The hedonic approach estimates the value attached (i.e. the implicit prices) to each of these attributes. The approach entails regressing the logarithm of the sales price on the relevant attributes. The standard hedonic model usually takes the following form:

$$\ln P_{it} = \sum_{t=1}^{T} \delta_t D_{it} + \sum_{j=1}^{J} \beta_{jt} X_{jit} + \sum_{k=1}^{K} \gamma_{kt} Z_{kit} + \epsilon_{it}$$

where P_{it} represents the price of item i at time t ($t = 1, ..., T$); D_{it} is a time dummy variable taking the value of 1 if item i is sold in period t and 0 otherwise, X_{jit} ($j = 1, ..., J$) is a set of j observed attributes of item i at time t; Z_{kit} is a set of k ($k = 1, ..., K$) unobserved attributes that also influence the price; and ϵ_{it} is a random (white noise) error term.

Table 3 reports the correlations in the growth rates between the various indices. The significant positive correlations between the regression-based indices indicate that their general trends are similar, and are different from the simple median. Allowing for time variation in the hedonic coefficients does not materially affect the results and neither does the exact value of the distance metric chosen.

4.3 Understanding the Gap between ps-RS Index and Hedonic Index Recognizing that the ps-RS model derives from the differentials of the hedonic model within the matched pairs is helpful to understand the gap between two the index series. The two transactions in a pair are quite homogeneous in physical and location attributes. The small within-pair differences in physical characteristics, such as floor number, unit size and number of rooms, are also easy to be identified. The ps-RS regression will drop out all observed and unobserved common attributes, thus we are able to mitigate the problem of omitted variables or variables that are difficult to measure reliably and obtain index results that are more robust. This is the advantage of the parsimony in the RS specification. In contrast, omitted variables or hedonic variables that are difficult to properly value or quantify can bias an index estimated from a hedonic model. Therefore, the gap between the hedonic model and the ps-RS model can be attributed to the unobserved variables cancelled out in the ps-RS model. Indeed, since the ps-RS model uses all transactions (unlike the classical RS model that can only use repeat-sales), it would seem that omitted (or mis-valued) hedonic variables must be the major source of any difference between the hedonic and the ps-RS indices in Figure 2. This is different from a typical comparison between hedonic and classical RS indices, as in that case the estimation databases are also different.

There are also two other sources of difference between the ps-RS and hedonic indices, which may partly explain the difference we observe in Figure 2. While the ps-RS model is based on all and only the same transactions as the hedonic model, the matching process generates a much larger (pseudo) sample size for the ps-RS model than what the hedonic model has to work with. This larger sample size should help the ps-RS model to be estimated more precisely, resulting in less noise in the index. Finally, a third source of difference between the two indices could arise from the use of the differential specification in the ps-RS model versus the undifferenced (levels) specification in the hedonic model. The ps-RS model directly estimates longitudinal price changes, whereas the hedonic model directly estimates price levels as of one point in time (and the hedonic index of longitudinal price changes is then

Table 1: Hedonic Regression results

	Dependent variable:
	lnprice
lnarea	0.405*** (0.004)
ah_codeAshbeys	0.114*** (0.028)
ah_codeBernardi	0.122*** (0.013)
ah_codeBonhams	1.198*** (0.026)
ah_codeChristies	1.152*** (0.065)
ah_codeRussell Kaplan	0.129*** (0.015)
ah_codeStephan Welz	0.590*** (0.013)
ah_codeStrauss	1.126*** (0.016)
med_codeWatercolour	0.553*** (0.017)
med_codeOil	1.435*** (0.014)
med_codeAcrylic	1.039*** (0.033)
med_codePrint/Woodcut	−0.759*** (0.015)
med_codeMixed Media	0.554*** (0.017)
med_codeSculpture	1.767*** (0.084)
med_codePhotography	−0.618*** (0.085)
med_codeOther	0.703*** (0.035)
lnsculpt_area	0.236*** (0.021)
dum_signed	0.208*** (0.016)
dum_dated	0.045*** (0.007)
nr_works	−0.096*** (0.003)
lnrep	0.949*** (0.003)
Constant	5.101*** (0.065)
Quarterly dummies	Yes
Observations	51,454
R^2	0.782
Adjusted R^2	0.782
Residual Std. Error	0.780 (df = 51369)
F Statistic	2,196.018*** (df = 84; 51369)
Note:	*p<0.1; **p<0.05; ***p<0.01

only constructed later from the differences in the hedonic model's time-dummy coefficients). The longitudinal differencing in the underlying ps-RS regression model may affect the results.

Returning to what we believe is the major source of difference between the ps-RS and hedonic indices, suppose the problematical omitted variable in the hedonic index is a positive attribute favored by households and the share of transactions with this attribute is rising over time. For example, suppose newer housing units built more recently have higher quality of the finishes on the flooring, walls and ceilings, or maybe higher quality of the heating and air conditioning systems, air and water filtration systems, or better kitchen/bathroom appliances, but the hedonic database does not have any information about quality improvement except of the number of rooms. Then the hedonic index will tend to overestimate the rate of price growth. It will in effect attribute the value of higher physical quality of housing units to the housing market condition (when in fact these represent the market for better physical quality of apartments). In such a case we would see the ps-RS index tending to track below the hedonic index. In reality, with such rapidly rising per capita income, it would seem likely that the new housing units have been incorporating more and more favorable attributes in terms of the physical characteristics within the units.

If, on the other hand, the share of the transactions with such favored (omitted) attributes is declining, then the ps-RS index will track above the hedonic index. In the case of negative (unfavorable) attributes, the situation is the opposite to what we have just described. One typical case is that the rapid urbanization has meant that location attributes may be inevitably tending to be less favorable (farther away from the CBD, although mitigated perhaps by transport infrastructure improvements and rising automobile ownership). It is possible that not all of these changes can be completely captured or accurately measured in the hedonic attributes database.

The repeat sales model can be derived from the hedonic model,[^5] if the hedonic model is differenced with respect to consecutive sales of items that have sold more than once in the sample period [@McMillen2012]. The standard model may be formulated as the change in the log of the sales price of item i that sold at time t and an earlier time s:

$$\ln P_{it} - \ln P_{is} = (\sum_{t=1}^{T} \delta_t D_{it} - \sum_{s=1}^{T} \delta_s D_{is}) + (\sum_{j=1}^{J} \beta_{jt} X_{jit} - \sum_{j=1}^{J} \beta_{js} X_{jis}) + (\sum_{k=1}^{K} \gamma_{kt} Z_{kit} - \sum_{k=1}^{K} \gamma_{ks} Z_{kis}) + (\epsilon_{it} - \epsilon_{is})$$

$$\ln P_{it} - \ln P_{is} = (\delta_t D_{it} - \delta_s D_{is}) + (\beta_{jt} X_{jit} - \beta_{js} X_{jis}) + (\gamma_{kt} Z_{kit} - \gamma_{ks} Z_{kis}) + (\epsilon_{it} - \epsilon_{is})$$

If the attributes (X and Z) of item i and the implicit prices (β and γ) are constant between sales, the equation reduces to the standard estimating equation:

$$\ln \frac{P_{it}}{P_{is}} = \sum_{t=1}^{T} \delta_t G_{it} + u_{it}$$

where P_{it} is the purchase price for item i in time t; δ_t is the parameter to be estimated for time t; G_{it} represents a time dummy equal to 1 in period t when the resale occurs, -1 in period s when the previous sale occurs, and 0 otherwise; and u_{it} is a white noise residual.

$$\ln P_{itb} - \ln P_{hsb} = \sum_{j=1}^{J} \beta_j (X_{itbj} - X_{hsbj}) + \sum_{t=0}^{T} \delta_t G_{it} + \epsilon_{ithsb}$$

where G_{it} is again a time dummy equal to 1 if the later sale in the pair occurred in quarter t, -1 if the former sale in the pair occurred in quarter s, and 0 otherwise; and ϵ_{srabj} again represents a white noise residual.

eliminated the time-invariant and unobserved effects:

$$\ln P_{it} - \ln P_{is} = (X_{it} - X_{is})\beta + (\delta_t - \delta_s) + (u_{it} - u_{is})$$