I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12129

Improving Estimation of Labor Market Disequilibrium Using Shortage Indicators, with an Application to the Market for Anesthesiologists

Matthew Baird
Lindsay Daugherty
Krishna Kumar

JANUARY 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12129

Improving Estimation of Labor Market Disequilibrium Using Shortage Indicators, with an Application to the Market for Anesthesiologists

Matthew Baird
RAND Corporation

Lindsay Daugherty
RAND Corporation

Krishna Kumar
RAND Corporation and IZA

JANUARY 2019

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Improving Estimation of Labor Market Disequilibrium Using Shortage Indicators, with an Application to the Market for Anesthesiologists*

While economic studies often assume that labor markets are in equilibrium, there may be specialized labor markets that are likely in disequilibrium. We develop a new methodology to improve the estimation of a reduced form disequilibrium model from the existing models by incorporating survey-based shortage indicators into the model and estimation. Our shortage-indicator informed disequilibrium model includes as a special case the foundational model of Maddala and Nelson (1974). We demonstrate the gains in information provided by our methodology. We show how the model can be implemented by applying it to the market for anesthesiologists, a profession susceptible to disequilibrium. In this application, we find that our new disequilibrium model informed by a shortage indicator fits the data better than the Maddala-Nelson model, and has better out-of-sample predictive power.

JEL Classification: J20, J44, I11, C18

Keywords: disequilibrium, labor demand, labor supply, shortage, maximum likelihood, health providers

Corresponding author:
Matthew Baird
RAND Corporation
4570 Fifth Avenue Suite 600
Pittsburgh, PA 15213
USA
E-mail: mbaird@rand.org

* This study was supported by the American Society of Anesthesiologists. We thank the American Society of Anesthesiologists advisory committee, participants at the Essen Health Conference, Western Economic Association International Annual Conference, and the RAND Labor & Population brownbag, and Jesse Matheson, Italo Gutierrez, Jeff Wenger, Jim Hosek, and Misha Dworsky for helpful comments.

1 Introduction

Labor markets are often assumed to be relatively flexible, with workers receiving wages close to the value of their marginal product of labor, and these wages adjusting in the aggregate to ensure that the supply and demand of labor are equilibrated. However, assumptions of wage flexibility and the resulting equilibrium are difficult to defend in some markets. For example, consider the case of highly specialized segments of the labor market that require years of training and subsequent licensing, resulting in very thin markets. Medical specialties are one important example of such exceptions. Barriers to entry to the profession are natural, arising from the rigors of qualifying, as well as regulations of the relevant associations of professionals to restrict labor supply. Moreover, government involvement in the reimbursement for services and the regulation of the provision of these services and of the facilities that provide them places restrictions on the demand for labor.

Disequilibrium models are difficult to model, perhaps explaining why there has not been extensive work in this area. Maddala and Nelson (1974) and Gourieroux, Laffont, and Monfort, (1980), and reemphasized in Gourieroux (2000), provided a reduced-form maximum likelihood approach that requires specification of a demand and supply equation. Lubrano (1985) and Lubrano (1986)'s model differs from these models by using a switching regression approach in a Bayesian framework (discussed later). Later models allowed for a dynamic aspect where supply and demand depend partially on prior quantity (Laroque and Salanie 1993 and Lee 1997 extending Maddala and Nelson 1974, and Bauwens and Lubrano 2007, extending the switching regression models). Further, disequilibrium models are not only used in labor markets; at times they are applied to credit markets (such as in

Bauwens and Lubrano 2007 and Hubbs and Kuethe 2017).

We expand on the literature of disequilibrium models by starting again from Maddala and Nelson's (1974) base model using an innovative strategy that incorporates additional information from surveys into the likelihood function. Specifically, we propose a disequilibrium model that directly uses indicators correlated with shortage or surplus to improve estimation (henceforth, shortage indicators). Our model is reduced-form in line with the prior research, and the Maddala-Nelson model is a special case of our more general model. As a by-product of our model, we also get useful information about the relationship between the shortage indicators and actual shortage, such as the average level of a shortage indicator in equilibrium. This gives insight into the natural rate of the indicator (say, the proportion of workers whose employees are actively attempting to hire more of the same type of workers) in a given industry. We also get information on the extent to which increased actual shortage affects the shortage indicator. This helps researchers and policymakers understand how indicators of interest (say, the fraction of workers working in an office where medical procedures are delayed due to a lack of anesthesiologists) might be expected to fluctuate with changes in economic trends (recessions) and policy variables (number of medical residencies in the country).

We demonstrate an application of the model to the labor market for anesthesiologists. The anesthesiology labor market provides an appropriate context in which typical assumptions of flexible labor markets may fail and where our ability to better evaluate labor market conditions is likely to be important for policy decisions. Shortages in such a critical specialty would have important implications for access to care, leading to waits in hiring, delays in necessary medical procedures, and potentially increased medical expenditures. On the other hand, a surplus of medical specialists can lead to highly capable, trained, and productive physicians being underutilized, leading to inefficient allocation of human capital, without necessarily improving health outcomes (Baiker and Chandra 2004, Phillips et al. 2005). There are also several reasons, discussed below, why this market may be more likely to be in disequilibrium at any given point in time.

We test four different shortage indicators using our new methodology, as well as two aggregators of the four shortage indicators. We find that the Maddala-Nelson estimator is statistically significantly different from our model that is informed by a shortage indicator. Given Maddala-Nelson is a special case of the more general model, the value of the likelihood function is greater for the shortage indicator informed model, and it fits the data better. We also find that it yields better out-of-sample predictions. Taken together, these outcomes suggest that our new model is a better representation than the Maddala-Nelson framework for evaluating a labor market that may be in disequilibrium.

The rest of the paper proceeds as follows. Section 2 presents the economic setting. Section 3 discusses details of our econometric approach to model disequilibrium and contrasts it with the existing disequilibrium model. Section 4 applies the model to data from our surveys and secondary data sources to analyze the labor market for anesthesiologists. Section 5 discusses and concludes.

2 Economic Setting

We consider a market where there are unobserved Walrasian trade offers, Q^D and Q^S for demand and supply, respectively. Following Maddala and Nelson (1974), in the labor for market m in year t, let the quantity of labor demanded and supplied be given by:

$$Q^D_{mt} = X^D_{mt}\beta^D + \varepsilon^D_{mt} \tag{1}$$

$$Q^S_{mt} = X^S_{mt}\beta^S + \varepsilon^S_{mt} \tag{2}$$

Here, Q^D and Q^S denote the total number of full-time equivalent (FTE) workers demanded and supplied respectively. X^D and X^S include factors influencing demand and supply, most importantly, the wage. Under equilibrium, $Q^D = Q^S = Q$ is observed, and we can estimate the unknown coefficients using standard regression techniques, such as 3SLS, instrumenting for the endogenous wage using excluded variables in each equation. However, if the market is in disequilibrium, then we are unable to observe both the quantity demanded and quantity supplied jointly for each market in each year; we only observe the minimum of the labor demand and labor supply.

An alternative to the above model would be the one introduced by Ginsburgh and Zang (1975) and Ginsburgh, Tishler, and Zang (1980).

$$Q_{mt} = \min(X^D_{mt}\beta^D, X^S_{mt}\beta^S) + \varepsilon_{mt} \tag{3}$$

As discussed by Lubrano (1985), this model represents trade offers that are ex-ante plans;

the observed quantity is the function of these ex-ante plans plus an ex-post unanticipated disturbance. While this model has several favorable statistical properties, as discussed by Richard (1980) and Lubrano (1985), our focus is on extending Maddala-Nelson. Further, we feel that the switching regression framework mentioned earlier is based on an assumption that is less credible; namely, that there is a perfect correspondence between the elements that affect the ex-ante trade plans versus the ex-post disturbances and which elements are observed by the econometrician and which are not. Therefore, we instead focus on the Maddala and Nelson's (1974) basic framework.

Several later models additionally allowed for a dynamic nature where demand and supply are also functions of prior values (either prior demand and supply, respectively, or prior quantity), such as Laroque and Salanie 1993, Lee 1997, and Bauwens and Lubrano (2007). Additionally, Lubrano (1985) and Bauwens and Lubrano show how to estimate these disequilibrium models (focusing on the switching regression, static and dynamic respectively) in a Bayesian framework. We focus on extending the simpler Maddala Nelson framework as a first step; however, we hypothesize that a similar extension done here of incorporating shortage indicators could be done for these other models.

There is more than one potential reason why a labor market could be in disequilibrium, such that there is no market-clearing wage in the short run. This is perhaps especially true for specialized health care providers. Some of the potential reasons for disequilibrium in this context are:

- In the face of a labor demand shock, the inability of labor supply to immediately respond given the finite stock of anesthesiologists and long lead time inherent in training new anesthesiologists can lead to an imbalance in demand and supply.

- Since anesthesiologists typically employ their services during surgeries, which are performed in a limited number of facilities in an HRR, monopsony power in these local labor markets might limit the ability for wages to adjust appropriately to demand shocks.

- Sticky wages arising from long-term contracts could prevent an equilibrium from being reached.

- Fixed reimbursement based on Medicare or other schedules could be another reason for wages to be rigid, preventing market clearing.

- The sluggish nature of adjustments on the extensive margin, as mentioned above, and a similarly low ability to adjust on the intensive margin among such specialized labor (given long hours of work), and potentially low rates of providers to population ratios, could combine to make it difficult for the markets to clear.

We do not take a particular stance on why the labor market for anesthesiologists could be in disequilibrium as much as note that it is natural to conceive disequilibrium in this market given the above reasons. Our estimation method is agnostic to the exact reason for disequilibrium.

Assume that we observe at least one indicator A, which we designate as a "shortage indicator." What makes it is a shortage indicator is that it is a function of the actual shortage or surplus. Consider for example that each facility or employee was asked whether their facility needed to hire more individuals to cover current demand. With higher levels of shortage, we would expect more individuals to respond that they need to hire more individuals. For our purposes here, given shortage is in counts and our shortage indicators are rates, we scale the shortage by the population of the market to get per-capita shortage/surplus. If the shortage indicator were on the same scale, this could be omitted. We assume shortage is some function of the per capita shortage as well as an additively separable error term. Equation 3 presents our specification for general function $h(\cdot)$.

$$A_{mt} = h\left((Q^D_{mt} - Q^S_{mt})/p_m\right) + \nu_{mt} \tag{4}$$

Below we discuss the shortage indicators we use from our surveys, shown in Table 1. p_m is the population in the labor market m, so that the shortage indicator is normalized to depend on per-capita labor shortage. $h(\cdot)$ could be any function, but needs to be defined for all real numbers. Thus, while you could not use the natural log function, which is not defined for non-positive numbers (here, excess supply), we could use the inverse hyperbolic sine function, which mimics the natural log function but allows for negative numbers. In our application, we use the linear case, which offers an easy first step in this analysis.

$$A_{mt} = \gamma_0 + \gamma_1 (Q^D_{mt} - Q^S_{mt})/p_m + \nu_{mt} \tag{5}$$

γ_0 captures the value the shortage indicator is expected in equilibrium. γ_1 describes the relationship between labor shortage and the observed indicator, measuring the average increase in the shortage indicator for each additional FTE worker demanded in excess of supply, per capita. There may be more than one indicator available to the researcher.

Table 1: Shortage indicator definitions

1	Fraction of respondents in the HRR answering "yes" to survey question: "To cover our current volume of cases, my group/practice would prefer to have more anesthesiologists"
2	Fraction of respondents in the HRR answering "yes" to survey question: "My group/practice could handle more cases if we could hire additional anesthesiologists"
3	Fraction of respondents in the HRR answering "Increased by less than 10%" or "Increased by more than 10%" (instead of decreasing by less than 10% or decreasing by more than 10%) to the survey question: "By what percentage have your work hours changed since [three years prior]?"
4	Fraction of respondents in the HRR that do not answer "I will increase my work hours if the compensation is high enough" to the survey question: "What is your attitude toward increased work hours (total hours-clinical, research, and administrative-rather than billable hours)?"
5	Arithmetic unweighted mean of indicators 1-4
6	Search-optimized weighted mean of indicators 1-4

3 Econometric Models

The primary econometric challenge is to discern whether the observed quantity of anesthesiologist labor is supply or demand or both (in the case of equilibrium). If $Q^D > Q^S$ then there is a situation of excess demand or shortage, and only Q^S is observed. If $Q^D < Q^S$ then there is a situation of excess supply or surplus, and only Q^D is observed. Assigning probabilities to the observed quantity being supply or demand is the primary concern of the models discussed in this section.

We contrast three reduced form models: equilibrium, basic disequilibrium models after the example of Maddala and Nelson (1974), and our shortage indicator informed disequilibrium models. Our primary comparison is between the basic class of disequilibrium models (hereafter referred to as MN, after Maddala and Nelson 1974) and our expanded model that includes information from shortage indicators (hereafter, the shortage indicator informed likelihood, or SII). In the literature on basic disequilibrium models, the focus is on the test of whether the market is in equilibrium or not; as this is not the focus of this paper, we rely simply on whether the expected aggregate excess demand is significantly different from zero as our inference for the hypothesis of overall equilibrium. Individual labor markets may be in disequilibrium even while the national market is on average in equilibrium.

Wage is endogenous to the system, and as such, it belongs in the likelihood function for the equilibrium model. However, to reduce the dimension of the parameter space for the maximum likelihood search algorithm, we estimate the first stage regressions of log wages on the determinants of wages (which include the excluded instruments of the demand and supply functions), and use the predicted wages in the maximum likelihood of observed quantities. Doing so with valid excluded variables (demand and supply shifters) removes the endogeneity that arises from the relationship between wages and hours. While the disequilibrium models don't have the same issue (the demand or supply weights show movements along the curves), we can include wages in the likelihood as well, and do for consistency. We estimate the confidence intervals using block bootstrapping, and given that the first stage is included in the bootstrapping procedure, the confidence intervals are still correct with this two-stage estimation procedure. From now on, we suppress the dependency on wages and consider it implied, and suppress the additional estimation of $\Pr(W = w)$, which is the same across the three classes of models.

3.1 Equilibrium

If the markets are in equilibrium, $Q^D_{mt} = Q^S_{mt} = Q_{mt}$ and the model may be estimated by full-information maximum likelihood or the closed-form solutions that GMM offers through Two Stage Least Squares and Three Stage Least Squares on the demand and supply equations given in (1) and (2). The usual approach of instrumenting is necessary to overcome the simultaneity of wage in the demand and supply equation. We estimate the equilibrium model as a frame of reference for the parameter values of the disequilibrium models rather than as a test of whether the markets are in equilibrium.

3.2 Maddala-Nelson Disequilibrium Model (MN)

The basic disequilibrium model assumes that $Q^D_{mt} \neq Q^S_{mt}$. We use our own notation to the problem as set forth by Maddala and Nelson (1974) and others. The likelihood function (suppressing m and t subscripts, and the conditioning on observed X, including log wages) without incorporating the shortage indicator may be expressed following Maddala and Nelson's (1974) derivation. Note that allowing for correlation between supply and demand unobservables requires computational integrals and significantly messier derivations. As a result, similar to Maddala and Nelson, we do not evaluate this case, in favor of the simpler case of independence between supply and demand, which allows us to highlight the contributions of our new model more clearly. In that case, we have

$$\begin{aligned}
\Pr(Q = q) &= \Pr(Q = q | Q^D > Q^S) \Pr(Q^D > Q^S) + \Pr(Q = q | Q^D < Q^S) \Pr(Q^D < Q^S) \\
&= \Pr(Q^S = q) \Pr(Q^D > q) + \Pr(Q^D = q) \Pr(Q^S > q) \qquad (6)
\end{aligned}$$

If we assume the error terms are normally distributed, substituting equations (1) and (2) into equation (6) yields

$$\frac{1}{\sigma_{\varepsilon S}}\phi\left(\frac{q-X^S\beta^S}{\sigma_{\varepsilon S}}\right)\left(1-\Phi\left(\frac{q-X^D\beta^D}{\sigma_{\varepsilon D}}\right)\right)+\frac{1}{\sigma_{\varepsilon D}}\phi\left(\frac{q-X^D\beta^D}{\sigma_{\varepsilon D}}\right)\left(1-\Phi\left(\frac{q-X^S\beta^S}{\sigma_{\varepsilon S}}\right)\right) \tag{7}$$

We then choose the parameters that maximize the log-likelihood function $\sum_{i=1}^{N}\ln(\Pr(Q = q))$. After we estimate the parameters of the model, we are able to estimate the probability of each market being in shortage as well as the expected shortage for each market as well as in aggregate. These derivations are provided in the Appendix.

Table 2: Comparisons of out of sample predictions

	MN	SII1	SII2	SII3	SII4	SII5	SII6
Mean Abs. Bias	95.14	95.24	93.48	101.81	79.77	89.32	76.24
MSPE	15,929	15,538	15,168	18,384	15,852	14,526	15,844

MSPE is the mean squared predicted error

3.3 Shortage Indicator Informed Disequilibrium Model (SII)

We can incorporate information from shortage indicators into our likelihood function, so that it is a function of observed quantity as well as shortage indicators. Recall from equation (4) that the shortage indicator is posited to be a function of the actual (unobserved) excess demand. The likelihood function expands on MN as follows (again, assuming independence across markets and between supply and demand unobservables, for simplicity):

$$\Pr(Q = q, A = a) = \Pr(Q = q|A = a)\Pr(A = a)$$
$$= [\Pr(Q^S = q|A = a)\Pr(Q^D > q|A = a) + \Pr(Q^D = q|A = a)\Pr(Q^S > q|A = a)]\Pr(A = a) \tag{8}$$

If we assume normality, then, as shown in the Appendix, this solves to

$$\begin{aligned}\Pr(Q = q, A = a) = &\left[\frac{1}{\sigma_{CS}}\phi\left(\frac{q - X^S\beta^S - \mu_{CS}}{\sigma_{CS}}\right)\left(1 - \Phi\left(\frac{q - X^D\beta^D - \mu_{CD}}{\sigma_{CD}}\right)\right)\right.\\ &\left. + \frac{1}{\sigma_{CD}}\phi\left(\frac{q - X^D\beta^D - \mu_{CD}}{\sigma_{CD}}\right)\left(1 - \Phi\left(\frac{q - X^S\beta^S - \mu_{CS}}{\sigma_{CS}}\right)\right)\right]\\ &\times \phi\left(\frac{a - E[A|X^D, X^S, p]}{\sigma_A}\right)/\sigma_A\end{aligned} \tag{9}$$

where for $j = \{S, D\}$

$$\mu_{Cj} = \frac{\sigma_{QjA}}{\sigma_A^2}(a - E[A|X^D, X^S, p]) \tag{10}$$

$$\sigma_{Cj} = \sigma_{\varepsilon j}\left(1 - \frac{\sigma_{QjA}^2}{\sigma_{\varepsilon j}^2\sigma_A^2}\right)^{.5} \tag{11}$$

$$\sigma_{QjA} = Cov(\varepsilon^j, h((X^D\beta^D + \varepsilon^D - X^S\beta^S - \varepsilon^S)/p)) \tag{12}$$

$$\sigma_A^2 = Var(h((X^D\beta^D - X^S\beta^S)/p)) + \sigma_\nu^2 \tag{13}$$

The difference between the Maddala Nelson (MN) likelihood function in equation (7) and our expanded likelihood function in equation (9) illustrate the gains of our shortage indicator informed disequilibrium model (SII). First, note that MN is a special case of SII where $\sigma_{QjA} = 0$, that is, when the error term of demand or supply is not correlated with the observed portion of the shortage indicator function. That is, the shortage indicator has no additional ability to explain the remaining unobserved portion of the demand and supply functions. For the linear case of h that we employ in our application, this reduces to a single parameter difference: $\gamma_1 = 0$ (i.e., the shortage indicator is not a function of shortage and thus gives no additional information).

If the shortage indicator is informative, then the additional information from the shortage indicator adjusts the likelihood in an intuitive way. For example, consider the element of the likelihood function representing whether a specific market is in shortage, $1 - \Phi\left(\frac{q - X^D\beta^D - \mu_{CD}}{\sigma_{CD}}\right)$. This is the probability that demand exceeds (observed quantity) supply. Higher values of $X^D\beta^D$ increase this probability, which is true for both MN and SII. However, for SII, higher values of μ_{CD} also increase the probability of the market being in shortage, and more weight being put on matching observed quantity to the demand supply equation. Note that this doesn't necessarily occur when a, the shortage indicator, is high, but only if it exceeds the predicted shortage indicator conditional on excess demand, if it provides additional information. In fact, if the shortage indicator is equal to the expected shortage indicator, then this portion of the likelihood is identical to the one for MN. However, if they are not equal, the there is additional information to be gleaned and the likelihood is adjusted. For example, consider the case when the observed shortage indicator in a market exceeds the expected shortage indicator given the supply and demand equations; there will be a higher value on the weight that will be put on matching $X^S\beta^S$

to q. This is because it assumes that the labor market is more likely to be in the state of excess demand, and the observed q is the labor supply, rather than the labor demand.

Each element of the likelihood function equations (7) and (9) have a similar potential adjustment depending on the shortage indicator for each labor market. If in fact the shortage indicator has the posited relationship with excess demand, then the expansion of the likelihood function from equation (7) to equation (9) yields more accurate measurements of the parameters of the model by more accurately discriminating between cases of shortage and surplus in each labor market, and matching the observed quantity to the appropriate independent variables. The shortage indicators can shift both the probability weights of being in shortage or surplus in equation (9) as well as the matching of the quantity observed and the predicted quantity from the supply and demand equations. There is also the last element of SII in equation (9), $\phi\left(\frac{a-E[A|X^D,X^S,p]}{\sigma_A}\right)/\sigma_A$, which differs from MN in equation (7). This serves to estimate the parameters of the shortage indicator function.

After we estimate the parameters, we also adjust how expectations of demand and supply are calculated, and thus the expected shortage. When we incorporate the shortage indicators, we have

$$E[Q^D|q,x^D,x^S,A] = E[Q^D|q,x^D,x^S] + \frac{\sigma_{Q^D,A}}{\sigma_A^2}(a - E[A|q,x^D,x^S]) \tag{14}$$

$$E[Q^S|q,x^D,x^S,A] = E[Q^S|q,x^D,x^S] - \frac{\sigma_{Q^S,A}}{\sigma_A^2}(a - E[A|q,x^D,x^S]) \tag{15}$$

Taking the expected quantity demanded, note that the first element $E[Q^D|q,x^D,x^S]$ is just the expectation under MN. The second element shifts the expectation depending on how the shortage indicator varies from the predicted shortage indicator. For example, if the observed shortage indicator exceeds the predicted one, and given we expect a positive correlation to exist between quantity demanded and the shortage indicator, then we would increase the expectation above that given by MN in this case. Similar to the likelihood function, if the observed shortage indicator is exactly equal to the predicted one by MN, then it contains no new information and we don't shift the expectation at all. The expected shortage is given by $E[Q^D|q, x^D, x^S] - E[Q^S|q, x^D, x^S]$, and the aggregate shortage is the sum of this difference across all of the markets.

Thus far, we have only considered the case where there is a single shortage indicator. This model may be expanded to include more than one shortage indicator by including a vector of shortage indicators that expand the likelihood. Alternatively, one can collapse multiple shortage indicators into a single indicator. We test the latter, and do so in two ways: a simple average of all of our indicators, and an indexed average where the weights are part of the parameter space and included in the search algorithm.

4 Application to Labor Market of Anesthesiologists

We use the example of the labor market of anesthesiologists to examine the differences between MN and SII. Anesthesiology provides an appropriate context because there are market imperfections that can lead to the anesthesiologist labor markets to not be in equilibrium at any given point in time. There is an open discussion concerning the direction and extent of shortages of medical specialties. Dall et al. (2013) project future demand and supply among various medical specialties, and predict a substantial increase in demand for physician services: a 14% increase in demand for FTE primary care physicians from 2013 to 2025, with an even larger increase for most specialties. They contend that insufficient attention to expanding supply of medical specialists could lead to shortages, causing longer wait times and reduced access to care.

Schubert et al. (2012) estimated a shortage of 2,000 anesthesiologists (the specialists of interest in this paper) in 2007.[1] Schubert et al. (2012) further conclude that there is evidence for persistence shortage at the national level, which seems to have been diminishing over the last two decades.[2] They suggest, albeit without making any quantitative statements, that increases in the number of new anesthesiologists, lower compensation and decreased demand due to the recession have led to the decrease in earlier estimated shortages, but warn that smaller residency graduation in the future along with demographic shifts (gender and age) related to willingness to work may exacerbate shortage in the future. Since our estimation procedure is agnostic about the reason for disequilibrium, if it exists, it allows for any of these factors to affect regional labor market conditions. Using a variety of factors to quantify shortage or surplus by state through econometric methods is the novelty of our approach.

One reason there may be disequilibrium in the anesthesiologist labor market is the lag with which supply is able to respond to perceived needs; it takes years for a potential anesthesiologist to go through medical school and complete a residency in anesthesiology. Should the market have a large demand shock for anesthesiologist services, even if hospitals can offer higher wages, it does not speed up the process of reaching a new equilibrium.

[1] See Baird et al. (2015) for a more detailed literature review and discussion on this topic.

[2] Schubert et al. (2012) also provide a good review of the literature regarding evidence for disequilibrium in the market for anesthesiologists. See also Daugherty et al. (2010).

Likewise, adjustment might be slow in the face of negative demand shocks, as anesthesiologists may be protected by long-term contracts (Stulberg, R. and A. Shulman (2013), Bierstein (2005), Cromwell (1999)). Adjustment is likely to be slow on the intensive margin as well. Our surveys reveal numerous cases where anesthesiologists were unable or unwilling to increase their number of hours, even with an increase in their pay. Around 27% of anesthesiologists we surveyed responded that they would not increase their hours because they did not have any more time available. Only around 37% said they would be willing to increase their hours if their compensation was high enough. When asked why they would not increase hours for any compensation, answers included personal reasons such as family reasons and the need for work-life balance. However, some of the replies indicated inability to increase hours due to institutional restrictions, including already operating at the maximum allowed number of hours.

There are also potential barriers to equilibrium from the demand side: Hospitals and the medical industry in general operate under heavy regulation (Melly and Puhani 2013, Daugherty et al. 2010, Abenstein 2004). HMOs and pay-for-service arrangements (e.g. fixed or capped prices for healthcare services, create wedges between market clearing wages and what can actually be offered to anesthesiologists (Robinson et al. 2004, Madison 2004, Hillman 1987).

Data for the variables contained in labor supply comes primarily from two surveys we administered to anesthesiologists in 2007 and in 2013. We refer to these surveys as the RAND Surveys. They are described in more detail in Section 4.4, as well as in Baird et al. (2015). We use Hospital Referral Regions (HRR) as the labor market unit of analysis. HRRs are geographical regions in the United States defined by the Dartmouth Atlas Project. They represent regional health care markets with at least one hospital that performs major cardiovascular procedures and neurosurgery.[3] There are currently 304 HRRs in the United States. We aggregate the data to the HRR labor market level by year. Data for variables contained in labor demand come primarily from external data sources, including the Area Health Resource File (AHRF) which we crosswalk to the HRR. Given our small number of observations (180 labor markets for which we have sufficient data in 2 different years for 360 observations), we aimed for parsimony in the labor demand and labor supply functions.[4] The results are not very sensitive to the inclusion of additional covariates, and the variables included seemed a priori to be the more relevant factors.

4.1 Labor Demand Function

The primary variable affecting demand is the average log wage of anesthesiologists in the HRR. Increased wages make anesthesiologists more expensive, decreasing demand for their services. The RAND Surveys provide us with wage data. We also include the log of the total number of surgeries in the HRR (irrespective of whether an anesthesiologist participated in the surgery or not). This is a good measure of demand for health services for which anesthesiologists would be required. We include the log of the population in the geography covered by the HRR as well as the log of median household income of that population. Increases in either population or income in the population in the market should increase the demand for anesthesiologist hours. Demand is also modeled as a function of the number of

[3] http://www.dartmouthatlas.org/tools/faq/researchmethods.aspx

[4] We keep an HRR if we have 5 or more ANs surveyed in the HRR or at least one quarter of all ANs that work in the HRR surveyed

Certified Registered Nurse Anesthetists (NAs), interacted with the opt-out status of state (states where NAs are able to perform anesthesia unsupervised). By including NAs we account for the complementarity in the production of anesthesia services. In opt-out states, NAs may serve as more of substitutes for anesthesiologists (Kalist et al. 2011, Kane and Smith 2004). The number of NAs is derived from the AHRF. Surgeries, population, income, and the number of NAs are taken from the AHRF. We also include the local unemployment rate, available from the Bureau of Labor Statistics Local Area Unemployment Statistics files. Finally, we also include a year dummy for 2013 to allow for different baseline aggregate demand.

4.2 Labor Supply Function

As with the labor demand function, the labor supply function for anesthesiologists contains variables that affect supply on the intensive or extensive margin. The first and primary variable is the average wage in the market. Higher wages induce current anesthesiologists to work more hours, and for more anesthesiologists to move to areas of high demand. The coefficient on log wages is a function of the labor supply elasticity. In the RAND Surveys, we asked each respondent for the wage increase necessary to induce a 10% increase in work, from which we can estimate an individual labor supply elasticity. See Daugherty et al. (2010) for details concerning how we estimate the elasticity from the questions and the resulting distribution of elasticities by state. Rather than estimate the coefficient on log wages in the supply function using appropriate demand shifters as instruments, we use the HRR-averaged survey elasticities directly. Thus, we do not need to rely on good instrumental variables, as we have direct estimates of the elasticity. Given our labor supply model is a level-log model, the elasticity is equal to the coefficient on log wages divided by the quantity, or equivalently, the coefficient on log wage is equal to the elasticity multiplied by the quantity. Thus, we multiply the elasticity, the quantity, and log-wages and subtract this product from the observed quantity for labor supply. The dependent variable is $Q_r^S(1 - elast_{Sr} * \log(wage_r))$ for health market r. This simplifies the analysis by requiring fewer assumptions on valid instruments. However, we note again that instruments are not necessary in the disequilibrium model, but also do no harm.

From the RAND surveys we include additional labor supply factors: the fraction of anesthesiologists that are male (male anesthesiologists are more likely to work more hours than females, Baird et al. 2015), the fraction of anesthesiologists working fewer than 30 hours (which reveals both work preferences of the local anesthesiologist population and the available capacity to increase labor hours), the fraction of anesthesiologists working in an urban area (making labor hour increases easier with smaller transportation costs, and also potentially related to anesthesiologist living preferences and thus the labor supply extensive margin). We also include the local unemployment rate, the log population in the HRR, and a dummy for the year 2013 to allow for overall time-dependent shifts in labor supply.

4.3 Identification of the Elasticity of Labor Demand

Wages and labor demand are jointly determined, so that estimation of the coefficient on wages in the labor demand function (which is proportional to the underlying labor demand elasticity) is endogenous. In equilibrium, movements in observed quantity and wage may be movements along a given curve (which trace out the slope and thus elasticity) or shifts in a curve, which do not trace out the desired relationship. We identify the coefficient based on the supply shifters: excluded variables that are in the labor supply function that serve to map out the slope of the labor demand function with respect to wages, and hence the coefficient on wages.

Our excluded instruments for the elasticity of labor demand are the fraction of anesthesiologists in the market that are working part-time, the fraction that are female, and the fraction working in an urban area. We argue that each of these has an effect on labor supply, as described in Section 4.2, but has no independent effect on labor demand. It is hard to imagine how the gender of those providing services or the part-time nature of work would directly affect the demand for anesthesiologist services. Justifying the fraction working in an urban area is potentially more difficult, as a higher urban concentration might increase demand for services. However, the primary avenues through which it would affect demand–higher population and lower income–are already included in the demand function. Furthermore, we find the results not sensitive to the inclusion of this instrument. In the 3SLS equilibrium model, the average elasticity with urban included as an excluded instrument is -2.8. If it is included in both demand supply functions, the average elasticity is estimated to be -2.9. These instruments are not necessary for identification in the disequilibrium model.

4.4 Data

We conducted surveys of members of the American Society of Anesthesiologists in 2007 and then again in 2013. Of the 29,158 American Society of Anesthesiologists members invited to respond for the 2013 survey, 6,825 of did so, which yielded a response rate of 23%. The 6,825 respondents represent a sample of the total of 42,230 anesthesiologists practicing in the United States. To correct for non-response bias in the survey and differences between American Society of Anesthesiologists members and the larger anesthesiologist population, we condition non-response on observed covariates, and create weights to aggregate to the state and national levels. Details about the survey respondents and their characteristics can be found in Baird et al. (2015).

Although there are 304 HRRs in the United States, we only include those for which we have sufficient number of observations to estimate the averages within the HRR. For our purposes, we only include HRRs for which we have at least 5 survey respondents or at least 25% of all anesthesiologists in the HRR responding to our surveys. We only keep HRRs for which we have data for both years of the survey. This leaves us with a final sample of 180 markets in 2 years, or 360 total market/year observations. We have each respondent's working zip code in the RAND Surveys; HRRs are defined as collections of zip codes, so we can easily aggregate the values up to the HRR level for each of these variables.

For the AHRF and Local Area Unemployment Series data, variables are defined at the county level. We crosswalk each HRR zip codes to the counties, and create a weighted aggregation depending on the relative populations of the counties included in the zip codes.

We also examine shortage indicators derived from our survey for our SII model. We test four separate shortage indicators, as well as their average and optimal-weighted average, for six separate tests of the SII model. The optimal-weighted estimator jointly searches for the parameters of the likelihood function as well as three weights for the first three shortage indicators (the fourth being normalized to 1 minus the other three). Table 1 describes these shortage indicators we use in this paper. Among the four shortage indicators, our prior is to trust best the first one, as this seems most representative of underlying shortage. However, overall we trust the unweighted average of the four indicators the most, as this uses the most information while not putting additional structure on the model.

In certain cases we report the results only for one shortage indicator for brevity; for this we use our ex-ante preferred single indicator, shortage indicator 1 in Table 1. The values range from 0 (no ANs in that HRR work in a facility trying to hire more ANs) to 1 (all ANs in that HRR work in a facility trying to hire more ANs). The average is about half, which is to say that the average HRR has half of the ANs working in such a facility. The standard deviation of the average is relatively large as well at around 0.2, suggesting a significant amount of variation in this variable across HRRs, providing good variation for our analysis and showing that there may be differences in the likelihood of a given market being in equilibrium, shortage, or surplus. Table 3 presents the market-level summary statistics.

4.5 Results

We estimate the models using maximum likelihood. We use the Nelder-Mead simplex search algorithm, using as initial values the equilibrium 3SLS parameter estimates as well

Table 3: HRR-level summary statistics

Variable	Mean	Std. Dev.	Min	Max
Total Anesthesiologist FTE	210.5	223	15.67	1685
Average wage	142.7	21.11	81.45	226.4
Total surgeries (1,000s)	135.4	111.8	17.36	645.2
Median household income	27423	8787	8707	71990
Nurse Anesthetists x opt-out state	0.034	0.0968	0	0.914
Nurse Anesthetists x non-opt-out state	0.163	0.189	0	1.094
Fraction working under 30 hours/week	0.0592	0.0649	0	0.348
Fraction female	0.211	0.121	0	0.509
Fraction working urban	0.929	0.14	0	1
Population	1.45E+06	1.30E+06	229360	1.02E+07
Unemployment rate	3.488	1.542	0.512	9.206
Elasticity of labor supply	0.355	0.146	0	0.599
Facility prefers more ANs to cover workload	0.482	0.194	0	1
Facility could handle more cases if more ANs were hired	0.371	0.191	0	0.914
Have increased hours in past 3 years	0.487	0.191	0	1
Would increase hours for sufficient increase in pay	0.397	0.162	0	1

Note: 360 HRR/year observations

as a random perturbation from these initial values.[5] We select the estimates between these two that has the largest likelihood value. We bootstrap all of the parameters by taking random draws of the RAND Survey respondents and reconstructing the HRRs with those respondents, following the same inclusion rules as before.

Table 4 presents the estimated coefficients of the demand and supply models; for the shortage indicator informed disequilibrium model, we only here present it for shortage indicator 1 for exposition. For the SII model, we have additional parameters related to the shortage indicator as seen in Equation 3. The parameter estimates for each shortage indicator sets are presented in Table 5. γ_0 is the expected value of the shortage indicator

[5]We tested starting from up to 10 different initial starting values but found no changes in the convergence points. In almost all tested cases, the second initial values starting yields the same converged parameters as the first. We include the second only as back up against a local maximum in the bootstrapping.

for a labor market in equilibrium. 0.472 for shortage indicator 1 is slightly lower than the observed value in Table 3 of 0.482, giving us our first indication that this labor market might in aggregate be in surplus. The parameter is significantly different from zero.

γ_1 tells us, for example for shortage indicator 1, that for each additional FTE shortage of ANs per 1,000 residents, we expect the shortage indicator to increase by around 0.27. Note that the average number of FTE ANs in an HRR per 1,000 residents is 0.14, with a minimum of 0.03 and a maximum of 0.47. Thus, a unit increase in the demand for number of FTE ANs per 1,000 is very large. An increase of 5% of the AN per 1,000 average (a moderate shift up in demand) is 0.007, for a marginal effect of an increase in the shortage indicator of 0.002. This seems a reasonable value. However, the result is not statistically significant. We do find statistical significance for indicator 4 as well as the indexed indicators, however. Note that the weights for the search-optimized algorithm are 0.08, 0.14, 0.22, and 0.55 for shortage indicators 1-4 respectively.

With the estimated coefficients and other parameters, we are able to estimate the elasticity of labor demand as well as expected excess demand. These are presented in Table 6. Both models estimate a surplus of ANs in 2007, with both significantly different from zero. However, the MN model predicts a substantially larger surplus. With a national working population of over 30,000, both models predict in 2007 there was a large surplus. For 2013, the MN model continues to estimate a statistically significant surplus of ANs, while the SII model estimates a very small and not statistically different from equilibrium shortage of ANs using shortage indicator 1. However, the other shortage indicators all find large surpluses in 2013, similar to the MN model. Supply elasticities, coming from the

Table 4: Estimated coefficients

		2SLS	3SLS	MN	SII1
Demand: Q^D	Log wage	-583.63	-574.02	-535.69	-686.33
		(201.05)	(198.24)	(381.58)	(379.61)
	Log surgery	53.46***	49.172***	81.826*	89.444
		(18.222)	(16.978)	(29.649)	(38.047)
	Log household income	-6.3964	7.8032	-34.707	-57.459
		(17.846)	(17.158)	(31.66)	(32.615)
	Nurse Anesthetists x opt-out	-64.474**	-9.3187	-149.98	-140.66
		(58.217)	(61.805)	(112.45)	(126.84)
	Nurse Anesthetists x non-opt-out	-82.721***	-44.927**	-176.98**	-177.64**
		(0.031576)	(0.032655)	(0.049526)	(0.057418)
	Log population	212.52***	210.37***	184.52***	161.8***
		(15.894)	(15.236)	(27.781)	(35.005)
	Unemployment rate	-17.569***	-19.698***	-19.207**	-13.972**
		(4.0375)	(4.1097)	(7.3796)	(7.288)
	Year 2013	-36.275	-33.865	-14.415	-24.475
		(15.507)	(15.37)	(30.655)	(31.3)
	Constant	-310.88*	-425.33**	-95.802	1106.1
		(1072.8)	(1055)	(2046)	(1990.8)
Supply: $Q^S_r * (1 - elast_{Sr} \log(wage_r))$	Part-time	-233.9	-188.89	73.753	37.158
		(96.526)	(88.931)	(83.665)	(82.803)
	Female	125.37**	108.46**	43.53	35.328
		(54.762)	(49.78)	(43.376)	(42.886)
	Urban	-45.979	-1.4138	16.818	-0.29226
		(41.559)	(35.742)	(29.754)	(33.344)
	Log population	-204.27***	-206.21***	-32.663	-24.654
		(17.321)	(17.383)	(17.389)	(16.987)
	Unemployment rate	14.599***	15.718***	8.5818	6.7736
		(6.1121)	(6.1678)	(7.2311)	(6.6734)
	Year 2013	-24.605	-25.78	-28.322	-26.286
		(17.504)	(17.449)	(19.745)	(18.606)
	Constant	2668***	2651.1***	372.65	286.08
		(222.96)	(222.13)	(221.72)	(215.92)

Bootstrapped standard errors in parentheses. $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$ from bootstrapped p-values. Elasticity of labor supply is directly observed from collected data, so for the supply equation is differenced off, leaving a dependent variable of the quantity after accounting for the demand through wages.

survey directly, are accurately measured. Demand elasticities, coming from our MLE, are relatively large although not statistically different from zero in any of the cases.

Table 5: Shortage equation estimated parameters

Shortage Indicator	γ_0	γ_1
1. Work in facility that prefers more ANs to cover current workload	0.472***	0.269
	(0.011)	(0.256)
2. Work in facility that could handle more cases if more ANs were hired	0.373***	-0.513
	(0.011)	(0.283)
3. Have increased hours in past 3 years	0.483***	0.310
	(0.011)	(0.287)
4. Would increase hours for sufficient increase in pay	0.668***	0.543***
	(0.017)	(0.043)
5: Average of 1-4	0.484***	0.276***
	(0.012)	(0.055)
6: Optimal-weighted average of 1-4	0.57502***	0.296***
	(0.016)	(0.053)

Bootstrapped standard errors in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$ from bootstrapped p-values

Table 6: Estimated parameters from models

Model and Shortage Indicator	Excess Demand 2007	Excess Demand 2013	Demand Elasticity	Supply Elasticity
MN (no shortage indicator)	-4883**	-2225**	-2.545	0.359***
	(4964.6)	(4313.6)	(1.744)	(0.0074)
SII1. Work in facility that prefers more ANs to cover current workload	-2956**	313.8	-3.261	0.359***
	(4728.1)	(4462.9)	(1.734)	(0.0074)
SII2. Work in facility that could handle more cases if more ANs were hired	-3347*	-7987.8*	-1.643	0.359***
	(6000.7)	(5667.9)	(1.771)	(0.0074)
SII3. Have increased hours in past 3 years	1585.2	-8586.9	-1.547	0.359***
	(34053)	(23310)	(1.676)	(0.0074)
SII4. Would increase hours for sufficient increase in pay	-17995***	-17052***	-2.401	0.359***
	(8434.5)	(8474.8)	(1.033)	(0.0074)
SII5: Average of 1-4	-4168.8*	-4689.9*	-2.840	0.359***
	(24874)	(24972)	(1.567)	(0.0074)
SII6: Optimal-weighted average of 1-4	-14752**	-15035**	-2.015	0.359***
	(27699)	(28416)	(1.008)	(0.0074)

Bootstrapped standard errors in parentheses. ***$p < .01$, **$p < .05$, *$p < .1$ from bootstrapped p-values

Figures 1 and 2 present the estimated shortage by HRR for the two models. The results are very similar, but do differ. The same general trends are present, but there is a lower level of surplus estimated and some sorting changes.

Figure 1: MN Estimated Expected Excess Demand for 2013

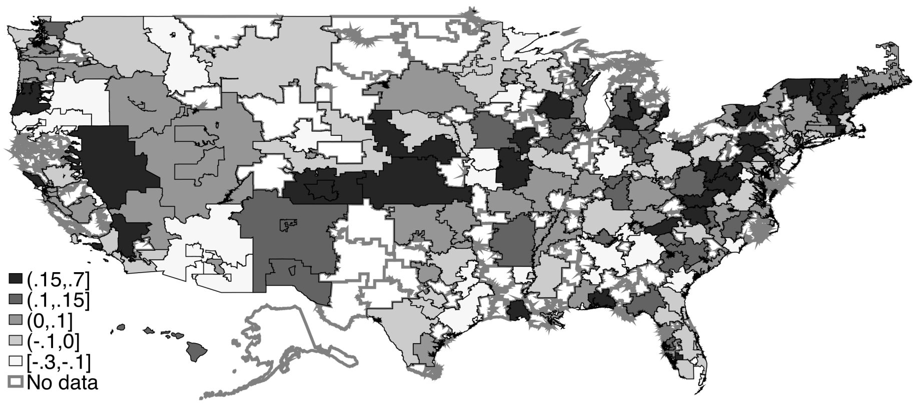
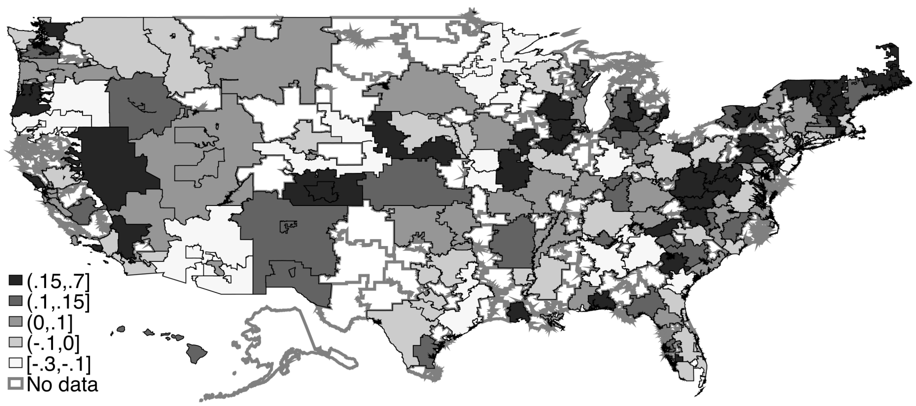


Figure 2: SII1 Estimated Expected Excess Demand for 2013



4.6 Post-Estimation Tests

In addition to comparing the coefficients and predictions of the model, we implement three different tests to compare the models after estimation. First, we do inference on whether the two models differ from each other empirically. We can do this in two ways. MN is a special case of SII where $\gamma_1 = 0$. In that case, the shortage indicator contains no additional information and the likelihood collapses to MN.[6] Thus, we can do a likelihood ratio test of the restricted (MN) and unrestricted (SII) models. Doing so yields a likelihood ratio χ^2 statistic above 100 in each case, as shown in Table 7. The critical value for the 1 percent significance level for the χ^2 distribution with one degree of freedom is 6.635; each of the SII models are significantly different than the MN model.

Table 7: Likelihood ratio test statistics comparing SII models to MN

Shortage Indicator	χ^2 value
1. Work in facility that prefers more ANs to cover current workload	135.4
2. Work in facility that could handle more cases if more ANs were hired	171.04
3. Have increased hours in past 3 years	161.5
4. Would increase hours for sufficient increase in pay	542.64
5: Average of 1-4	530.88
6: Optimal-weighted average of 1-4	739.64

Second, we estimate MN and the SII models only on 2007 data, and then use each model to predict what the labor demand and labor supply are given observables we see in 2013, and hence what the predicted labor quantity (as the minimum of predicted labor demand and predicted labor supply) is compared to the actual 2013 observed quantity. If SII estimates the supply and demand functions better, than we would expect better predictions. We estimate the average absolute bias as well as the Mean Square Prediction Error (MSPE) of the two predictions. Table 8 presents these results. SII1 has slightly worse mean absolute bias, but better MSPE than MN. However, the differences are small. All of the other SII models, with the exception of SII3, outperform MN in mean absolute bias.

[6] As a technical note, we additionally include in the likelihood the estimation of the mean and standard deviation of the shortage indicator to make the two comparable.

With regards to MSPE, all of the SII models outperform MN, except again in the case of SII3. SII5, the average of the shortage indicators, has the best MSPE across models, and has a lower mean absolute bias than the MN model, making it at least here our preferred shortage indicator version and overall preferred model.

Table 8: Comparisons of out of sample predictions

	MN	SII1	SII2	SII3	SII4	SII5	SII6
Mean Abs. Bias	95.14	95.24	93.48	101.81	79.77	89.32	76.24
MSPE	15,929	15,538	15,168	18,384	15,852	14,526	15,844

5 Conclusion

It is important to understand the extent of shortage or surplus in labor markets, such as in health provider labor markets, where shortage can lead to adverse health outcomes. Since the early model of Maddala and Nelson (1974), several disequilibrium models have been developed: in a switching regression model, in a Bayesian framework, and by allowing for a dynamic element where current supply and demand depend on prior supply and demand. Here, we develop a new disequilibrium estimation technique that uses shortage indicators as sources of additional information for shedding light on shortage or surplus in local labor markets. The shortage indicator informed disequilibrium model has an intuitive interpretation, wherein markets with higher (lower)-than-expected values of the shortage indicator put more weight on estimating observed quantity as labor supply (demand), and adjust the expected labor demand (supply) upwards and labor supply (demand) downwards, relative to the Maddala-Nelson model.

We apply the model to the labor market for anesthesiologists. The shortage indicator informed disequilbrium models estimated with different shortage indicators are each statistically different from the Maddala-Nelson model using a likelihood ratio test, which given the shortage indicator informed disequilibrium model contains as a special case the Maddala Nelson model implies that our model fits the data better than the Maddala-Nelson model. We also find better out-of-sample predictive power from the expanded models by estimating the parameters of the model using the 2007 survey and then predicting quantity supplied and demanded and predicted observed quantity, using the 2013 survey. Among the models, the unweighted average of the shortage indicators provides the best results in terms of out-of-sample predictions in terms of the MSPE.

There are also interesting by-products to our new approach, including estimated information about the shortage indicator such as its quantitative relationship with changes in shortage or surplus per capita, as well as what the equilibrium level of the shortage indicator is. This may be useful in many settings when analyzing labor markets for disequilibrium.

References

Abenstein, J. P., K. Long, B. McGlinch, and N. Dietz (2004), "Is physician anesthesia cost-effective?" Anesthesia and Analgesia, Vol. 98, 2004, pp. 750-757.

Baird, M., L. Daugherty, K. Kumar, and A. Arifkhanova (2015). "Regional and gender differences and trends in the anesthesiologist workforce." Anesthesiology, 123(5), 997-1012.

Baiker, K. and A. Chandra (2004). "Productivity of physician specialization: evidence from the Medicare." American Economic Review 94(2).

Bauwens, L. and M. Lubrano (2007). "Bayesian inference in dynamic disequilibrium models: an application to the Polish credit market." Econometric Reviews, 26(2-4), 469-486.

Bierstein, K. (2005). "Practice management: hospital contracts survey - 2004 data." American Society of Anesthesiologists Newsletter, 69(4).

Cromwell, J. (1999). "Barriers to achieving a cost-effective workforce mix: lessons from anesthesiology." Journal of Health Politics, Policy, and Law, 24(6).

Dall, M., P. Gallo, R. Chakrabarti, T. West, A. Semilla and M. Storm (2013). "An aging population and growing disease burden will require a large And specialized health care workforce by 2025." Health Affairs, 32, no.11 (2013).

Daugherty, L., R. Fonseca, K. B. Kumar, and P. Michaud (2010). "An analysis of the labor market for anesthesiology." RAND Health Technical Report.

Ginsburgh, V., A. Tishler, and I. Zang (1980). "Alternative estimation methods for two regime models." European Economic Review. 13:207-228.

Ginsburgh, V. and I. Zang (1975). "Price taking or price making behavior: An alternative to full cost price functions." Cowles Foundation discussion paper no. 403.

Gourieroux, C., Econometrics of Qualitative Dependent Variables, Cambridge University Press, 2000.

Gourieroux, C., J. Laffont, & A. Monfort (1980). "Disequilibrium econometrics in simultaneous equations systems." Econometrica, 75-96.

Hillman, A. (1987). "Financial incentives for physicians in HMOs. is there a conflict of interest?" New England Journal of Medicine, 317(27).

Hubbs, T. and T. Kuethe, (2017) "A disequilibrium evaluation of public intervention in agricultural credit markets." Agricultural Finance Review, Vol. 77 Issue: 1, pp.37-49.

Kalist, D., N. Molinari, and S. Spurr (2011). "Cooperation and conflict between very similar occupations: the case of anesthesia." Health Economics, Policy, and Law, 6(2).

Kane, M. and A. Smith (2004). "An American tale-professional conflicts in anesthesia in the United States: implications for the United Kingdom." Anesthesia, 59(8).

Laroque, G. and B. Salanie (1993). "Simulation based estimation of models with lagged latent variables." *Journal of Applied Econometrics*, 8(Suppl.):S119-S133.

Lee, L. (1997). "A smooth likelihood simulator for dynamic disequilibrium models." *Journal of Econometrics*, 78(2): 257-294.

Lubrano, M. (1985). "Bayesian analysis of switching regression models." *Journal of Econometrics*, 29(1-2), 69-95.

Lubrano, M. (1986). "Bayesian analysis of single market disequilibrium models: an application to unemployment in the US labour market." In Blundell, R., & Walker, I. (Eds.). (1986). Unemployment, search and labour supply. CUP Archive.

Maddala, G., and F. Nelson (1974). "Maximum likelihood methods for models of markets in disequilibrium." Econometrica, 42(6).

Madison, K. (2004). "Hospital-physician affiliations and patient treatments, expenditures, and outcomes." Health Services Research, 39(2).

Melly, B. and P. Puhani (2013). "Do public ownership and lack of competition matter for wages and employment? evidence from personnel records of a privatized firm." Journal of the European Economic Association, 11(4).

Phillips, R., M. Dodoo, and L. Green (2005). "Adding more specialists is not likely to improve population health: is anybody listening?" Health Affairs, W5-111.

Richard, J. (1980). "C-type distributions and disequilibrium models." Paper presented at the conference disequilibrium models held in Toulouse.

Robinson, J., S. Shortell, R. Li, L. Casalino, and T. Rundall (2004). "The alignment and blending of payment incentives within physician organizations." Health Services Research, 39(5).

Schubert, A., G. Eckhout, A. Ngo, K. Tremper, and M. Peterson (2012). "Status of the anesthesia workforce in 2011: evolution during the last decade and future outlook." Anesthesia Analgesia, 115(2).

Stulberg, R. and A. Shulman (2013). "Protecting physicians through employment contracts: A guide to the basic terms and conditions." New York State Bar Association Labor and Employment Law Journal, 38(1).

Online Appendix: Derivations

MN additional parameters derivations

After estimating the parameters, we can estimate the following additional functions of these parameters as follows:

The probability of a market having excess demand (shortage) is given by

$$\pi^{ED} = \Pr(Q^D > q) = q - \Phi\left(\frac{1 - X^D\beta^D}{\sigma_{\varepsilon D}}\right) \tag{A.1}$$

The expected quantity demanded is given by

$$\begin{aligned}
E[Q^D] &= E[Q^D|Q^D > Q^S]\Pr(Q^D > Q^S) + E[Q^D|Q^D < Q^S]\Pr(Q^D < Q^S) \\
&= E[X^D\beta^D + \varepsilon^D|X^D\beta^D + \varepsilon^D > q]\pi^{ED} + q(1 - \pi^{ED}) \\
&= \pi^{ED}\left(X^D\beta^D + \frac{\sigma_{\varepsilon D}\phi\left(\frac{q - X^D\beta^D}{\sigma_{\varepsilon D}}\right)}{1 - \Phi\left(\frac{q - X^D\beta^D}{\sigma_{\varepsilon D}}\right)}\right) + (1 - \pi^{ED})q
\end{aligned} \tag{A.2}$$

We similarly can calculate for expected quantity supplied (using the same equations, substitute S for D, and estimate the expected shortage in a market $E[Q^D_{mt} - Q^S_{mt}]$. This we can aggregate up to multiple markets by summing over all of the markets.

SII additional parameters derivations

Here we provide the derivation of the SII likelihood functions. We start by examining the elements of (8). Considering the elements $\Pr(Q^S = q|A = a)$ and $\Pr(Q^D = q|A = a)$. These can be expressed as, for $j = S$ or D,

$$\Pr(Q^j = q|A = a) = \Pr(X^j\beta^j + \varepsilon^j = q|h((Q^D - Q^S)/p) + \nu = a)) \tag{A.3}$$

We assume independence of the error terms. In that case, we can take advantage of the following relationship of a conditional normal distribution

$$Q^j|A \sim N\left(\mu_{Qj} + \frac{\sigma_{QjA}}{\sigma_A^2}(a - \mu_A), \sigma_{Qj}^2\left(1 - \frac{\sigma_{QjA}^2}{\sigma_{Qj}^2\sigma_A^2}\right)\right) \tag{A.4}$$

Where

$$\mu_{Qj} = X^j\beta^j \tag{A.5}$$

$$\sigma_{QjA} = Cov(X^j\beta^j + \varepsilon^j, h((X^D\beta^D + \varepsilon^D - X^S\beta^S - \varepsilon^S)/p)) \tag{A.6}$$

$$\sigma_A^2 = Var(h((X^D\beta^D - X^S\beta^S)/p)) + \sigma_\nu^2 \tag{A.7}$$

$$\mu_A = E[A|X^D, X^S, p] = E[h((X^D\beta^D - X^S\beta^S)/p)|X^D, X^S, p] \tag{A.8}$$

$$\sigma_{Qj}^2 = \sigma_{\varepsilon j}^2 \tag{A.9}$$

This suggests we can express the probability above as follows

$$\Pr(Q^j = q|A = a) = \phi\left(\frac{q - X^j\beta^j - \mu_{Cj}}{\sigma_{Cj}}\right)/\sigma_{Cj} \tag{A.10}$$

Where

$$\mu_{Cj} = \frac{\sigma_{QjA}}{\sigma_A^2}(a - E[A|X^D, X^S, p]) \tag{A.11}$$

$$\sigma_{Cj} = \sigma_{\varepsilon j}\left(1 - \frac{\sigma_{QjA}^2}{\sigma_{\varepsilon j}^2\sigma_A^2}\right)^{.5} \tag{A.12}$$

We also need to define $\Pr(Q^D > q|A = a)$ and $\Pr(Q^S > q|A = a)$. This is done in the same way as above, using the nature of the conditional normal distribution. Thus, the equation comes out to be

$$
\begin{aligned}
&[\Pr(Q^S = q|A = a) \Pr(Q^D > q|A = a) + \Pr(Q^D = q|A = a) \Pr(Q^S > q|A = a)] \Pr(A = a) \\
&= \left[\frac{1}{\sigma_{CS}} \phi \left(\frac{q - X^S\beta^S - \mu_{CS}}{\sigma_{CS}} \right) \left(1 - \Phi \left(\frac{q - X^D\beta^D - \mu_{CD}}{\sigma_{CD}} \right) \right) \right. \\
&\left. + \frac{1}{\sigma_{CD}} \phi \left(\frac{q - X^D\beta^D - \mu_{CD}}{\sigma_{CD}} \right) \left(1 - \Phi \left(\frac{q - X^S\beta^S - \mu_{CS}}{\sigma_{CS}} \right) \right) \right] \\
&\times \phi \left(\frac{a - h((X^D\beta^D - X^S\beta^S)/p)}{\sigma_A} \right) / \sigma_A \qquad \text{(A.13)}
\end{aligned}
$$

For the case of $h(\cdot)$ being a linear function, as given in equation (5), we have

$$
\mu_{Cj} = \frac{-1^{j==S}\gamma_1\sigma^2_{\varepsilon j}/p}{\gamma_1^2(\sigma^2_{\varepsilon S} + \sigma^2_{\varepsilon D})/p^2 + \sigma^2_\nu} \left(a - \gamma_0 - \gamma_1(X^D\beta^D - X^S\beta^S)/p \right) \qquad \text{(A.14)}
$$

$$
\sigma^2_{Cj} = \sigma^2_{\varepsilon j} \left(1 + \frac{\gamma_1^2\sigma^2_{\varepsilon j}/p}{\gamma_1^2(\sigma^2_{\varepsilon S} + \sigma^2_{\varepsilon D})/p^2 + \sigma^2_\nu} \right) \qquad \text{(A.15)}
$$

Next, we want to solve what the expectations of supply and demand are, i.e. $E[Q^D|q, x^D, x^S, A]$ and $E[Q^D|q, x^D, x^S, A]$. Using the same identity expressed in equation (A.4), we have

$$
E[Q^D|q, x^D, x^S, A] = E[Q^D|q, x^D, x^S] + \frac{\sigma_{Q^D,A}}{\sigma^2_A}(a - E[A|q, x^D, x^S]) \qquad \text{(A.16)}
$$

$$
E[Q^S|q, x^D, x^S, A] = E[Q^S|q, x^D, x^S] - \frac{\sigma_{Q^S,A}}{\sigma^2_A}(a - E[A|q, x^D, x^S]) \qquad \text{(A.17)}
$$

$E[Q^j|q, x^D, x^S]$ we have already derived, as given in equation (A.2), and the other parameters are as defined above.

If we assume linear function for $h(\cdot)$, as given in equation (5), then using the law of total covariance we further have (taking as an example Q^D; the case for Q^S follows similar steps)

$$
\sigma_{QDA} = Cov(Q^D, A|q, x^D, s^S) = E[Cov(Q^D, A|q, x^D, x^S)] + Cov(E[Q^D|q, x^D, x^S], E[A|q, x^D, x^S]|q, x^D, x^S) \qquad \text{(A.18)}
$$

The first term becomes

$$E[Cov(Q^D, A|q, x^D, x^S)|Q^D > Q^S]\Pr(Q^D > Q^S) + E[Cov(Q^D, A|q, x^D, x^S)|Q^D < Q^S]\Pr(Q^D < Q^S)$$

$$= E[Cov(X^D\beta^D + \varepsilon^D, \gamma_0 + \gamma_1(X^D\beta^D + \varepsilon^D - X^S\beta^S - \varepsilon^S)/p|q, x^D, x^S)|Q^D > Q^S]\pi^{ED}$$

$$+ E[Cov(q, \gamma_0 + \gamma_1(X^D\beta^D + \varepsilon^D - X^S\beta^S - \varepsilon^S)/p|q, x^D, x^S)|Q^D < Q^S](1 - \pi^{ED})$$

$$= \gamma_1^2\sigma_{\varepsilon D}^2\pi^{ED}/p^2 \tag{A.19}$$

The second term is equal to zero. Substituting in, we have

$$E[Q^D|q, x^D, x^S, A] = E[Q^D|q, x^D, x^S] + \frac{\gamma_1^2\sigma_{\varepsilon D}^2/p^2}{\gamma_1^2(\sigma_{\varepsilon D}^2 + \sigma_{\varepsilon S}^2)/p^2 + \sigma_\nu^2}\pi^{ED}(a - (\gamma_0 + \gamma_1(X^D\beta^D - X^S\beta^S)/p)) \tag{A.20}$$