Empir Econ (2010) 39:1–25
DOI 10.1007/s00181-009-0292-8

ORIGINAL PAPER

Estimation of hedonic price functions with incomplete information

Subal C. Kumbhakar · Christopher F. Parmeter

Received: 3 February 2008 / Accepted: 2 April 2009 / Published online: 1 May 2009

Abstract Existence of persistent price dispersion suggests that some buyers find lower prices through search and information acquisition, while some sellers charge higher prices by gathering information on potential buyers. If buyers are not fully informed of the lowest price available in the market they end up paying a price higher than if they had full information. Similarly, if sellers are not fully informed about the highest price they could charge, they too suffer by receiving a price lower than had they had full information. This paper develops a hedonic price model that incorporates the effects of incomplete information on both sides of the market and obtains estimates of the discrepancies between market prices and buyers' maximum willingness to pay and sellers' minimum willingness to accept. Correlates of such price discrepancies are also explored. We apply the technique to a data set constructed from the American Housing Survey, and find that incomplete information has had a significant impact on housing prices.

Keywords Hedonic price model · Information tax · Information deficiency · Two-tier frontier

JEL Classification C21 · D82 · D83

The authors thank Casey Dawkins, Daniel J. Henderson and two anonymous referees, as well as participants at the first annual New York Camp Econometrics, the NAPW 2006 and the APPC 2006 for helpful comments.

S. C. Kumbhakar (✉)
Department of Economics, Binghamton University, Binghamton, NY 13902-6000, USA
e-mail: kkar@binghamton.edu

C. F. Parmeter
Department of Ag and Applied Economics, Virginia Tech, Blacksburg, VA 24061-0401, USA
e-mail: parms@vt.edu

1 Introduction

In his seminal paper, Stigler (1961) attributed price variation to two overarching components: heterogeneity and ignorance (lack of information). As an example, consider the online textbook market. Different sellers sell the same textbook on the internet at various prices. The product being sold is identical and yet different prices are observed. Thus, a buyer pays a higher price (or alternatively a seller receives a lower price) for a homogeneous product due to his/her ignorance. Moving to the traditional hedonic price literature, markets are sufficiently "thick" so as to eliminate price dispersion for identical goods. However, as a good becomes increasingly unique, markets become "thin" and the establishment of a single price (given characteristics) for the product does not occur. Thus, prices may vary due to heterogeneity of goods and/or ignorance on the part of the buyer/seller. The uniqueness of goods becomes less of a reason for price dispersion as more characteristics are taken into account. As a result, ignorance becomes the prime factor causing price variations. Whatever its cause, price dispersion can be taken as an empirical regularity.

When purchasing goods, buyers face search costs in obtaining price information from different sellers. These search costs lead the buyer to search optimally in the sense that the marginal gain from another search is equal to its marginal cost (shadow price of search). If buyers knew of the seller(s) with the lowest willingness to accept (WTA), as would be the case if the market was located at one central location and perfectly competitive, there would be no price variation except for the inherent randomness of the market and seller heterogeneity. This claim may be disputed by considering the study of Pratt et al. (1979) who find significant price dispersion, in terms of the coefficient of variation, for 39 goods in what many would consider competitive markets. More recent studies such as Sorensen (2000) and Lach (2002) find price dispersion in "brick and mortar" retail markets while Häring (2003) and Cooper (2006) find price dispersion in online markets. On the other hand, if sellers knew the buyer(s) with the highest maximum willingness to pay (WTP), as is true for a first degree price discriminating monopolist, there would be no variation in prices except for noise in the data and heterogeneity among buyers.[1]

However, neither of the situations described above are common or even likely in today's marketplace. What is common is the existence of informational conduits through which both buyers and sellers can obtain information; the classified ads and the internet are but a few. While there is no way of knowing the highest WTP or the lowest WTA without some explicit advertisement of those prices, both buyers and sellers can gain information through search. However, given that search costs exists, market participants most likely will not become fully informed and price variations due to ignorance will exist even after controlling for product characteristics.

The lack of information on behalf of both buyers and sellers is suggestive that those with more information may gain an advantage in the differentiated goods market. Thus, these markets may be inefficient for certain types of buyers and sellers. In the housing market, researchers have found that this is not the case across either

[1] That is, buyers are not expected to have the same maximum WTP for the same product and so prices should vary across buyers.

buyers (Turnbull and Sirmans 1993) or sellers (Turnbull et al. 1990). However, few studies have looked at the advantage a buyer may possess over a seller or vice-versa simultaneously; for some insight into this issue (see Gaynor and Polachek 1994) for an analysis of the market for physicians services and Harding, Rosenthal, and Sirmans (2003) HRS hereafter, for the housing market when bargaining takes place. Our results suggest that there does not seem to be a group, either buyer or seller, that has an advantage in the housing market, lending credibility to the idea of an efficient market from the standpoint of information across agents.

The extent of the price reductions obtained by buyers (from sellers) and the price increases received by sellers (from buyers) depends upon how well informed they are. Above, the lack of information was termed ignorance. However, it may better our discussion to use a milder terminology to characterize agents who possess less than full information. Here, we call buyers and sellers who lack full information, deficient. As opposed to ignorance, deficient means that buyers and sellers may want to gather more information through search, but further search is costly and time consuming, and so the incomplete information that each possesses is sufficient to enter the market.[2]

The contributions of this paper are three-fold. First, we graphically formulate the classic hedonic price model to account for incomplete information on both sides of the market. After this we show how to formally estimate this model by placing the hedonic equilibrium in a two-tier stochastic frontier (2TSF) framework (see Polachek and Yoon 1987, 1996). We borrow the decomposition technique proposed by Kumbhakar and Parmeter (2009) to measure the impact of buyers' and sellers' information deficiency on observed market prices. Second, we generalize the method to allow information deficiency to depend on buyer and seller characteristics.[3] Third, we apply these techniques to estimate a hedonic price function for a detailed data set on houses and examine the effects that home owner/buyer characteristics, such as being a first time buyer, being from out of town, having kids, and the like, have on housing prices. Since the impact of information deficiency on observed prices can be computed for each transaction, one can use these estimates to discern which types of buyers (sellers) are the most deficient, on average, in terms of paying (receiving) higher (lower) prices.

Our empirical results are based on data from the American Housing Survey (AHS) which records standard characteristics of a home such as square footage and number of bathrooms, but also has detailed information on the homeowners themselves. By comparing information of the home owner prior to and directly after the house sold we can obtain data on the characteristics of the buyers and sellers and then use this to parse out insight into the impact that these characteristics have on the effect of incomplete information in the housing market. Our results suggest that both buyers and sellers in the AHS are information deficient. This is not surprising given that our data come from the period 1986 to 1993 which represents a time when gathering information was costlier than it is today with the advent of the internet and the creation of multiple listing services.

[2] We thank Solomon Polachek for suggesting this taxonomy to us.

[3] The classic "determinants of efficiency" approach can be found in the standard stochastic frontier literature (Kumbhakar and Lovell 2000).

The remainder of the paper is organized as follows. Section 2 discusses the hedonic price model and demonstrates graphically how incomplete information effects prices. In Sect. 3 we show how to incorporate incomplete information into a hedonic model (in a reduced form setting), which fits into the 2TSF model. Section 4 reviews the 2TSF model, focuses on how to construct measures of price efficiency (and the cost of information deficiency) for both buyers and sellers, and extends the model to allow for determinants of information deficiency. Section 5 presents the data and some results for the housing market while Sect. 6 ties our analysis in with previous studies of market pricing efficiency. A summary of results and conclusions are given in Sect. 7.

2 Incomplete information and hedonic price functions

Here we re-evaluate the hedonic price setup popularized by Griliches (1971) and Rosen (1974). In these models, a good is composed of a set of characteristics. The affine transformation between these characteristics and their implicit values (known as shadow prices) determine the market value of the good. The locus of market values traces out the hedonic price function. In what follows we consider the Rosen (1974) theoretical model.

Bid functions are constructed to discern the price a buyer would pay for a good with certain attributes for a fixed level of utility and income. It indicates the highest price a buyer is willing to pay for the good. Utility is maximized when the bid function is equal and tangent to the market price function, the minimum price a buyer must pay in the market. Similarly, firms have offer functions for the good in question, given a certain level of profit. The offer functions indicate the lowest price they will accept for the product in question. If the market is competitive, sales take place where an individual seller's offer function "kisses" an individual buyer's bid function. In other words, the locus of the points of tangency between the offer functions $\varphi(z;\pi)$ (for different levels of profit, π) and bid functions $\psi(z; y, U)$ (for different levels of income, y and utility, U) traces the hedonic price function $P(z)$, of the good with scalar characteristic z. This is shown in Fig. 1.

However, this model implicitly assumes that there is full information in the market, as well as perfect competition, implying that bid and offer functions are tangent.[4] In this setup there would be no rationale for price variation except that which is due to inherent heterogeneity of the good.[5] With full information, every buyer knows of the lowest WTA and every seller knows of the highest WTP. This forces the market to generate a unique price for each z in the market. However, with incomplete information the market price is affected by the levels of information that buyers and sellers possess specifically because there is no reason for all bid functions to be alike and all offer functions to be the same at a given point. That is, for a given z there are numerous bid and offer functions due to differing levels of information. The multiplicity of bid and offer functions for a specific characteristic results in a gap between the highest bidding buyer and the lowest offering seller. This gap is the source of price

[4] It also assumes that all bid (offer) functions at a given point are exactly the same for all buyers (sellers).

[5] This would be the error term from the regression of prices on the characteristics.

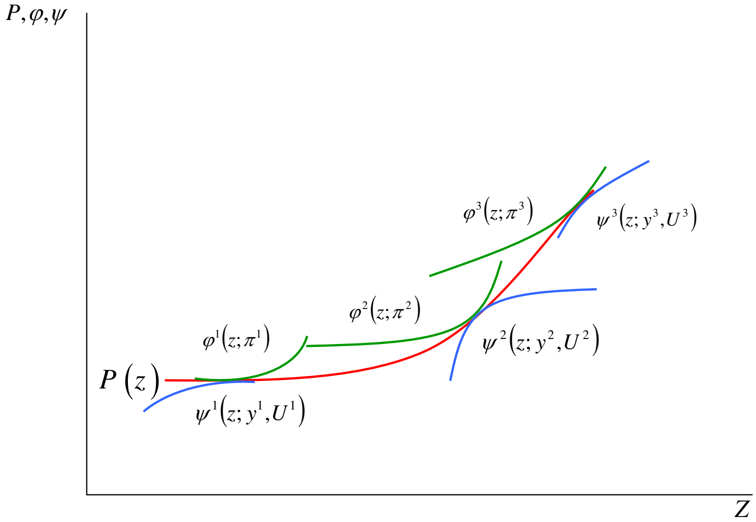

Fig. 1 Hedonic price equilibrium with full information

variations that are observed in the market, after controlling for product characteristics. Equilibrium occurs when a buyer-seller pair align themselves such that their levels of information translate into tangency between their bid and offer functions as in the traditional hedonic setup with full information.[6]

The above account of how information affects price variation is detailed below. For clarity we describe each side of the market separately before continuing on to equilibrium.

A hedonic market with incomplete information on the buyers' side is shown in Fig. 2. Here we have a hedonic price frontier for buyers, $P_b(z)$, that is defined as the upper envelope for the bid functions ψ^1, ψ^2, and ψ^3, similar to Rosen's framework. However, there are other buyers in the market that are looking to purchase the good at lower bids, ψ^4, ψ^5, and ψ^6. In a world of perfect competition and full information these lower bids are not relevant to sellers as they know of bids ψ^1, ψ^2, and ψ^3. It is this upper envelope that sellers with full information would align themselves with.

The situation on the sellers' side of the market is similar and is depicted in Fig. 3. Here we have a hedonic price frontier for sellers, $P_s(z)$, that is defined as the lower envelope for the offer functions φ^1, φ^2, and φ^3, and is, again, similar to Rosen's framework. However, there are other sellers in the market that are looking to provide the good at lower offers, φ^4, φ^5, and φ^6. In a world of perfect competition and full information these lower offers are not relevant to buyers as they know of offers φ^1, φ^2, and φ^3. In Rosen's model we would have $P_b(z) = P_s(z)$ and price fluctuations would only exist due to heterogeneity. However, with incomplete information, $P_b(z) \neq P_s(z)$ and a more detailed analysis is required to understand what is happening in the differentiated goods market.

[6] It is as though information is being directly incorporated into the utility and profit functions.

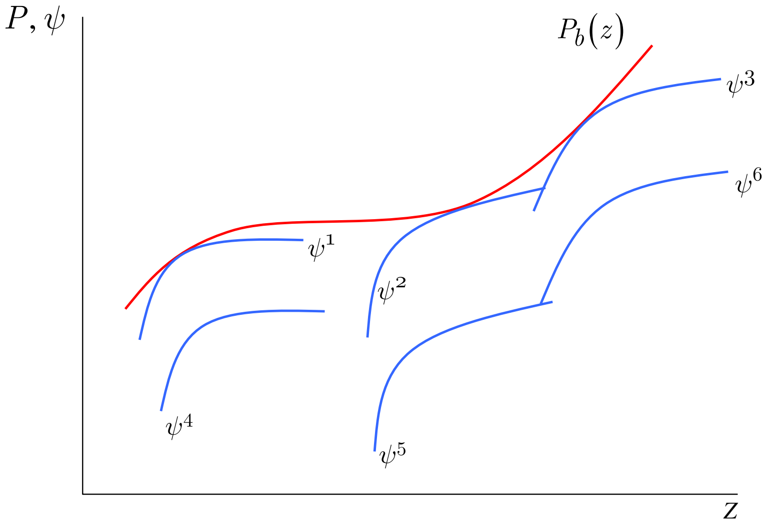

Fig. 2 Sellers' viewpoint of the buyers' side of the market with incomplete information

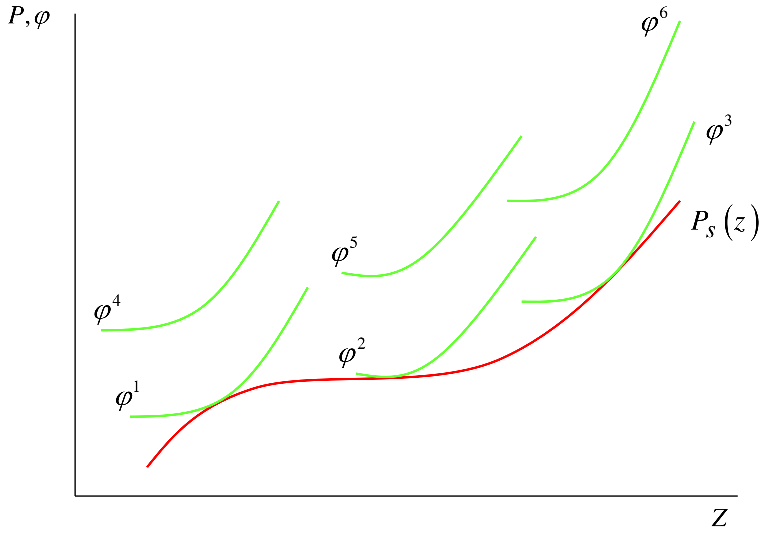

Fig. 3 Buyers' viewpoint of the sellers' side of the market with incomplete information

Equilibrium in the differentiated goods market with incomplete information is depicted in Fig. 4. Figure 4 captures the essence of Figs. 2 and 3 by placing them on top of one another. Since it is assumed that buyers and sellers search optimally, tangency of bid and offer functions occurs where the marginal cost of searching, along with the marginal bid and offers, are equivalent for both parties. Three such tangencies are shown for product levels z', z'', and z'''.

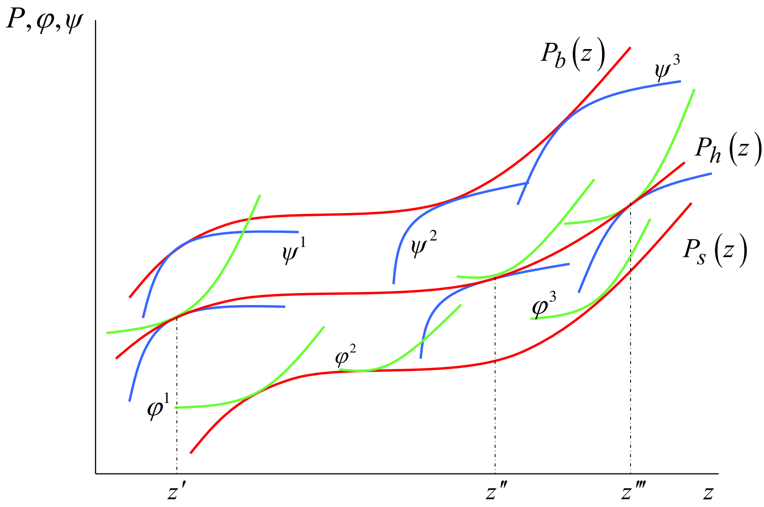

Fig. 4 Structure of a differentiated good market when incomplete information exists

It can be seen from this representation that instead of having a unique price for a given level of a characteristic (as in Fig. 1), there is a price surface that lies between the highest WTP (bid curve) and the lowest WTA (offer curve). Where a point of sale originates in this space is dictated by the levels of information possessed by both parties. An equilibrium is still achieved though as tangency between bid and offer functions occurs where marginal costs of information arise.[7] The observed market price will depend on the characteristics of the good, the levels of information possessed by both the buyer and seller, as well as any inherent randomness (unexplained product heterogeneity) that may be present. It should be noted at this time that we are not specifically categorizing the sources of the incomplete information. It could be attributed to the actual characteristics of the house, characteristics of the seller, characteristics of the buyer, or a combination of the three.

What do these price frontiers, $P_b(z)$ and $P_s(z)$, imply about the market equilibrium price that is observed? $P_b(z)$ indicates that equilibrium prices cannot be higher than this function, otherwise no sale will result. Sellers wish to charge prices corresponding to this upper envelope. However, incomplete information prevents sellers from doing so. Additional information about potential buyers can be obtained through market research and further interactions, which may lead to an increase in the prices received. The same story holds on the buyers' side. Here buyers attempt to find the "lowest prices" available in the market, i.e., prices corresponding to points along $P_s(z)$. The presence of search costs limits their ability to determine the lowest price for the product in the market. Buyers get additional information about market prices by soliciting prices from more sellers.

[7] This is very different from the setup of HRS who also assume a gap between bid and offer functions, but there is no rationale for an equilibrium to occur as bid and offer functions never become tangent to one another.

At this point it is natural to ask if buyers' know sellers' willingness to accept given that asking prices are printed in newspapers and posted on multiple listing services. One could use these posted prices to develop a hedonic model that incorporates them directly into the analysis. We avoid that here for several reasons. First, since one of our main interests lies with understanding how a home buyer or seller's characteristics influence the level of information deficiency they face we use a dataset that provides access to various buyer and seller attributes.[8] The use of this data comes at a cost since it does not provide the exact location of a house which makes tracking down the list price of the house next to impossible. Additionally, it is highly unlikely that list prices correspond to sellers' willingness to accept and as such use of list prices is likely to induce issues with measurement error. However, this is a fruitful avenue for research that takes a different direction than that proposed here and we draw attention to its potential for future hedonic research.

If one were to simply consider the hedonic price function without taking account of the highest and lowest prices just discussed, nothing can be said about the impact of the level of information that any particular buyer or seller possesses on the observed price. Thus, an estimation strategy that incorporates information deficiencies of agents would seem appropriate. In the following section we show how the 2TSF approach can allow one to take advantage of the fact that incomplete information exists in the market.

3 Hedonic models with incomplete information

Information and its impact are, as is clear from the previous discussion, important to take account of in a hedonic analysis. What we present here is not a formal hedonic model accounting for incomplete information. Instead, this is an empirical approach that will allow us to think about studying a formal hedonic model, such as Rosen (1974), with incomplete information where our main interest is how we go about identifying the difference between full-information and incomplete-information prices. Our methodology will allow us to estimate a model that is consistent with an incomplete information set up in the housing market.

Consider the standard hedonic price equation that is typically specified as

$$P_h = h(z) + v, \tag{1}$$

where P_h is the hedonic price (or the logarithm of the hedonic price), z is a vector of product characteristics influencing the overall value of the good in the market, and v represents random noise and measurement error in price. The hedonic price function in (1) corresponds to the full information model developed by Rosen (1974). Nowhere in this setup are the buyers with the highest WTPs or the sellers with the lowest WTAs accounted for.

[8] See Parmeter and Pope (2009) for more on the different type of housing data sets that are available and their pros and cons for use in an empirical analysis.

If we recast the hedonic price model to take account of these price bounds then from a seller's standpoint the price received (P_m^s) can be represented as

$$P_m^s = P_b - u, \tag{2}$$

where P_b represents the highest WTP in the market and $u \geq 0$ can be viewed as the loss to a seller for being information deficient. From a buyer's point of view the price paid (P_m^b) can be written as

$$P_m^b = P_s + w, \tag{3}$$

where P_s represents the lowest WTA in the market and $w \geq 0$ represents the cost the buyer bares for being information deficient. While the previous modeling approach may seem ad hoc, conceptually, a structural model that incorporates information into the budget constraints of buyers and the cost(profit) functions of sellers, should lead to a similar framework. One could think of incorporating information directly into the original theoretical hedonic model of Rosen (1974) to derive the same set of conditions found there.[9]

Given that market prices are bounded from above and below, we need a technique that takes these bounds into account and provides estimates concerning the effects of buyer and seller information deficiencies. In the 2TSF approach, the effects of incomplete information are captured because the market price is not identical to the hedonic price of the good. The observed market price must be adjusted for the incomplete information that is present at the time of the transaction. However, given that the levels of information possessed by buyers and sellers are unknown, an alternative strategy is needed in order to correctly interpret the problem at hand. We incorporate the effects of incomplete information into the standard hedonic price framework as follows.

For a market transaction to occur we require that the market price, P_m equals both the price paid by a buyer and the price received by a seller, $P_m^b = P_m^s \equiv P_m$, which yields the following equality

$$P_s + w = P_m = P_b - u. \tag{4}$$

We exchange u and w across the equality which gives,

$$P_s + u = P_m + u - w = P_b - w. \tag{5}$$

The left-hand side can be viewed as the adjusted hedonic price of the seller while the right-hand side can be viewed as the adjusted hedonic price of the buyer. These adjusted hedonic prices account for information that is possessed by market participants at the

[9] That is, instead of assuming full information for buyers and sellers, information is now a characteristic with an explicit "cost" for buyers and sellers. Showing that an equilibrium exists in this setting would be an interesting contribution to the literature.

time of an exchange. Given that u, w, P_s, and P_b are all unobserved, estimating either the LHS or the RHS of (5) is infeasible without additional assumptions.

To obtain a regression equation from (5), we set the hedonic price of the good, $h(z) + v$, equal to the middle term of (5). Thus, we can express prices observed in the market as:

$$h(z) + v = \underbrace{P_m + u - w}_{\text{Full information price}} \Rightarrow P_m = h(z) + v + w - u. \tag{6}$$

Equation (6) shows that the market price of a good is constituted of (i) the implied value of the characteristics $h(z)$, (ii) inherent heterogeneity and random noise (v), and (iii) the costs of incomplete information to the buyers (w) and sellers (u).[10] With this representation one can find not only marginal values of attributes, but the impact of incomplete information on prices (labeled as the loss associated with information deficiency). The market price should be the same as the hedonic price (aside from v) when (i) there is no information deficiency on the part of either buyers or sellers, or, (ii) the cost of buyers incomplete information completely offsets the cost of sellers incomplete information (i.e., $w - u = 0$). The standard hedonic price models take the market price as the true valuation of the good. In doing so the asymmetry in the costs of the incomplete information possessed by the buyers and sellers is discarded and consequently the only explanation of price dispersion is heterogeneity.[11] Our model incorporates information deficiency from both sides of the market, as suggested by Rothschild (1973), which should help characterize the nature of price dispersion better than the standard model that assumes away information deficiency.

Collecting the three separate error components in (6), we write the estimating equation as,

$$P_m = h(z) + v + w - u = h(z) + \varepsilon. \tag{7}$$

Here $\varepsilon = v + w - u$ is a three component composite error term.

The parameters of $h(z)$ in (7) can be obtained from a simple regression once a parametric form of $h(z)$ is chosen. Since u and w are one-sided, $E(\varepsilon)$ may not be zero, even if $E(v) = 0$. Consequently, the OLS estimate of the intercept will be biased.[12] The OLS procedure will give unbiased and consistent estimates of the slope coefficients so long as the error components are uncorrelated with any of the goods characteristics. Direct inspection reveals that if one is interested in discerning the presence and effect of incomplete information on prices then estimation of Eq. (1) is inappropriate for this purpose. One needs to estimate a model that has the impacts of incomplete information built in. This is what the model in Eq. (7) is suited for. If the factors influencing

[10] Or, alternatively, that the market price needs to be adjusted for consideration in a traditional hedonic setting.

[11] Indeed, as pointed out by HRS, the fact that housing prices display seasonal variation suggests a weakness of the standard hedonic price model for the housing market.

[12] Note that although $E(u)$ and $E(w)$ are non-zero, $E(w - u)$ might be zero. If this happens then the OLS estimator of the intercept will also be unbiased.

a home's price, such as structural characteristics and environmental amenities, are uncorrelated with those factors influencing the level of information deficiency, such as income and being from out of town, then the presence of informational deficiencies should not impact the marginal effects that are typically at the center of any empirical hedonic analysis. Thus, the marginal effects are not likely to differ depending on whether one estimates the model in Eq. (1) or (7).[13] Since our objective in this paper is to estimate both the marginal effects *and* the impact of incomplete information, we have to use the model in Eq. (7).

4 The two-tier frontier method

4.1 The log likelihood function

Standard (single-tier) stochastic frontier models include either u (for production, revenue, and profit functions) or w (for a cost function) depending on the assumed behavior of firms. In the present case, we have two frontiers, viz., an upper frontier P_b, which represents the highest prices that buyers are willing to pay, and a lower frontier P_s, which represents the lowest prices that sellers are willing to accept. These two frontiers are imbedded in (7).

To estimate (7) we use the maximum likelihood (ML) method. The ML method we propose is based on the following distributional assumptions of the error components, viz., v, u and w: (i) $v_i \sim$ i.i.d. $N(0, \sigma_v^2)$, (ii) $u_i \sim$ i.i.d. $\text{Exp}(\sigma_u, \sigma_u^2)$[14] (iii) $w_i \sim$ i.i.d. $\text{Exp}(\sigma_w, \sigma_w^2)$, along with the assumption that each of these error components is distributed independently of one another and from each of the regressors. The exponential distributions for u and w capture the fact that the probability of low costs of incomplete information for buyers and sellers are high (the area near zero values of u and w being the highest), which one would expect if markets are close to competitive.

Based on the above distributional assumptions, it is straightforward (but tedious) to derive the pdf of ε_i, $f(\varepsilon_i)$, which is[15]

$$f(\varepsilon_i) = \frac{\exp\{\alpha_i\}}{\sigma_u + \sigma_w}\Phi(\beta_i) + \frac{\exp\{a_i\}}{\sigma_u + \sigma_w}\Phi(b_i), \tag{8}$$

where $a_i = \frac{\sigma_v^2}{2\sigma_w^2} - \frac{\varepsilon_i}{\sigma_w}$, $b_i = \frac{\varepsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_w}$, $\alpha_i = \frac{\varepsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}$, $\beta_i = -\left(\frac{\varepsilon_i}{\sigma_v} + \frac{\sigma_v}{\sigma_u}\right)$ and $\Phi(.)$ is the cumulative distribution function of the standard normal variable. The likelihood function for a sample of n observations is the product of the $f(\varepsilon_i)$ in (8), $\prod_{i=1}^{n} f(\varepsilon_i)$.

---

13 This might not be the case if the uncorrelatedness assumption is dropped.

14 Here $\text{Exp}(\sigma_z, \sigma_z^2)$ denotes a random variable z that is exponentially distributed with mean σ_z and variance σ_z^2.

15 See Kumbhakar and Parmeter (2009) for a formal derivation.

The log likelihood function for a sample of n observations identically distributed according to (8) is given as

$$\ln L\,(x;\theta) = -n\ln\left(\sigma_u + \sigma_w\right) + \sum_{i=1}^{n} \ln\left[e^{\alpha_i}\Phi(\beta_i) + e^{a_i}\Phi(b_i)\right], \tag{9}$$

where $\theta = \{\delta, \sigma_v, \sigma_u, \sigma_w\}$ and δ represents the parameters associated with the hedonic function $h(z)$. The estimates of all the parameters can be obtained by maximizing the log likelihood function. It should be noted that identification of σ_u and σ_w is achieved due to the fact that they appear in the likelihood equation in distinct parts. We use the exponential–exponential–normal set up here for tractability, although one could use distributions such as truncated normal or gamma, as is done in the traditional single-tier frontier models. The gamma–gamma–normal setup would provide a useful robustness check, even though a closed form solution does not exist. Tsionas (2006) proposed several techniques for estimating a gamma–gamma–normal 2TSF. His results suggest that the differences in the parameter estimates do not vary substantially suggesting that the choice of distributions do not have a large effect on the qualitative results for his empirical analysis. In single tier stochastic frontier models the consensus is that the distributional assumptions do not change the qualitative results substantially.[16]

It is well known in the single-tier stochastic frontier literature that the choice of distribution has a noticeable effect on the raw estimates but not on the qualitative implications (see page 90 of Kumbhakar and Lovell 2000). Here we suggest prudence in the raw interpretation of results since we are calculating conditional means. These conditional means will change as the distribution and the associated parameters change. However, if one was interested in the impact of a specific characteristic on the conditional mean of our information deficiency distributions, income of buyers say, then it is likely that the insights we draw will not vary from distribution to distribution. Still it would prove useful as future research for the derivation of 2TSF likelihood equations under a variety of distributional assumptions, such as truncated normal of half-normal, to quantify our insights.

4.2 Buyer's and seller's costs of incomplete information

The primary goal of estimating a stochastic frontier function is to obtain estimates of observation-specific inefficiency. In the present case we wish to determine by how much more a buyer pays and how much less a seller receives for having incomplete information.[17] To explain these concepts we reconsider Eqs. (2) and (3), and note that a buyer's price efficiency is the ratio of the lowest WTA to the observed price. That is, if prices are expressed in logarithmic form,

$$\mathrm{PE}_{\mathrm{Buyer}_i} = \mathrm{WTA}_i/\text{observed price}_i = \exp\{-w_i\}, \tag{10}$$

[16] Since there are no empirical/Monte Carlo studies on the two-tier models, we advise the readers to view the empirical results with caution.

[17] Note that these are all potential in the sense that they are based on WTP and WTA.

while a seller's price efficiency is the ratio of the observed price to the highest WTP and is represented by

$$\mathrm{PE}_{\mathrm{Seller}_i} = \text{observed price}_i/\mathrm{WTP}_i = \exp\{-u_i\}. \tag{11}$$

Following the single-tier frontier approach (Jondrow et al. 1982), we estimate (10) and (11) using their conditional means, viz., $E\left(e^{-u_i} \mid \varepsilon_i\right)$ and $E\left(e^{-w_i} \mid \varepsilon_i\right)$. These formulae are[18]

$$E\left(e^{-u_i} \mid \varepsilon_i\right) = \frac{\lambda}{1+\lambda}\frac{1}{\chi_{2i}}\left[\Phi\left(b_i\right) + \exp\left\{\alpha_i - a_i\right\}\exp\left\{\sigma_v^2/2 - \sigma_v\beta_i\right\}\Phi\left(\beta_i - \sigma_v\right)\right] \tag{12}$$

and

$$E\left(e^{-w_i} \mid \varepsilon_i\right) = \frac{\lambda}{1+\lambda}\frac{1}{\chi_{1i}}\left[\Phi\left(\beta_i\right) + \exp\left\{a_i - \alpha_i\right\}\exp\left\{b_i\sigma_v - \sigma_v^2/2\right\}\Phi\left(b_i + \sigma_v\right)\right]. \tag{13}$$

To estimate the percentage cost (in terms of over- or underpayment) of information deficiency for a particular buyer or seller, we compute $1 - \mathrm{PE}_{\mathrm{Buyer}_i} = 1 - e^{-w_i}$ and $1 - \mathrm{PE}_{\mathrm{Seller}_i} = 1 - e^{-u_i}$, using the above formulae for e^{-w_i} and e^{-u_i}. Note that if a buyer (seller) were fully informed he/she would find the lowest price regardless of sellers' (buyers') incomplete information. One can interpret w (u) (when multiplied by 100) as the percentage increase (decrease) in price that a buyer (seller) pays (receives), especially when w (u) is small.

4.3 Heterogeneity in buyer's and seller's deficiency

So far we have assumed that the distribution of w (u) is identical for all buyers (sellers). Since the costs associated with incomplete information are likely to differ across buyers and sellers, the 2TSF model has to be extended to incorporate the possibility of systematic differences in w and u. In view of this, our objective in this section is to allow the parameters of the distribution of w (u) to depend on buyers' (sellers') characteristics, so that the loss associated with possessing incomplete information can differ systematically. For example, if it is believed that first time home buyers have less information than a repeat home buyer (Turnbull and Sirmans 1993), one incorporates this information directly into the hedonic price function by adding a dummy variable and performs a statistical test. In our framework, we can go further and assert that the distribution of w is different for repeat buyers and first time buyers. This can be done by allowing the mean (standard deviation) of w (u) to depend on buyers' (sellers') characteristics.

To allow for the possibility that a vector of exogenous variables can influence the cost of incomplete information, we allow the means of w and u to be functions of buyers' and sellers' characteristics (z_w and z_u variables), respectively. Thus, we specify σ_w and σ_u as

[18] Derivation of these conditional means are given in Kumbhakar and Parmeter (2009, Appendix).

$$\sigma_w = e^{\delta_w' z_w} \quad \text{and} \quad \sigma_u = e^{\delta_u' z_u}, \tag{14}$$

where z_w and z_u are vector of buyers' and sellers' characteristics (including intercepts).[19] If we believe that a particular characteristic lowers (increases) the expected cost of incomplete information, such as being a repeat buyer, then we would expect the associated coefficient to be negative (positive). The standard likelihood ratio (LR) test can be used to test whether some of these characteristics influence σ_w and σ_u. Thus, the benchmark model becomes a special case of this extended model, when the coefficients associated with z_w and z_u are jointly zero (except the intercepts). The log likelihood function of this extended model is the same as the benchmark model. The only difference is that σ_u and σ_w in (9) are to be replaced by the functions in (14). Similarly, to estimate e^{-w_i} and e^{-u_i} in the extended model, we use the formulae in (12) and (13) but replace σ_u and σ_w by the functions given in (14).

5 A housing market application

We apply the 2TSF techniques developed in the preceding sections to examine the cost of incomplete information on housing prices. Our data[20] for this study comes from the AHS which was recently investigated by HRS.

5.1 Data

The data for the AHS is collected every 2 years and contains not only the traditional characteristics of the house, such as the number of rooms and square footage, but characteristics of the primary homeowner. Thus, attributes of buyers or sellers that may be believed, a priori, to influence the extent of search that takes place, such as having kids, wealth, not being from the local area, being a first time buyer, and the like, can be incorporated into the analysis in a parsimonious manner that will allow us to investigate the impact of information deficiency on housing prices.

We use a mix of house attributes and local area characteristics in the hedonic price function. The attributes used to explain housing prices are: square footage of the floor space, an indicator of whether the floor space variable was top coded, the total number of rooms in the house, the total number of bathrooms in the house, dummy variables for the age of the house, a quality control value that indicates if the house is deemed inadequate prior to being sold and dummy variables that capture the effect of the house being either a single family attached or a single family detached house. The local area characteristics are: dummy variables for the area the house is located in (city center, urban suburban, rural, other urban area), dummy variables for the population of the metropolitan statistical area (MSA) the house is located in (greater than 7 million people, between 3 and 7 million, between 1 and 3 million, and not in a MSA), and a climate control variable that ranges from 1 to 6. We also included time dummies (the

[19] This formulation is similar to the scaling method of Wang and Schmidt (2002) in a single-tier frontier model.

[20] We thank Stuart Rosenthal for providing us with the data set.

AHS runs from 1986 to 1993) to capture any year effects that may be effecting the housing market as well as MSA codes to account for any spatial structure which may be present. For a more detailed description of the data we refer the readers to HRS.

We use this set of variables in the benchmark model to estimate the hedonic price and then determine the cost of incomplete information for buyers and sellers. We also consider an extended model in which characteristics such as race, marital status, gender, having children, having a college education, owning a business, age, income, being a first time buyer, and being a buyer from out of town are allowed to affect prices indirectly through the means of u and w. In other words, the losses associated with information deficiency are allowed to vary systematically across buyers and sellers.

Before conducting our explicit variance analysis we take a look at the raw standard deviation of prices across different attributes of the house, buyer, and seller. These measures of spread are reported in Table 1. It is interesting to note that the standard deviations are quite large and vary across all three categories of attributes. One interesting anomaly is the decrease in spread of prices as one looks at houses further from the city center. While no means a formal analysis of price variation, these numbers do confirm the fact that there is a substantial amount of variation in prices. The sources of this variation are the starting point for our analysis of information deficiency in what follows.

5.2 Results from the benchmark model

Estimates of a log-linear hedonic price function from the benchmark model and the parameters associated with the distributions of v, u, and w are presented in Table 2. The coefficient estimates are comparable to those in HRS (Table A2–column 2) which may suggest that information is not linked to house attributes.[21] Our main interest is the estimates of the parameters, σ_u and σ_w, which are individually statistically significant, suggesting that both buyers and sellers in the housing market are incompletely informed. It is also important to note the size of the information variances relative to the "heterogeneity" variance. Our estimates suggest that a simple regression of price on characteristics implies that the variation attributed to heterogeneity is too large. By controlling for incomplete information we notice that a majority of the variance of the residual is being attributed to information deficiency and not heterogeneity, in line with Stigler's argument. It is also in accord with efficiency studies that typically find that the signal to noise ratio is quite large with respect to the inefficiency component and the heterogeneity component (see Kumbhakar and Lovell 2000).

Table 3 presents the mean and quartile values of several measures based on the point predictors of $E(e^{-u}|\varepsilon)$ and $E(e^{-w}|\varepsilon)$, which are measures of sellers' and buyers' price efficiency. From these point predictors we find that, on average, buyers are 72% efficient and sellers are 70% efficient. That is, on average, buyers paid 28% above the lowest WTA, $(1 - \hat{E}(e^{-w}|\varepsilon))$. The median and quartile values show large

[21] MSA indicators were included as well, but given that there are 142 or them in our sample we do not report them here to save space. They are available upon request.

Table 1 Variance of selling price across attributes

Attribute	SD	n	Attribute	SD	n
House attributes					
Square footage $< 2{,}000$	55,597	3,034	Square footage$\geq 2{,}000$	80,775	1,928
# Baths ≤ 2.5	60,961	4,569	# Baths > 2.5	99,438	393
# Rooms < 6	51,547	1,502	# Rooms ≥ 6	71,015	3,368
City center	67,163	1,236	Urban-suburban	77,265	2,068
Other urban	35,448	411	Rural	39,153	542
Age < 5	77,405	396	$5 \leq$ Age <10	69,350	699
$10 \leq$ Age < 15	71,898	730	$15 \leq$ Age < 30	67,067	1,397
Age ≥ 30	70,480	1,740	MSA pop. > 7 mill.	87,374	266
$3 <$ MSA pop. ≤ 7 mill.	77,738	326	$1 <$ MSA pop. ≤ 3 mill.	79,667	1,159
MSA pop. ≤ 1 mill.	61,534	646	Not in MSA	56,178	2,565
Buyer attributes					
Out of town	77,404	563	Local	69,647	4,399
First time	53,516	1,950	Repeat	77,560	3,012
Income $< 50{,}000$	52,573	3,422	Income $\geq 50{,}000$	80,837	1,540
Owns business	82,152	698	College educated	81,568	1,817
Age < 35	61,914	2,460	Age ≥ 35	77,322	2,502
Black	47,640	225	White	71,486	4,737
Married	72,783	3,555	Single	61,682	1,407
Kids	74,400	2,542			
Seller attributes					
Income $< 50{,}000$	60,250	3,614	Income $\geq 50{,}000$	79,169	1,348
Owns business	84,257	718	College educated	78,832	1,753
Age < 35	56,236	1,630	Age ≥ 35	76,004	3,332
Black	56,654	121	White	71,027	4,841
Married	71,278	3,422	Single	67,745	1,540
Kids	68,878	2,335			
Overall variance	70,793	4,962			

variations in information costs[22] paid by the buyers. The last row of Table 3 shows that, on average, sellers received 30% less than the highest WTP, $(1 - \hat{E}(e^{-u}|\varepsilon))$. The quartile values of sellers' incomplete information costs are quite similar to those of the buyers.

Aside from investigating the price efficiency of individual buyers and sellers, one can also examine whether buyers are benefiting from sellers' incomplete information or vice-versa. Using the estimated values of $E(e^{-u}|\varepsilon)$ and $E(e^{-w}|\varepsilon)$ for each transaction, we construct an estimate of $100[e^{w-u} - 1] = 100[e^{-u}/e^{-w} - 1]$, which is the net effect of incomplete information on the price in percentage terms. Table 3 presents the mean and quartile values of the net information cost on price. The first row of Table 3

[22] Information deficiency and the cost of incomplete information are interchangeably used.

Table 2 Parameter estimates of the benchmark hedonic price function

Variable	Estimate	Variable	Estimate
Constant	6.8087 (0.000)	Rural	−0.3282 (0.000)
Square footage	0.3731 (0.000)	Deemed inadequate	−0.0555 (0.392)
Square footage top coded	−0.1310 (0.002)	Degree days code (1–6)	−0.0128 (0.014)
Number of bathrooms	0.3024 (0.000)	Sale year: 1987	0.0383 (0.148)
Number of rooms	0.0302 (0.000)	Sale year: 1988	0.0835 (0.002)
Single family attached	0.7917 (0.000)	Sale year: 1989	0.1443 (0.000)
Single family detached	0.8599 (0.000)	Sale year: 1990	0.2136 (0.000)
Structure age $\leq$ 5 years	0.2545 (0.000)	Sale year: 1991	0.1490 (0.000)
Structure age 5–10 years	0.1402 (0.000)	Sale year: 1992	0.2073 (0.000)
Structure age 10–15 years	0.0741 (0.001)	Sale year: 1993	0.2169 (0.000)
Structure age $\geq$30 years	−0.0408 (0.037)	MSA $>$ 7 million	0.7063 (0.000)
Central city	−0.1074 (0.000)	MSA 3–7 million	0.2216 (0.000)
Urban/suburban	0.0025 (0.921)	MSA 1–3 million	0.2289 (0.000)
Other urban	−0.3047 (0.000)	Not in MSA	−0.0844 (0.001)
σ_v	0.1454 (0.000)	σ_u	0.4288 (0.000)
		σ_w	0.3889 (0.000)

The natural logarithm of the selling price is used as the dependent variable in the regression
Asymptotic *p* values are reported in parentheses next to each estimate. There are 4,962 observations

Table 3 Cost of incomplete information on housing prices (benchmark model)

Measure	Mean (%)	Q_1 (%)	Median (%)	Q_3 (%)	$Q_3 - Q_1$ (%)
$\hat{E}((e^{w-u}) \mid \varepsilon) - 1$	11.6	−23.7	−2.3	25.8	49.4
$1 - \hat{E}(e^{-w} \mid \varepsilon)$	28.1	17.0	19.9	34.1	17.1
$1 - \hat{E}(e^{-u} \mid \varepsilon)$	30.0	17.1	21.7	36.6	19.5

Given that our dependent variable is in natural logarithms we use the point predictors of e^{-w} and e^{-u} to measure buyer's and seller's price efficiency

shows that in more than 50% of the transactions buyers benefited in the sense that the information costs paid by the sellers (which lowered the price) exceeded those of the buyers (which increased the price). The median net effect is −2.3%, thereby meaning that information deficiency led to a decline in house prices of 2.3% or more for over half the houses in our sample. For a quarter of the transactions the net effect was a reduction of price by at least 23.7%, while in another quarter of all transactions (those in the third quartile) the net information cost led to an increase in prices by 25.8% or more. These results show that the impact of incomplete information on housing prices is quite large and varied substantially across transactions (the inter-quartile range, i.e., the difference between the third and the first quartile values, is 49.5%).

Although, at the median, the cost of incomplete information imposed on buyers and sellers nearly cancel out, there remain some buyers and sellers who are benefitting from the other party's information deficiency. While we have considered differences

Table 4 Cost of buyer's information deficiency (benchmark model)

Buyer type	Mean (%)	Q_1 (%)	Median (%)	Q_3 (%)	$Q_3 - Q_1$ (%)
Out of town buyer	30.0	17.2	22.5	39.2	22.0
First time home buyer	26.9	17.0	18.6	30.7	13.7
Owns a business	29.9	17.7	23.0	37.7	20.0
Buyer is black	30.1	17.4	21.4	38.5	21.1
Buyer is married	28.8	17.1	20.4	35.7	18.6
Buyer is single female	29.9	17.7	23.2	37.8	20.1
Buyer is college educated	27.5	17.0	19.7	33.0	16.0
Buyer has kids	26.8	17.0	19.3	31.7	14.7

We use $1 - \hat{E}\left(e^{-w} \mid \varepsilon\right)$ to calculate all entries in the table

Table 5 Cost of seller's information deficiency (benchmark model)

Seller type	Mean (%)	Q_1 (%)	Median (%)	Q_3 (%)	$Q_3 - Q_1$ (%)
Owns a business	27.1	17.0	20.1	32.3	15.2
Seller is black	39.4	19.7	33.0	53.3	33.6
Seller is married	28.4	17.1	21.4	34.3	17.2
Seller is single female	33.0	17.1	23.8	44.6	27.5
Seller is college educated	25.4	17.0	19.1	28.5	11.6
Seller has kids	29.6	17.3	22.6	36.7	19.3

$1 - \hat{E}\left(e^{-u} \mid \varepsilon\right)$ is used to calculate all entries in the table

between buyers and sellers, it is also of interest to investigate deficiency for different groups of buyers and sellers.

To explore the issue of differences in information deficiency across groups of buyers and sellers, we evaluate the costs of information deficiency across several groups. The results are presented in Tables 4 and 5 using point predictors of $1 - e^{-w}$ and $1 - e^{-u}$, which (when multiplied by 100) are the percentage increase/decrease in prices due to buyers'/sellers' information deficiency.

The entries in Table 4 represent the percent by which different types of buyers overpay relative to the minimum WTA. We find that, on average, the cost of information deficiency for first time home buyers and buyers with kids, are the least, while the cost to out of town buyers, blacks, single females, and those who own a business are somewhat higher. The first quartile values suggest almost no difference across different groups of buyers, while the third quartile values suggest that buyers who are from out of town, black or single female pay a higher price due to information deficiency. This result may be justified for those who own a business by arguing that their opportunity cost of time is higher. It is worth noting that substantial variation in the costs of incomplete information (measured by the interquartile range) are observed across different groups of buyers.

The entries in Table 5 represent the percent by which sellers' prices are reduced in relation to the maximum WTP (labeled as sellers' cost of incomplete information).

Table 6 Estimates of parameters in the σ_u and σ_w functions

Variable	Estimate	Variable	Estimate
σ_v	0.1401 (0.000)		
σ_w		σ_u	
Constant	−1.1277 (0.000)	Constant	−0.4606 (0.000)
Out of town buyer	−0.0309 (0.713)		
1st time buyer	0.0287 (0.655)		
Buyer's income	−0.0257 (0.561)	Seller's income	−0.1721 (0.000)
Buyer has business	−0.0362 (0.637)	Seller has business	−0.0013 (0.984)
Buyer's age	0.0965 (0.330)	Seller's age	0.0064 (0.933)
Buyer is black	−0.3537 (0.0139)	Seller is black	0.2272 (0.074)
Buyer is married	−0.1262 (0.102)	Seller is married	−0.2787 (0.000)
Buyer is single female	0.0228 (0.813)	Seller is single female	−0.0030 (0.700)
Buyer has college education	−0.0006 (0.991)	Seller has college education	−0.2512 (0.000)
Buyer has kids	−0.0896 (0.123)	Seller has kids	−0.0655 (0.233)

Asymptotic p values are reported in parentheses next to each estimate

On average, blacks and single females face the biggest price reductions (face the largest losses due to information deficiency), while those who are college educated or own a business face a smaller reduction in price. Comparing Tables 4 and 5, we find that the variation in incomplete information costs is much smaller for sellers in every category.

While these results are interesting, a general model that accounts for buyer's and seller's attributes directly into the hedonic price function, and indirectly through the means of u and w (i.e., σ_u and σ_w) is desirable. In this framework incomplete information costs are allowed to vary systematically across buyers and sellers. Results from this generalized model are reported next.

5.3 Results from the generalized model

Our generalized hedonic model includes the same characteristics and attributes as the benchmark model, except we now include buyer attributes as determinants for the mean of buyer information deficiency and seller attributes as determinants for the mean of seller deficiency. To conserve space we do not report the coefficients from the hedonic function.[23]

In Table 6 we report the estimate of σ_v and the parameters associated with the σ_u and σ_w functions. So far as buyers' characteristics are concerned, we find that only the black dummy is significant. The negative sign on the black dummy in σ_w suggests that the incomplete information costs for black buyers is lower (everything else being the same). However, given that only 2.4% of all buyers are black in our sample, this result may not be representative of the underlying population of black home buyers.

[23] The estimates are almost identical to those of the benchmark model and are available upon request.

Table 7 Cost of incomplete information on housing prices (extended model)

Measure	Mean (%)	Q_1 (%)	Median (%)	Q_3 (%)	$Q_3 - Q_1$ (%)
$\hat{E}((e^{w-u}) \mid \varepsilon) - 1$	1.3	−23.9	−4.6	15.5	39.4
$1 - \hat{E}(e^{-w} \mid \varepsilon)$	23.3	15.0	17.5	26.3	11.3
$1 - \hat{E}(e^{-u} \mid \varepsilon)$	28.1	15.5	20.4	35.1	19.6

Given that our dependent variable is in natural logarithms we use the point predictors of e^{-w} and e^{-u} to measure buyer's and sellers' price efficiency

Next, we look at the sellers' side of the market. Unlike buyers, several variables are found to be significant in explaining σ_u. A seller's income, being married, and having a college education are all significant at the 1% level, while the black dummy is significant at the 10% level. All four of these characteristics have negative coefficients suggesting that the presence of these characteristics (given everything else) reduce the information cost paid by these sellers, on average. A likelihood ratio test failed to reject the null hypothesis that the buyer and seller attribute variables were jointly insignificant within the information deficiency distributions.

As mentioned earlier, since both buyers and sellers face losses due to incomplete information and one party's loss is other party's gain, it is more meaningful to examine the net effect of buyers and sellers information costs on prices. The results are reported in Table 7. Computations of all the measures reported here are the same as those in Table 3.

Similar to the benchmark model, we find that in more than 50% of the transactions buyers benefited in the sense that their loss relative to sellers was smaller, resulting in a price reduction. The median net effect shows a decline in price by 4.6% due to buyer and seller incomplete information. For a quarter of all transactions, the net effect was a reduction of price by at least 23.9%, while in another quarter of the transactions, net incomplete information costs led to an increase in prices by 15.5% or more. These results show that even after taking buyers' and sellers' characteristics into account within the hedonic price function (as has been done in previous research), the impact of buyers' and sellers' incomplete information on housing prices does not vanish. It is clear from Table 7 that variation in net incomplete information costs across transactions is quite large (the inter-quartile range is 39.5%), although somewhat smaller compared to the benchmark model (see Table 3).

In sum, the cost of information deficiency led to a price decrease (at the median) by 4.6%, as opposed to 2.3% in the benchmark model (reported in Table 3). This suggests that buyers, at the median, are benefiting from sellers' incomplete information. Intuitively this makes sense because buyers can search over a broad array of houses for very low cost, while sellers have to keep their house on the market longer if they wish to gather information from prospective buyers. Keeping a house on the market longer can result in a lower selling price as it may be evidence that the house is of poor quality (see Anglin 2005 for more on this).

Next we consider the cost of incomplete information across different types of buyers and sellers. The results are reported in Tables 8 (for buyers) and 9 (for sellers).

Table 8 Cost of buyer's information deficiency (extended model)

Buyer type	Mean (%)	Q_1 (%)	Median (%)	Q_3 (%)	$Q_3 - Q_1$ (%)
Out of town buyer	22.9	14.4	17.4	26.3	11.8
First time home buyer	23.3	15.1	17.5	26.3	11.1
Owns a business	23.5	14.9	17.2	27.2	12.3
Buyer is black	18.0	12.8	14.5	17.9	5.1
Buyer is married	22.4	14.5	16.6	25.4	10.9
Buyer is single female	25.8	16.7	18.9	29.2	12.5
Buyer is college educated	23.2	14.3	17.2	27.8	13.5
Buyer has kids	21.9	14.4	16.4	24.8	10.5

$1 - \hat{E}\left(e^{-w} \mid \varepsilon\right)$ is used to calculate all entries in the table

It can be seen from Table 8 that differences in losses due to information deficiency across groups of buyers disappear (except for the blacks) when the group characteristics are taken into account in the hedonic price function as well as in σ_u and σ_w. Like the benchmark model, the costs of information deficiency for black home buyers, ceteris paribus, is found to be lower.[24] Turning to the sellers in Table 9, we find that cost of information deficiency for blacks and single females are the highest (evidenced by the mean and quartile values). Costs of information deficiency for sellers with college education, having kids, being married, and owning a business are the least. Again this is evidenced by the mean and quartile values. Here the intuition is that sellers with the above characteristics can afford to leave their house on the market longer (thereby having more information about the prospective buyers) which brings down the cost of information deficiency. In summary, we find that the incomplete information for both buyers and sellers are nontrivial and these costs are quite similar for each group. Moreover, the variations in these costs for each group, measured by the inter-quartile range, are also similar.

While the results from both the benchmark and generalized model suggest that the majority of price dispersion was due to information deficiencies on *both* sides of the market and that there were quite large gaps between selling prices and limits prices, the AHS dataset used is pre-internet. That is, existing information conduits for buyers and sellers via the internet either did not exist or were so infrequently used that it is possible these gaps really exist. That being said, an avenue for future research would be to look at the variation in informational deficiencies pre and post internet to determine the extent of the information provided economic agents.

5.4 An attempt at controlling the spatial dimension

An interesting issue that arises separately in hedonic housing modeling is the spatial nature of the data. Even if every attribute of the house was controlled for, the analyst may still find persistent price variations across regions given that people have a

[24] Given that there are only 225 black buyers in our sample one must interpret this result with care.

Table 9 Cost of seller's information deficiency (extended model)

Seller type	Mean (%)	Q_1 (%)	Median	Q_3 (%)	$Q_3 - Q_1$ (%)
Owns a business	26.6	15.1	20.2	33.6	18.5
Seller is black	33.2	16.4	23.6	42.0	25.7
Seller is married	25.2	14.6	18.6	29.6	15.1
Seller is single female	35.7	18.0	28.6	48.5	30.5
Seller is college educated	22.7	13.6	16.9	27.2	13.7
Seller has kids	25.6	14.5	18.6	30.4	15.9

$1 - \hat{E}\left(e^{-u} \mid \varepsilon\right)$ is used to calculate all entries in the table

natural preference for one area over another. This preference will lead to higher prices in certain areas and lower prices in others. To assuage the thought that our estimates are picking up this sort of variation as opposed to variation due to incomplete information we included a set of indicators that controlled for the MSA the house was located in, the smallest regional identifier within the AHS, in estimation of both our benchmark and generalized models.[25]

Upon estimation of both the benchmark and generalized hedonic models *without* the MSA indicators we found that the while the slope coefficients were only marginally different our heterogeneous variation estimates were increased significantly with our information mean/variance parameters only insignificantly changed.[26] While this may seem counter intuitive, it has a rather natural explanation. Our results suggest that there are no regions which have more information in the housing market than others, i.e., the level (or lack thereof) of information is not dictated by the MSA in which one buys/sells a house. The heterogeneity variance increasing is more in accord with the above anecdotal evidence that there are certain properties of MSAs which are a cause of price variation, but not in regards to information.

Overall, having access to buyer and seller characteristics through use of the AHS has reduced the spatial resolution of the data. The lack of adequate spatial resolution may cause omitted variable biases given that many housing characteristics are "local amenities" not typically captured in a housing dataset (school quality for example). Unfortunately, given that the AHS only provides housing locations up to an MSA this is the finest spatial grid with which we can attempt to control for unseen spatial variations across our sample while including home owner characteristics. It is also hard to assess the likely direction of any bias arising from such spatial omissions. Houses in bad neighborhoods, everything else equal most likely sell for less whereas houses in good neighborhoods, everything else equal sell for more. However, if buyers and sellers know this it is likely that some of these neighborhood effects are already part of the information set used when making a purchsing/selling decision.

[25] This spatial resolution is not fine enough however to tell us if the house is in a good or bad neighborhood within the MSA however.

[26] A likelihood ratio failed to reject that these estimates were the same across the two different covariate sets.

Thus, it is imprudent for us to assign exact effects of biases from using too small a spatial resolution structure for our analysis.[27]

6 Implications for price efficiency

Our findings should be placed in the context of market efficiency and housing studies of the past. We want to relate our findings of significant differences in buyer and seller limit prices to the concept of an efficient market. It is quite natural that all parties are incompletely informed which does not imply that the housing market is inefficient. Rather, it seems that the levels of incomplete information are consistent across buyers and sellers, as well as across characteristics of both of them.

So why does incomplete information persist in the housing market? A house is the most durable of durable goods and has to be looked at in a different light. As a product wears our consumers go back to the market to purchase the item again. As the length of time increase between successive visits, the information gathered previously dissipates, leaving one in virgin territory. What this suggests is that by the time a home buyer (or seller) goes back to the market, it has changed sufficiently enough to render most of the acquired information unhelpful in finding a lower (higher) price on another house. While there have been myriad studies documenting the forecastability of housing prices (suggesting the market is inefficient), they do not relate to whether buyers have access to better information than sellers. See Case and Shiller (1988, 1989, 1990) for documentation on the inefficiency of the single family housing market. Evidence that is suggestive of near equality of information across buyer/seller types can be found in Turnbull et al. (1990) for houses sold by corporations versus those sold by noncorporate owners and Turnbull and Sirmans (1993) for first time home buyers versus repeat home buyers as well as local buyers versus out-of-town buyers. There has been no study investigating differences in information across buyers and sellers however, which is what our research accomplishes.

7 Conclusion

In this paper we use a two-tier stochastic hedonic price function to decompose price variations into those explained by observed variables, unobserved product heterogeneity, and information deficiency, which is further decomposed into buyer and seller components. Since obtaining information is costly, and these costs are likely to vary among buyers and sellers, the presence of incomplete information imposes a cost on both buyers and sellers. Such costs may be optimal (efficient). That is, a buyer might pay a higher price for a good, knowing that the search costs (opportunity cost of time) associated with gathering additional information necessary for obtaining a lower price are too high. The same is true for a seller who might sell the good at a lower price instead of waiting longer to gather more information about prospective buyers who might be willing to pay more. The two-tier stochastic frontier approach used in this

[27] See Kuminoff et al. (2009) for more on the impact of omitted variable bias in hedonic models.

paper enabled us to obtain estimates of price efficiency and cost of information deficiency for each buyer and seller. We also extended the model to allow information costs to depend on buyers' and sellers' characteristics. This formulation allowed systematic variation in incomplete information costs, estimates of which are used to analyze differences in the cost of incomplete information across different types of buyers and sellers.

An application of the model to AHS data showed that the impact of incomplete information on market prices is not negligible for either buyers or sellers. On average, buyers were found to be 72% price efficient while the price efficiency of sellers was 70%. That is, because of information deficiency, buyers, on average, paid 28% above the lowest available WTA, while sellers, on average, received 30% less that the highest WTP. In the extended model that takes buyers' and sellers' characteristics into account, these figures are 23.3 and 28.1%, respectively. That is, information costs led to a net price decrease (at the median) of 4.6%, as opposed to 2.3% in the benchmark model. These results suggest that buyers, at the median, are benefiting from sellers' incomplete information.

The estimation of the costs of incomplete information may prove useful in many other economic settings where buyers (sellers) seek the lowest (highest) price and they may not share the same information. Some examples include price determination in auctions and used car markets, dowries in marriage markets, wages in labor markets, and premiums in insurance markets.

References

Anglin PM (2005) The selling process: if, at first, you don't succeed, try, try, try again. Working Paper, University of Windsor

Case KE, Shiller RJ (1988) The behavior of home buyers in boom and postboom markets. N Engl Econ Rev November/December:29–46

Case KE, Shiller RJ (1989) The efficiency of the market for single family homes. Am Econ Rev 79(1): 125–137

Case KE, Shiller RJ (1990) Forecasting prices and excess returns in the housing market. J Am Real Estate Urban Econ Assoc 18(3):253–273

Cooper JC (2006) Prices and price dispersion in online and offline markets for contact lenses. Working Paper No. 283, Bureau of Economics, Federal Trade Commission

Gaynor M, Polachek SW (1994) Measuring information in the market: an application for physician services. South Econ J 60(4):815–831

Griliches Z (1971) Price indices and quality change. Harvard University Press, Cambridge

Harding JP, Rosenthal SS, Sirmans CF (2003) Estimating bargaining power in the market for existing homes. Rev Econ Stat 85(1):178–188

Häring J (2003) Different prices for identical products? Market efficiency and the virtual location in B2C e-commerce. Discussion Paper No. 03-68, ZEW (Centre for European Economic Research)

Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econom 19:233–238

Kumbhakar SC, Lovell CAK (2000) Stochastic frontier analysis. Cambridge University Press, New York

Kumbhakar SC, Parmeter CF (2009) The effects of bargaining on market outcomes: evidence from buyer and seller specific estimates. J Prod Anal 31(1):1–14

Kuminoff N, Parmeter CF, Pope JC (2009) Assessing the efficacy of modern hedonic estimators in addressing omitted variable bias. AAEC Working Paper (2009)-02, Virginia Tech

Lach S (2002) Existence and persistence of price dispersion: an empirical analysis. Rev Econ Stat 84(3):433–444

Parmeter CF, Pope JC (2009) Quasi-experiments and hedonic property value methods. SSRN Working Paper 1283705
Polachek S, Yoon BJ (1987) A two-tiered earnings frontier estimation of employer and employee information in the labor market. Rev Econ Stat 69(2):296–302
Polachek S, Yoon BJ (1996) Panel estimates of a two-tiered earnings frontier. J Appl Econom 11(2): 169–178
Pratt J, Wise D, Zeckhauser R (1979) Price differences in almost competitive markets. Q J Econ May: 189–211
Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. J Political Econ 82(1):34–55
Rothschild M (1973) Models of market organization with imperfect information: a survey. J Political Econ 81(6):1283–1308
Sorensen AT (2000) Equilibrium price dispersion in retail markets for prescription drugs. J Political Econ 108(4):833–850
Stigler G (1961) The economics of information. J Political Econ 69(3):213–225
Tsionas EG (2006) Maximum likelihood estimation of stochastic frontier models by the fourier transform. Working Paper, Athens University of Economics and Business
Turnbull GK, Sirmans CF (1993) Information, search and house prices. Reg Sci Urban Econ 23(4):545–557
Turnbull GK, Sirmans CF, Benjamin JD (1990) Do corporations sell houses for less? A test of housing market efficiency. Appl Econ 22:1389–1398
Wang H, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. J Prod Anal 18:129–144