2}$$

$$= \gamma Y_j \frac{dY_j}{dQ_k} - Y_j Y_k \left(\beta + \gamma \frac{dY_k}{dQ_k}\right); \qquad (A1)$$

hence,

$$\frac{dY_j}{dQ_k} = -Y_j Y_k \left(\beta + \gamma \frac{dY_k}{dQ_k}\right) \cdot (1 - \gamma Y_j)^{-1}. \qquad (A2)$$

This is the marginal effect of the quality of provider $k$ on the market share of provider $j$; i.e., the marginal substitution effect of competitor quality on own market share

To analyze entry with a similar approach, we include distance in the simple model from Section 2 so that entry can be represented as a (substantial) decrease in the distance between provider $k$ and the patients in the market. Suppose that all the patients and all the providers, except provider $k$, are at the same location, and let the variable $X_j$ be the distance between provider $j$ and all the patients ($X_j = 0$ for all providers except $k$). The model is now:

$$Y_j = \frac{\exp(\beta Q_j + \delta X_j + \gamma Y_j)}{\sum_l \exp(\beta Q_l + \delta X_l + \gamma Y_l)}, \qquad (A3)$$

with $\beta > 0$, $\delta < 0$, and $\gamma < 0$. The marginal effect of the distance between provider $k$ and the patients in the market, on the market share of some provider $j \neq k$, is found as follows:

$$\frac{dY_j}{dX_k} = \frac{\exp(\beta Q_j + \delta X_j + \gamma Y_j)\,\gamma\,\frac{dY_j}{dX_k}}{\sum_l \exp(\beta Q_l + \delta X_l + \gamma Y_l)} - \frac{\exp(\beta Q_j + \delta X_j + \gamma Y_j)\exp(\beta Q_k + \delta X_k + \gamma Y_k)\left(\delta + \gamma\frac{dY_k}{dX_k}\right)}{[\sum_l \exp(\beta Q_l + \delta X_l + \gamma Y_l)]^2}$$

$$= \gamma Y_j \frac{dY_j}{dX_k} - Y_j Y_k \left(\delta + \gamma \frac{dY_k}{dX_k}\right); \qquad (A4)$$

hence,

$$\frac{dY_j}{dX_k} = -Y_j Y_k \left( \delta + \gamma \frac{dY_k}{dX_k} \right) \cdot \left( 1 - \gamma Y_j \right)^{-1}. \qquad (A5)$$

By contrast, in a model without a congestion effect, the predicted marginal effect would be: $\frac{dY_j}{dX_k} = -Y_j Y_k \hat{\delta}$, where $\hat{\delta}$ is the estimated coefficient on distance. As with the marginal effects of competitor quality, the key difference between these marginal effects in the standard model and the congestion model is the factor $\left( 1 - \gamma Y_j \right)^{-1}$ in (A5). Given some provider $k$, (i.e., the one entering the market), this factor makes the magnitude of the substitution effect be relatively smaller for providers with larger market shares. Thus when a new provider "enters" the market (i.e., gets closer to the patients) the substitution away from higher quality providers, which have larger market shares, is relatively smaller in the congestion model, compared to what is predicted by the standard model.

# Appendix B. Supplemental Tables

The tables that follow contain the full sets of parameter estimates for the multinomial logit and mixed logit models estimated in each regional market.

**Table B1. Parameter Estimates, Central PA Market**

| Variables | ML/MSL Estimation | | | 2-Step Estimation | | | 2-Step Estimation | | |
| | Mult. Logit | Mixed Logit | | Mult. Logit | Mixed Logit | | Mult. Logit | Mixed Logit | |
| | | Mean | Std. Dev. | | Mean | Std. Dev. | | Mean | Std. Dev. |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Risk-adjusted mort. rate | -0.110 | -0.136 | | -0.379 | -0.396 | | -0.661 | -0.716 | |
| | (0.061) | (0.063) | | (0.263) | (0.261) | | (0.303) | (0.283) | |
| Elixhauser x RAMR | -0.037 | -0.039 | | -0.038 | -0.039 | | -0.035 | -0.037 | |
| | (0.010) | (0.010) | | (0.010) | (0.011) | | (0.010) | (0.011) | |
| Surgeon's market share | | | | | | | -0.198 | -0.238 | |
| | | | | | | | (0.083) | (0.088) | |
| Medicaid x mkt. share | | | | | | | -0.073 | -0.074 | |
| | | | | | | | (0.031) | (0.032) | |
| Patient-surgeon dist. | -0.058 | -0.075 | -0.032 | -0.017 | -0.032 | 0.020 | -0.017 | -0.032 | 0.020 |
| | (0.006) | (0.008) | (0.018) | (0.010) | (0.012) | (0.011) | (0.010) | (0.012) | (0.011) |
| Card.-surgeon dist. | -0.095 | -0.178 | 0.091 | -0.145 | -0.202 | -0.067 | -0.144 | -0.202 | -0.067 |
| | (0.007) | (0.017) | (0.014) | (0.012) | (0.019) | (0.011) | (0.012) | (0.019) | (0.011) |
| Card.-surg. same med. sch. | -0.171 | 0.066 | | 0.202 | 0.230 | | 0.219 | 0.250 | |
| | (0.492) | (0.485) | | (0.484) | (0.491) | | (0.484) | (0.492) | |
| …same hospital | 0.413 | 0.221 | | 0.585 | 0.601 | | 0.585 | 0.601 | |
| | (0.223) | (0.234) | | (0.240) | (0.253) | | (0.241) | (0.254) | |
| …same hospital sys. | -0.171 | -0.280 | | 0.421 | 0.203 | | 0.417 | 0.196 | |
| | (0.203) | (0.207) | | (0.261) | (0.265) | | (0.261) | (0.265) | |
| CABG mort. not reported | -0.478 | -0.255 | | -1.270 | -1.084 | | -1.912 | -1.802 | |
| | (0.167) | (0.176) | | (0.725) | (0.766) | | (0.683) | (0.655) | |
| Valve mort. not reported | -0.264 | -0.272 | | 0.757 | 0.675 | | 0.580 | 0.393 | |
| | (0.151) | (0.152) | | (0.666) | (0.705) | | (0.476) | (0.478) | |
| First-stage F-stat. | | | | | | | 49.94 | 36.46 | |
| Observations | 17,601 | 17,601 | | 17,601 | 17,601 | | 17,601 | 17,601 | |

Notes: Estimated standard deviations of random coefficients in mixed logit models are reported in separate columns, as indicated. First-stage F-statistics refer to 2SLS estimation of equation (6). Standard errors shown in parentheses.

**Table B2. Parameter Estimates, Southeastern PA Market**

| Variables | ML/MSL Estimation Mult. Logit (1) | Mixed Logit Mean (2) | Std. Dev. (3) | 2-Step Estimation Mult. Logit (4) | Mixed Logit Mean (5) | Std. Dev. (6) | 2-Step Estimation Mult. Logit (7) | Mixed Logit Mean (8) | Std. Dev. (9) |
|---|---|---|---|---|---|---|---|---|---|
| Risk-adjusted mort. rate | 0.057 | 0.057 | | 0.098 | 0.044 | | 0.116 | 0.070 | |
| | (0.024) | (0.026) | | (0.090) | (0.090) | | (0.156) | (0.168) | |
| Elixhauser x RAMR | -0.005 | -0.005 | | -0.010 | -0.008 | | -0.009 | -0.008 | |
| | (0.004) | (0.004) | | (0.005) | (0.005) | | (0.005) | (0.005) | |
| Surgeon's market share | | | | | | | -1.852 | -2.389 | |
| | | | | | | | (0.750) | (1.076) | |
| Medicaid x mkt. share | | | | | | | -0.285 | -0.302 | |
| | | | | | | | (0.052) | (0.055) | |
| Patient-surgeon dist. | -0.083 | -0.125 | 0.085 | -0.079 | -0.136 | -0.077 | -0.080 | -0.137 | -0.078 |
| | (0.004) | (0.006) | (0.009) | (0.005) | (0.009) | (0.016) | (0.005) | (0.009) | (0.016) |
| Card.-surgeon dist. | -0.081 | -0.148 | -0.081 | -0.117 | -0.208 | -0.119 | -0.116 | -0.207 | -0.119 |
| | (0.003) | (0.007) | (0.007) | (0.005) | (0.011) | (0.010) | (0.005) | (0.011) | (0.010) |
| Card.-surg. same med. sch. | -0.057 | -0.094 | | -0.300 | -0.322 | | -0.282 | -0.305 | |
| | (0.133) | (0.139) | | (0.139) | (0.143) | | (0.139) | (0.144) | |
| …same hospital | 0.706 | 0.482 | | 0.779 | 0.538 | | 0.785 | 0.541 | |
| | (0.086) | (0.093) | | (0.112) | (0.127) | | (0.112) | (0.127) | |
| …same hospital sys. | 0.616 | 0.700 | | 1.160 | 1.292 | | 1.149 | 1.282 | |
| | (0.089) | (0.093) | | (0.122) | (0.136) | | (0.121) | (0.136) | |
| CABG mort. not reported | 0.160 | 0.238 | | 1.191 | 1.052 | | 1.781 | 1.623 | |
| | (0.076) | (0.079) | | (0.389) | (0.460) | | (0.868) | (1.047) | |
| Valve mort. not reported | -0.589 | -0.751 | | -0.305 | -0.641 | | -3.089 | -4.260 | |
| | (0.060) | (0.066) | | (0.442) | (0.447) | | (1.361) | (1.870) | |
| First-stage F-stat. | | | | | | | 11.43 | 6.95 | |
| Observations | 143,613 | 143,613 | | 143,613 | 143,613 | | 143,613 | 143,613 | |

Notes: Estimated standard deviations of random coefficients in mixed logit models are reported in separate columns, as indicated. First-stage F-statistics refer to 2SLS estimation of equation (6). Standard errors shown in parentheses.

**Table B3. Parameter Estimates, Western PA Market**

| Variables | ML/MSL Estimation Mult. Logit (1) | ML/MSL Mixed Logit Mean (2) | ML/MSL Mixed Logit Std. Dev. (3) | 2-Step Estimation Mult. Logit (4) | 2-Step Mixed Logit Mean (5) | 2-Step Mixed Logit Std. Dev. (6) | 2-Step Estimation Mult. Logit (7) | 2-Step Mixed Logit Mean (8) | 2-Step Mixed Logit Std. Dev. (9) |
|---|---|---|---|---|---|---|---|---|---|
| Risk-adjusted mort. rate | -0.014 | 0.007 | | -0.006 | 0.055 | | 0.020 | 0.092 | |
| | (0.032) | (0.033) | | (0.074) | (0.096) | | (0.143) | (0.131) | |
| Elixhauser x RAMR | 0.001 | -0.001 | | 0.001 | -0.003 | | 0.001 | -0.003 | |
| | (0.005) | (0.005) | | (0.005) | (0.006) | | (0.005) | (0.006) | |
| Surgeon's market share | | | | | | | -1.160 | -0.705 | |
| | | | | | | | (0.749) | (0.459) | |
| Medicaid x mkt. share | | | | | | | -0.072 | -0.069 | |
| | | | | | | | (0.065) | (0.069) | |
| Patient-surgeon dist. | -0.042 | -0.066 | -0.049 | -0.051 | -0.109 | 0.064 | -0.051 | -0.109 | 0.064 |
| | (0.003) | (0.005) | (0.006) | (0.004) | (0.009) | (0.010) | (0.004) | (0.009) | (0.010) |
| Card.-surgeon dist. | -0.042 | -0.084 | -0.047 | -0.050 | -0.126 | -0.076 | -0.050 | -0.126 | -0.076 |
| | (0.002) | (0.006) | (0.006) | (0.003) | (0.012) | (0.010) | (0.003) | (0.012) | (0.010) |
| Card.-surg. same med. sch. | -0.084 | -0.194 | | -0.343 | -0.278 | | -0.346 | -0.283 | |
| | (0.255) | (0.266) | | (0.272) | (0.288) | | (0.272) | (0.288) | |
| …same hospital | 1.149 | 1.075 | | 1.425 | 1.314 | | 1.426 | 1.315 | |
| | (0.107) | (0.113) | | (0.127) | (0.137) | | (0.127) | (0.137) | |
| …same hospital sys. | 0.411 | 0.620 | | 0.712 | 0.797 | | 0.713 | 0.798 | |
| | (0.106) | (0.116) | | (0.140) | (0.156) | | (0.140) | (0.156) | |
| CABG mort. not reported | -0.192 | -0.113 | | -0.757 | -0.714 | | -1.187 | -0.863 | |
| | (0.185) | (0.187) | | (0.905) | (1.146) | | (1.464) | (1.431) | |
| Valve mort. not reported | -0.904 | -0.914 | | -0.524 | -0.481 | | -1.799 | -1.449 | |
| | (0.139) | (0.141) | | (0.600) | (0.645) | | (1.217) | (0.956) | |
| First-stage F-stat. | | | | | | | 3.98 | 6.57 | |
| Observations | 67,039 | 67,039 | | 67,039 | 67,039 | | 67,039 | 67,039 | |

Notes: Estimated standard deviations of random coefficients in mixed logit models are reported in separate columns, as indicated. First-stage F-statistics refer to 2SLS estimation of equation (6). Standard errors shown in parentheses.