WORKING PAPER – PLEASE DO NOT QUOTE WITHOUT AUTHOR'S PERMISSION.

Value Follows Form:
The Shape of Buildings as a Predictor of Property Values

by
Thies Lindenthal
University of Cambridge, Department of Land Economy
htl24@cam.ac.uk

Keywords: Spatial hedonic regressions, Big Data, property prices, building codes, shape distributions, mass appraisals

Popular wisdom advises not to judge a book by its cover. For a building, however, the opposite holds true: The outer shell of a structure reveals a wealth of information about the place (Jensen & Cowen, 1999). A passer-by can quickly estimate the type and spaciousness of a dwelling, the number of floors and rooms within it and assess maintenance levels and upkeep. Trained observers are able to deduct the year of construction from the architecture, the height of rooms from the location of windows and other element in the facade, or estimate the energy efficiency of the building by inspecting the windows and thermal insulations from a distance. In addition, the choice of exterior building materials oftentimes relates to the exclusivity of interior finishes and additional amenities like garages, balconies are directly observable. Finally, certain shapes might be perceived as more aesthetically pleasing than others and therefore carry a direct architectural premium.

So far, large-scale empirical studies have not tapped into the shape of buildings as a source of property-level data since three-dimensional building data and methods to analyze them were lacking. Instead, research of property values and market participants rely solely on on-site inspections or land registry records to obtain information on easily observable property characteristics such as the interior floor space or number of bedrooms. Less palpable dimensions like the layout of floorplans or the silhouette of the building remain unrecorded and ignored.

In recent years, urban planners and commercial data providers have developed extensive three-dimensional maps of entire cities (Kolbe, 2009; Shiode, 2001). This paper is the first to explore 3D city models from a real estate finance perspective as it links a building's shape to property characteristics and its economic value. First, it refines, explores and verifies a method to convert three-dimensional building shape data into a numerical representation that can be fed into mass appraisal systems. The second part of the paper tests whether these quantitative representations of shape can help to explain recent transaction prices using real-life property data from the Dutch city of Rotterdam.

Before delving into any details, a fundamental question needs to be addressed: Why could a measure of property shapes be a useful predictor of property values? The early 20th century modernist movement in art and architecture postulated that "form follows function" (Sullivan, 1896): The intended function or use of a building determines its shape. Simultaneously, the value of a property also depends on these functions or services provided. In combination, if the *form* of a building stems from the same set of functions that also influence its *price*, then shape data might contain information useful for the valuation of buildings.

Tapping into large-scale data on building shapes unlocks a wealth of property-level data for *all* buildings in a city. This drastically increases the number of observations available for analysis compared to subsamples of properties based on sales, mortgage originations or valuations for tax purposes. Furthermore, the direct context of neighboring buildings can be analyzed which is not always possible when relying on samples instead of the universe of structures.

Additionally, considering the shape of properties in empirical price estimations can put a price tag on certain architectural forms or the value of homogeneity at the block level. Finally, controlling for shape in hedonic regressions leads to more accurate coefficient estimates for attributes like housing type, year of construction or location. Once the shape is separately accounted for, the marginal prices for correlated variables become less of an overall price for a bundle of different attributes (including shape) and more an attribute-specific price.

The remainder of the paper will first position the study in the current academic discourse, suggest a methodology to analyze shapes numerically, introduce the data, present the results and discuss the findings in a conclusion.

Earlier Work

Advances in the interpretation of remotely sensed data has lead to a surge of large and spatially consistent data sets with detailed three-dimensional information at building level. New York, Paris, Singapore, Tokyo and many other cities can be explored digitally, while the municipalities of Berlin or Rotterdam even openly share semantic city models[1]. So far, these models have been put to use in a wide range of research areas, including urban planning (Ranzinger & Gleixner, 1997; Wu, He, & Gong, 2010), disaster management (Kwan & Lee, 2005), policing (Wolff & Asche, 2009), navigation (Rakkolainen & Vainio, 2001), facility management and building information models (Nagel, Stadler, & Kolbe, 2009), or emission and other environmental modeling (Nichol & Wong, 2005). However, quantitative research linking a building's shape to its prices is still lacking.

Existing work provides clues that form indeed is a determinant of value. At least architecture, as a crude proxy for form, is. For residential homes the architectural style of buildings can partially explain sales prices (Asabere, Hachey, & Grubaugh, 1989). For row houses in Boston's South End, a premium on distinct architectural features of the facade has been documented for historical sales (Moorhouse & Smith, 1994). In 19th century Boston, it is good to be different: The architecture-premia observed become smaller with each additional

1 3D city data for Berlin is available for download at http://www.businesslocationcenter.de/en/downloadportal and for Rotterdam at http://www.rotterdam.nl/rotterdam_3d.

building in the proximity sharing the same architectural style. For commercial properties, buildings by star architects command higher rents and values (Fuerst, McAllister, & Murray, 2011). At the neighborhood level, perceived beauty of the built environment, is one of the main determinants of the resident's satisfaction, alongside economic factors, school quality and the perceived opportunity of social interactions (Florida, Mellander, & Stolarick, 2009).

A building's shape also influences real estate values in its surroundings. Positive externalities of historic landmark buildings have been confirmed for the US in general (Listokin, Listokin, & Lahr, 1998), Texan cities (Leichenko, Coulson, & Listokin, 2001), or the German capital Berlin (G. M. Ahlfeldt & Maennig, 2010). Similarly, dwellings designed by famous architects contribute to surrounding property values. Homes within 50 m to a residential building by Frank Lloyd Wright in Oak Park, Illinois, enjoy a price premium of 8.5 percent. It remains unclear, however, if this premium is caused by the aesthetic properties of the buildings or by the star-status of the architect (G. Ahlfeldt & Mastro, 2012). Buildings can also impair the value surrounding buildings very directly, just by offering or blocking a sought-after view. An unobstructed sea view (estimated at property level from a 3D city model) will increase property prices by 15% in Singapore (Yu, Han, & Chai, 2007). Positive values for viewsheds on nature and historical buildings have also been documented for Kyoto (Yasumoto, Jones, Nakaya, & Yano, 2011).

In previous work, it is unlikely that the estimated price effects for architecture are solely determined by the form of the buildings. Instead, any premium is a combined estimate for a diverse combination of multiple attributes like year of construction, size or the micro-location within a city. A more concise measure to single out the direct effect of shape is still wanting.

Methodology

The shape of buildings remains difficult to describe and information that can be analyzed on a large scale is not readily available. Within the constraints imposed by topography, climate, construction costs, urban planning, lot sizes and lot shapes, the form of buildings varies extensively. Individual developer preferences, architectural creativity and amendments during the life-time of a building are a source of diversity. At the same time, economies of scale during construction, architectural preferences for harmony and overall fashion trends induce similarities between buildings while lot specific factors,

The heterogeneity and multi-dimensionality of building shapes renders their classification a non-trivial challenge. Broad categories can describe roof forms, the 2D shape of the ground plates, or overall dimensions, angles and ratios. Still, classifications relying on a manageable number of categories cannot provide a finely grained view and the variation in shape within each of the classes remains high. This paper circumvents the problem of finding a meaningful classification system for property forms by estimating pairwise shape similarities between buildings instead.

Methods on measuring shape similarity both in 2D and 3D have been researched extensively in computer graphics, computer vision, biology and other disciplines. For a general review please refer to Cardone, Gupta, & Karnik (2003) or Tangelder & Veltkamp (2008). This paper builds on the *shape distribution* approach put forward by Osada, Funkhouser, Chazelle, & Dobkin (2001). A large number of random points are drawn from the surface of each shape and pairwise distances between these points are calculated. The estimated probability density functions (EDF) of these distances represent building-specific *shape signatures* that can be stored and compared efficiently for large numbers of buildings. The distributions can be normalized by dividing by the average distance for each shape.

======= Insert Figure 1 about here =========

Figure 1 illustrates that differences in the shape configuration lead to distinct differences in the density functions. Three simple shapes are constructed by combining two base shapes, cubes and triangular prisms. The shape distribution of a single cube exhibits a single distinct peak while the distribution for a cuboid, formed by joining two cubes, has a long tail to the right. Adding a roof to the cube changes the resulting shape distribution yet again: The "house with saddle roof" representation differs strongly from the other two examples.

While it is easy to reduce 3D objects to univariate shape distributions, the reverse direction is not. The skewness of the distribution gives a rough indication of the overall compactness of a structure but backing out shape details from shape distributions is not feasible.

Similar shapes will lead to similar distributions. Intuitively, if the area between two plots of shape distributions is small, then the original shapes can be considered similar. A pairwise measure of similarity $S_{i,j}$ for shapes i and j is calculated from the respective EDFs (similar to Osada et al., 2001):

$$S_{i,j} = 1 - \int_{d=0}^{d=D} \left| edf_i(d) - edf_j(d) \right| \quad (1)$$

Obviously, $S_{i,j} = S_{j,i}$.

Shape distributions possess several advantageous characteristics: they can be calculated for solid and non-solid 3D shapes like surfaces and 2D shapes alike and are tolerant to errors in the underlying geometries (Ohbuchi, Minamitani, & Takei, 2005). This robustness is crucial when working with shape data for large numbers of buildings that have been automatically derived from areal scans and oftentimes comprise of non-solid shapes for individual buildings (Alam, Wagner, Wewetzer, Coors, & Pries, 2013), due to small gaps between walls or missing walls between adjacent buildings. In a sense, the building models that will be later used in this study are drafty. If one printed the building models on a 3D printer only few houses would be reasonably airtight. The share of non-solid building-level models derived from 3D city models has been documented to be as high as 95% (Boeters, 2013), which rules out any approach requiring input shapes to be solid.

The accuracy and relevance of the suggested estimate of shape similarity S is first tested directly: Are buildings, that are known to have identical forms, recognized as being similar? In real cities, the most basic form is probably a cube, which is also the easiest to identify based on their typical geometric characteristics. Cube-buildings feature exactly four walls, a roof and a ground plate which are all squares of the same area. For a subset of cube buildings, the estimate of pairwise similarity *S* is expected to be close to 1, with 1 representing perfect identity. Among dissimilar shapes, *S* is hypothesized to be significantly smaller.

The mapping of shapes to shape distributions is not a bijective function. Shape distributions are invariant to rotation, mirroring and, if normalized, also to scaling (Osada et al., 2001). While a shape is converted into exactly one shape distribution, one distribution can be the shape signature of multiple 3D shapes. For example, a cube balancing on one of its corners will have exactly the same distribution as one resting flat on one face. Combining the three-dimensional similarity measure with an estimate of similarities of the 2D ground plates, estimated in the same way as *S* but in two dimensions only, reduces the odds of false positives when searching for similar shapes. In addition, other dimensions like the overall size of the properties can also be (re-)introduced to account for large deviations in scale.

Furthermore, human perceptions of similarity are likely to be a nonlinear function of *S*. For example, a decrease in *S* from a high 0.95 to 0.85 might change the perceived similarity of two buildings dramatically while moving from 0.35 to 0.25 might not. Translating *S* into a binary variable that classifies pairs of buildings as either similar or dissimilar accounts for non-linearities effectively.[2]

A similarity matrix ***WS*** contains the pairwise similarity estimate for all $n \times n$ pairwise combinations of buildings in a sample of size n . Each element $ws_{i,j}$ is defined to be 1 if buildings *i* and *j* are sufficiently similar in shapes (high *S* in 3D), ground plates (high *S* in 2D) and volumes, or 0 otherwise:

$$ws_{i,j} = \begin{cases} 1, \text{ if 3D-S}_{i,j} > a \wedge \text{2D-S}_{i,j} > b \wedge \frac{Volume_i}{Volume_j} \in [v_{low}, v_{high}] \\ 0 \end{cases} \quad (2)$$

The similarity estimate is symmetrical, as $ws_{i,j} = ws_{j,i}$. The volume of a building can be approximated by the average distance between random points on the surface of each building.

To achieve a similar/dissimilar classification that resembles the perceptions of shape similarity by humans as closely as possible, values for the parameters a, b, v_{low} and v_{high} are selected using responses from a dedicated web-based survey on shape similarity. In this survey, students are repeatedly presented pairs of 3D model visualizations of actual buildings and asked to classify them as either “rather similar” or “rather dissimilar”[3]. Drawing from this unique dataset of similarity perceptions, threshold values are selected that lead to a good fit between human classifications and the algorithm based classifications in ***WS***. With pre-compiled 2D- and 3D-shape signatures, a pairwise similarity matrix ***WS*** can be estimated fast and without consuming excessive computing resources even for large samples.

2 An either/or classification also resonates well with the vocabulary our language offers when describing similarity of shapes: We only have words for the extremes and cannot describe “somewhat similar” or other more nuanced degrees of similarity with single words.

3 Details on the survey design and all response data are available from the author on request.

Next, the economic relevance of shape analysis is investigated in an empirical estimation of property transaction values. Does the classification of property shapes help in explaining actual sales prices? In an ad-hoc test, all properties are assigned to 10 broad shape categories applying the k-means clustering algorithm to the shape similarity matrix ***WS***. The distributions of prices and hedonic attributes for properties across these shape clusters are compared.

Finally, a hedonic spatial error model (SEM) estimates the relationship between transaction prices for single family homes and set of explanatory variables including property characteristics, the time of transaction, the location of each building and the transaction prices of similar properties in a generalized method of moments (GMM) regression:

$$\ln(P_i) = \alpha + \boldsymbol{B}\,\boldsymbol{X}_i + \boldsymbol{G}\,\boldsymbol{Year}_i + \mu_i \quad (3)$$

$$\mu = \lambda_1 \boldsymbol{W} \mu + \lambda_2 \boldsymbol{WS}\, \mu + \epsilon$$

The natural logarithms of transaction prices P for building i is explained by a vector of hedonic attributes $\boldsymbol{X}_i$ and a vector of dummies variables $\boldsymbol{Year}_i$ for the year of transaction. The vectors ***B*** and ***G*** contain regression coefficients. The error terms μ are correlated with one another for nearby observations and for similar shapes.

The elements in the $n \times n$ spatial weight matrix ***W*** are defined to be 1 for all corresponding properties which are closer than 100 m and 0 otherwise. The coefficient λ_1 is expected to be positive, since properties that are geographically close share the same unobserved location amenities. In a similar spirit, the error terms of similar buildings (indicated by ***WS***) are expected to be correlated as well, since they share unobserved attributes. If the coefficient of shape correlation λ_2 is found to be significant and positive, then prices paid for properties that share the same shape are correlated beyond the factors explained by hedonics, time, or location.

LeSage (2014) advises to “avoid the pitfall of multiple weight matrices” in spatial models, since, among other concerns, covariances between multiple weight matrices are restricted to be zero (LeSage & Pace, 2011). When estimating Eq. (3), alternative specifications of ***WS*** are therefore tested that explicitly have a correlation of zero with the spatial weight matrix ***W***, circumventing any covariance restrictions.

Data

This paper relies on three sources of data. First, the Dutch city of Rotterdam provides a three-dimensional model of all buildings in the city[4], which has been calculated from surface scanning data captured from helicopters in April 2010. The accuracy of the spatial data is high: At least 30 points per square meter have been scanned in the city center and 65 percent of these points are within 10 cm of the true location (95 percent within 15 cm), and the confidence intervals around height estimates are even narrower (City of Rotterdam, 2015).

4 Available for download at http://www.rotterdam.nl/links_rotterdam_3d

The virtual representation of Rotterdam is distributed in the CityGML (Level of Detail 2) format, which is an open data model for the storage and exchange of three-dimensional city information. A building's shape is defined by a set of polygons, each representing a wall, part of a roof or the ground plate. One can compare this to building a model of a house by cutting two-dimensional shapes out of cardboard and gluing them together: any structure can be approximated but fine architectural nuances are lost. Demarcations of buildings that share walls have been added based on land registry records (City of Rotterdam, 2015). After dropping small structures with a ground plate of less than 3 m^2, 185,914 properties remain in the database.

Second, data on residential transactions in Rotterdam is acquired through the Association of Dutch Realtors (NVM). About 70% of all transactions in the Netherlands are facilitated by members of the NVM[5]. The NVM database contains 29.948 observations for Rotterdam in the years 2006-2013. For each sale, the sales price, the exact address and a basic set of quality attributes for the property like interior floor space, dwelling type, year of construction, number of bedrooms, number of bathrooms/WC and the building's volume are recorded. The street address can be translated into geographic coordinates using the geocoding service of the Dutch land register[6]. Based on these coordinates, sales can be matched with buildings in the 3D model.

Finally, the Dutch land registry maintains a national register of all buildings (*Basisregistraties Adressen en Gebouwen,* BAG) which offers information on the number of units within each building (among other attributes).

Combining the 3D data, the sales database and the building registry gives a sample of 6,717 transactions of individual structures that contain only one unit. Multi-unit buildings are excluded because their 3D shape cannot be assigned to individual sales reliably. Further, observations with extreme or wrongly coded values are deleted whenever the transaction price is below 30,000 EUR or above 1 million EUR, the value for interior floor space is below 30 m^2 or above 500 m^2, a lot size above 5,000 m^2 or an estimate of the building's volume below 30 m^3 or above 5,000 m^3 has been recorded. The cleansed final sample comprises of 6,126 transactions.

==== *Insert Figure 2 about here* ====

Figure 2 gives an overview of the spatial distribution of the sample within the borders of the Rotterdam municipality. The gray areas indicate all buildings from Rotterdam's 3D city map, including residential, industrial and commercial properties The black areas represent the final sample of single family homes for which transaction data is available in 2006-2013. Solid lines mark the official neighborhood boundaries. The majority of residential transactions can be found in the residential neighborhoods in Rotterdam proper in the east, while the west is dominated by harbor, infrastructure, warehouses and industrial properties.

---

5 https://www.nvm.nl/over_nvm/english.aspx

6 More information on the geocoding webservice is available at https://www.pdok.nl/nl/service/openls-bag-geocodeerservice

Results

For all buildings in Rotterdam, the 3D shape distributions and 2D shape distributions of ground plates are calculated and stored.[7] The computation of the shape distributions for a single building takes only a fraction of a second on a contemporary PC.

To verify that the suggested shape similarity measure *S* holds up in a real world application, the distribution of *S* for buildings that are known to be similar is compared to the overall distribution of *S*. Cube-shaped buildings can be easily identified as they have exactly four walls, a roof and a ground plate which are all squares of the same area. For 1,229 (out of 185,914) buildings, these conditions are met reasonably well. For all pairwise combinations of cube-buildings, the average value of *S* is 0.95, which is close to the ideal value of 1. In contrast the distribution of *S* for all buildings has a mean of 0.76. It is re-assuring that the difference in means between cube and non-cube buildings is large and statistically significant (t-value: 3,828). Overall, *S* passes the initial test of being able to tell similar from distinct shapes.

==== *Insert Table 1 about here* ====

==== *Insert Figure 3 about here* ====

The 6,129 shape distributions displayed in *Figure 3* exhibit substantial heterogeneity, indicating a large diversity in the shapes of the single-family homes in Rotterdam. At the same time, the darker areas in the figure show a clustering around typical distributions – despite all uniqueness, building exteriors appear to be variations of a limited number of typical architectural forms.

Similarity in shapes comes with similarity in hedonic attributes (Table 2). When dividing the buildings into 10 dominant shape clusters (the dark thick traces in Figure 3) using the k-means algorithm, stark contrasts in building attributes can be observed. For instance, Cluster 8 features the most affordable transaction price (EUR 194,000), the smallest average interior floor space (101 m^2), a low volume (281 m^3) and the most recent average year of construction (1975). It comprises almost exclusively of terraced houses (99.1%). Cluster 10, in contrast, features the highest share of detached homes (11.7%), the highest average values for sales price (EUR 365,000) and volume (450 m^3), and high values for interior floor size (153 m^2). The test for equal means in a one-way layout shows that the differences in cluster means are statistically significant, with F-values of 22 and higher (num. df = 9, denom. df = 2,163). Also, housing types are not equally distributed across clusters (X^2 =3333.4, df = 36).

==== *Insert Table 2 about here* ====

Finding a link between shapes and building characteristics corroborates the general idea of

7 Both the code to draw large numbers of random points from the exterior of a building model and the stored shape distributions are available from the author on request.

this paper: Shape information can be used as a proxy for observable – and more interestingly – otherwise unobservable building attributes. However, shape-related estimates remain difficult to interpret as they represent both an effect for a specific shape and jointly the contribution of unobserved hedonic variables correlated with specific shapes.

A further complication stems from the legendary rigor and detailedness of Dutch urban planning. Strict zoning laws in combination with economies of scale in large developments of multiple units with similar designs enforce a high degree of homogeneity in buildings' forms and appearances at block or street level. This combination of strong regulations and market forces induces high levels of spatial correlation in any measure of building shape for Rotterdam. The data support this expectation: The odds of observing buildings from identical shape clusters within 100 m off each other are 2.8 times higher than expected under the assumption of random spatial distributions. The same-shape joint count test statistics with nonfree sampling (Cliff & Ord, 1981; Upton & Fingleton, 1985) are highly significant. With such strong spatial correlations present in shape, strong spatial controls are indispensable in the subsequent analysis.

==== *Insert Table 3 about here* ====

Reassuringly, the automatic classification of buildings into similar and dissimilar pairs corresponds well with the perception of building similarity by human. Overall, 374 combinations of Rotterdam building models have been presented to students, who were then asked whether they would consider these buildings as being "pretty much the same" or "different"[8]. The automatic classification suggested in this paper can predict the human classifications well: 116 out of 125 combinations that have been classified as being "rather similar" by human survey respondents are also classified as similar in WS.[9] Only 9 (or 7%) are not. For pairs that are perceived as being different by humans, the match is a little lower: 193 out of 249 combinations flagged as "rather dissimilar" by humans are also considered dissimilar by the automatic classification (76%). A highly significant chi-squared statistic of 163 (with 1 degree of freedom) confirms that the automatic identification of similar buildings is highly correlated with human classifications.

==== *Insert Table 4 about here* ====

Table 5 presents coefficients for four independent GMM regressions. First a reduced version of Equation 3 is estimated in Model I, which explains the natural logarithm of transaction prices by dummies for the year of transaction and a traditional spatial weight matrix W (in which element $w_{i,j}$ is set to one if buildings i and j are less than 100 m apart, and 0 otherwise) only. The fit of this rudimentary model is surprisingly good (adj. R^2: 0.716) due to the fine-grained spatial weights capturing the variation in location amenities and building characteristics. The coefficient of spatial correlation, λ_W, is large (0.8) and statistically significant.

8 An example of the survey is presented in Appendix 1.

9 The 98th percentiles for 3D-S and 2D-S are used as thresholds a and b, and v_{low} and v_{high} are set to 0.83 and 1.2, respectively, when calculating WS.

==== *Insert Table 5 about here* ====

Adding a second weight matrix *WS* based on shape similarity (Model II) boosts the explanatory power further. The adj. R^2 reaches 0.74, reducing the unexplained variation by 8.5% ((1-0.74)/(1-0.716)=0.915). The coefficient of similar-shape correlation, λ_{WS}, is relevant in size (0.286) and also found to be significantly different from zero. In the sample, three quarter of all variation in transaction prices can be attributed to the overall market, location and similarities in shapes – all variables that can be remotely observed without on-site inspections and which represent low-hanging fruits for mass appraisal systems.

The coefficient estimates for the hedonic variables in III and IV do not surprise: Detached homes are valued most as all other types carry significant negative discounts. Terraced houses, for instance, are about 25% more affordable. Interior floor space and volume have positive elasticities, which add up to a little below 1. The elasticity of lot size is a low 0.039. The 1960s through 1980s vintages carry a significant discount, while newer homes command a premium over historic homes built before 1906.

Interestingly, the spatial correlation coefficient λ_W is the highest (0.805) in Model I, and drops sizably (to 0.699) after controlling for shape similarities in Model II. This suggests that spatial correlations in a traditional SER model does not exclusively capture micro-location related amenities but also includes a sizable share of property attribute information. Similarly, adding hedonic control variables directly (Model III) reduces the magnitude of the spatial correlation estimate (0.637), while the "purest" spatial correlation estimate (0.609) can most likely be observed in Model IV, which includes both direct hedonic variables and the indirect controls for unobserved variables through the shape similarity weight matrix *WS*.

Including hedonic control variables reduces the coefficient for shape similarity λ_{WS}, by more than half, again indicating that form and function are correlated. Still, even with strong hedonic controls, finely-grained spatial weight matrices and strong model fits (R^2 exceeding 0.8!), shape similarity correlation estimates remain statistically significant (p-value of 0.03).

Robustness tests find that buildings of similar shapes to exhibit a common structure in regression error terms even if observations are located far apart: setting all elements of *WS* for buildings that are less than 5 km from each other to 0 does not change the λ_{WS} estimates substantially. This is interesting for buyers and sellers of properties: Using shape information, one can identify relevant comparables, even if they are at the other end of town.

Finally, to rule out that the shape similarity matrix *WS* is not solely a disguised fixed effect for buildings by the same developer (which happen to have similar designs), the weight matrix is again manipulated. Assuming that buildings from in different vintages have been realized by different developers, all elements of *WS* for buildings built less than 15 years from each other are set to 0. Again, the λ_{WS} estimate remains robust in magnitude. As long as shapes are similar, differences in building age do not matter when looking for comparables.

Conclusion

This paper shows that it is not only feasible but also worthwhile to empirically analyze the shape of buildings. Existing research on property values has eschewed three-dimensional building models as an information source since these data do not come in convenient bite-size formats but have unwieldy "Big Data" properties. City-wide shape data sets tend to be massive in size, exceeding the computational limits of traditional regression-style empirics. Furthermore, the data is unstructured and needs interpretation before derived information on shapes can be linked to other property characteristics.

Extracting shape information is not "Big Data" wizardry, however. Condensing building models to shape distributions reduces the complexity while preserving sufficient information to estimate the degree of similarity between properties. These algorithm-based similarity estimates are good predictors of human perceptions of similarity (Table 4). This opens up new avenues of research not only in real estate finance and economics, but also in the domain of architecture, urban planning or sustainability. For instance, the market response to changes in building codes relating to e. g. the homogeneity of buildings at block level could be estimated directly, which could guide urban planners when drafting regulations.

In the property industry, shape information partially explains transaction prices of buildings and thereby improves the performance of automatic mass appraisal systems. Shape data can be collected remotely and consistently for large areas, which could help municipalities when determining and reassessing property tax levels. In real estate markets, shape information can also be used to identify relevant comparables and to arrive at more informed bid and asking prices. Buyers and sellers could quickly find buildings that are similar in shape – currently, buildings can only be searched according to location, size, number of bedrooms and other basic variables.

Finally, linking shape data to transaction prices could help developers and architects to evaluate the market potential of specific designs. It remains surprising that something as influential and omnipresent as the architecture of the buildings that constantly surround and house us is not frequently (re)evaluated in a coherent, large-scale and data-driven fashion.

References

Ahlfeldt, G. M., & Maennig, W. (2010). Substitutability and Complementarity of Urban Amenities : External Effects of Built Heritage in Berlin. *Real Estate Economics*, *38*(2), 285–323.

Ahlfeldt, G., & Mastro, A. (2012). Valuing Iconic Design: Frank Lloyd Wright Architecture in Oak Park, Illinois. *Housing Studies*, *27*(8), 1079–1099. doi:10.1080/02673037.2012.728575

Alam, N., Wagner, D., Wewetzer, M., Coors, V., & Pries, M. (2013). Towards automatic validation and healing of CityGML models for geometric and semantic consistency. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. II, pp. 27–29). Istanbul.

Asabere, P. K., Hachey, G., & Grubaugh, S. (1989). Architecture , Historic Zoning , and the Value of Homes. *Journal of Real Estate Finance and Economics*, *2*(3), 181–195. doi:10.1007/BF00152347

Boeters, R. (2013). *Automatic enhancement of CityGML LoD2 models with interiors and its usability for net internal area determination*. TU Delft.

Cardone, A., Gupta, S. K., & Karnik, M. (2003). A Survey of Shape Similarity Assessment Algorithms for Product Design and Manufacturing Applications. *Journal of Computing and Information Science in Engineering*, *3*(2), 109. doi:10.1115/1.1577356

City of Rotterdam. (2015). Rotterdam 3D. Retrieved from http://www.rotterdam.nl/rotterdam_3d

Cliff, A. D., & Ord, J. K. (1981). *Spatial Processes: Models & Applications*. London: Pion.

Florida, R., Mellander, C., & Stolarick, K. (2009). *Beautiful Places : The Role of Perceived Aesthetic Beauty in Community Satisfaction*.

Fuerst, F., McAllister, P., & Murray, C. B. (2011). Designer buildings: estimating the economic value of “signature” architecture. *Environment and Planning A*, *43*(1), 166–184. doi:10.1068/a43270

Jensen, J., & Cowen, D. (1999). Remote sensing of urban suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, *65*(5), 611–622.

Kolbe, T. H. (2009). Representing and Exchanging 3D City Models with CityGML. *3D Geo-Information Sciences*, 15–31. doi:10.1007/978-3-540-87395-2_2

Kwan, M.-P., & Lee, J. (2005). Emergency response after 9/11: the potential of real-time 3D GIS for quick emergency response in micro-spatial environments. *Computers, Environment and Urban Systems*, *29*(2), 93–113. doi:10.1016/j.compenvurbsys.2003.08.002

Leichenko, R., Coulson, E., & Listokin, D. (2001). Historic Preservation and Residential Property Values: An Analysis of Texas Cities. *Urban Studies*, *38*(11), 1973–1987. doi:10.1080/00420980120080880

LeSage, J. P. (2014). What regional scientists need to know about spatial econometrics. *Available at SSRN*, 1–31. doi:10.2139/ssrn.2420725

LeSage, J. P., & Pace, R. K. (2011). Pitfalls in higher order model extensions of basic spatial regression methodology. *Review of Regional Studies*, *41*(1), 13–26.

Listokin, D., Listokin, B., & Lahr, M. (1998). The contributions of historic preservation to housing and economic development. *Housing Policy Debate*, *9*(3), 431–478. doi:10.1080/10511482.1998.9521303

Moorhouse, J. C., & Smith, M. S. (1994). The Market for Residential Architecture: 19th century Row Houses in Boston's South End. *Journal of Urban Economics*, *35*(3), 267–277. doi:10.1006/juec.1994.1016

Nagel, C., Stadler, A., & Kolbe, T. H. (2009). Conceptual Requirements for the Automatic Reconstruction of Building Information Models from Uniterpreted 3D Models. In *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*.

Nichol, J., & Wong, M. S. (2005). Modeling urban environmental quality in a tropical city. *Landscape and Urban Planning*, *73*(1), 49–58. doi:10.1016/j.landurbplan.2004.08.004

Ohbuchi, R., Minamitani, T., & Takei, T. (2005). Shape-similarity search of 3D models by using enhanced shape functions. *International Journal of Computer Applications in Technology*, *23*(2/3/4), 70. doi:10.1504/IJCAT.2005.006466

Osada, R., Funkhouser, T., Chazelle, B., & Dobkin, D. (2001). Matching 3D models with shape distributions. In *Proceedings International Conference on Shape Modeling and Applications* (pp. 154–166). IEEE Computer Society. doi:10.1109/SMA.2001.923386

Rakkolainen, I., & Vainio, T. (2001). A 3D City Info for mobile users. *Computers & Graphics*, *25*(4), 619–625. doi:10.1016/S0097-8493(01)00090-5

Ranzinger, M., & Gleixner, G. (1997). GIS Datasets for 3D Urban Planning. *Computers, Environment and Urban Systems*, *21*(2), 159–173.

Shiode, N. (2001). 3D Urban Models: Recent Developments in the Digital Modeling of Urban Environments in Three-Dimensions. *GeoJournal*, *52*(3), 263–269.

Sullivan. (1896). The Tall Office Building Artistically Considered. *Lippincott's Magazine*.

Tangelder, J. W., & Veltkamp, R. C. (2008). A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*, *39*(3), 441–471. doi:10.1007/s11042-007-0181-0

Upton, G., & Fingleton, B. (1985). *Spatial data analysis by example: point pattern and* *quatitative data*. Chhichester, UK: John Wiley.

Wolff, M., & Asche, H. (2009). Towards Geovisual Analysis of Crime Scenes – A 3D Crime Mapping Approach. In M. Sester, L. Bernard, & V. Paelke (Eds.), *Advances in GIScience.* (pp. 429–448). Berlin and Heidelberg: Springer. doi:10.1007/978-3-642-00318-9

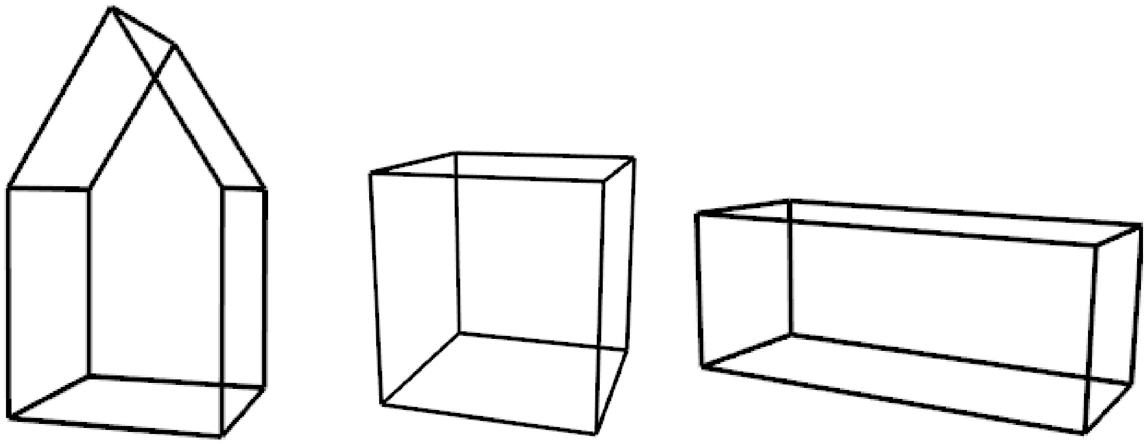
Wu, H., He, Z., & Gong, J. (2010). A virtual globe-based 3D visualization and interactive framework for public participation in urban planning processes. *Computers, Environment and Urban Systems*, *34*(4), 291–298. doi:10.1016/j.compenvurbsys.2009.12.001

Yasumoto, S., Jones, A. P., Nakaya, T., & Yano, K. (2011). Computers , Environment and Urban Systems The use of a virtual city model for assessing equity in access to views. *Computers, Environment and Urban Systems*, *35*(6), 464–473. doi:10.1016/j.compenvurbsys.2011.07.002

Yu, S.-M., Han, S.-S., & Chai, C.-H. (2007). Modeling the value of view in high-rise apartments: a 3D GIS approach. *Environment and Planning B: Planning and Design*, *34*(1), 139–153. doi:10.1068/b32116

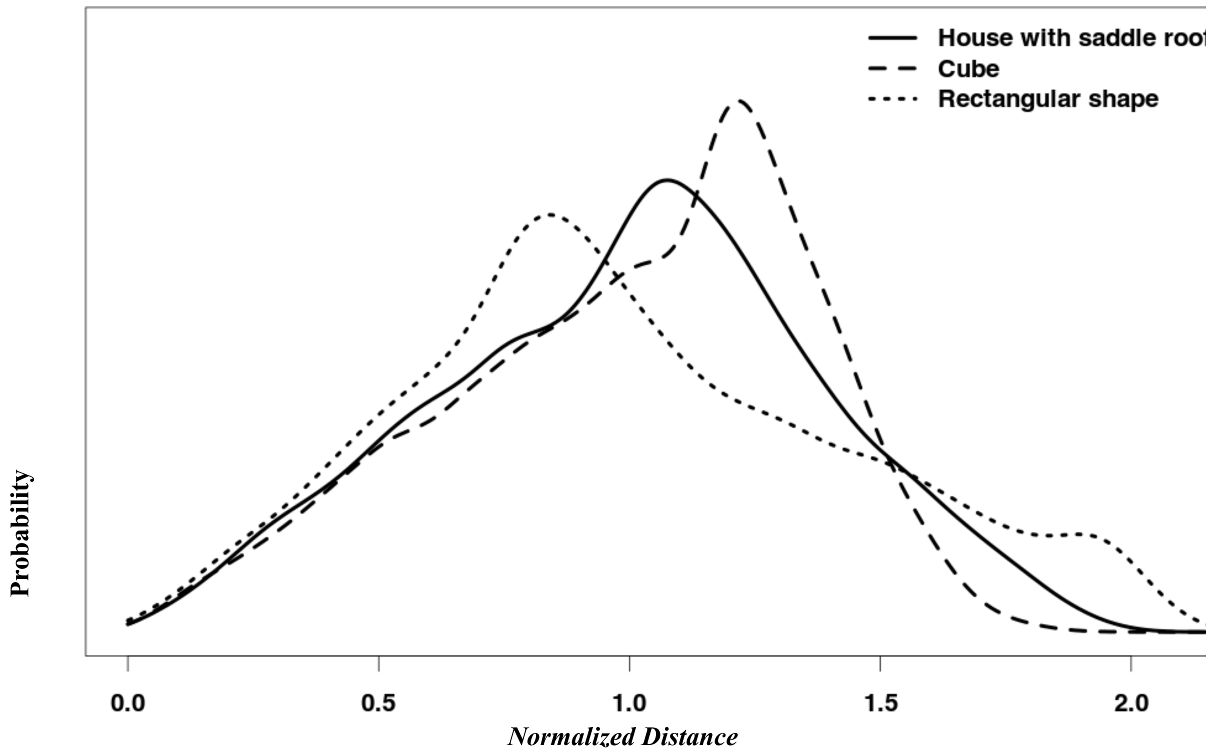
Figures and Tables

Figure 1: Basic solid geometries and their representation as a shape distribution



House with saddle roof | Cube | Rectangular form

Estimated kernel density functions for distances between randomly selected points on hull of shape

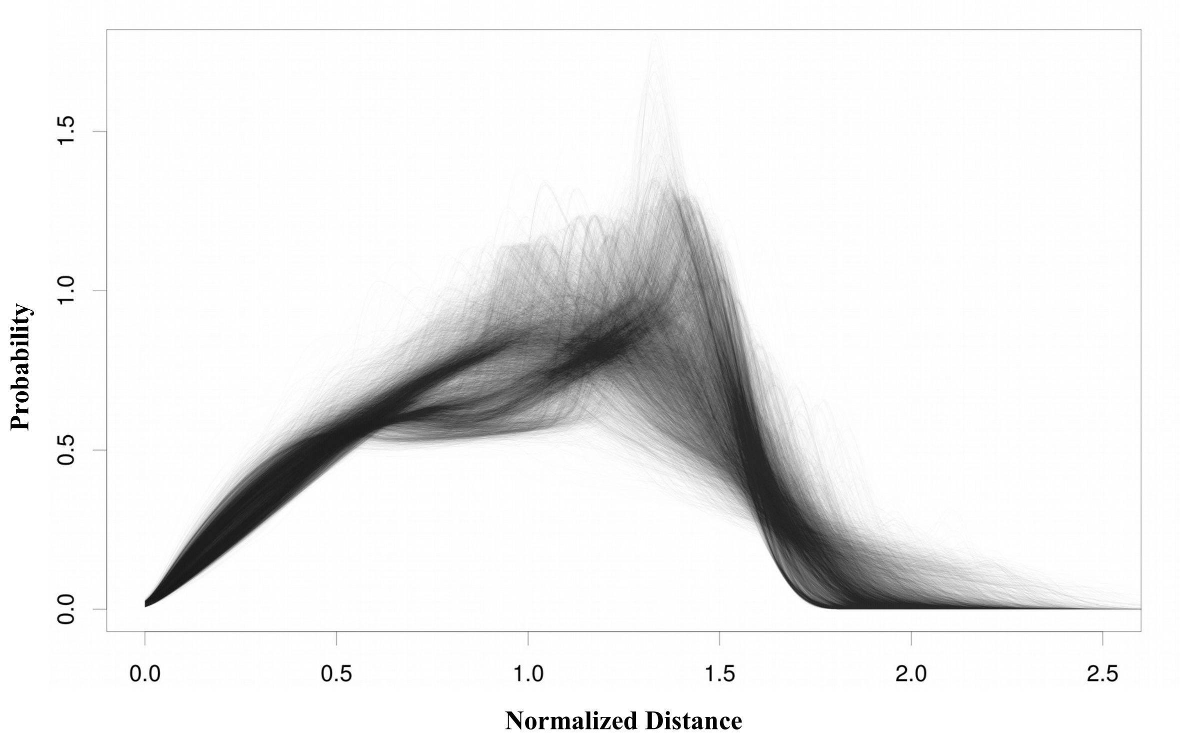


Notes: The distributions are normalized by dividing all distances by average distance per shape before estimation of the density functions. The density functions of three basic shapes show very distinct profiles with cubes having the most pronounced peak. Rectangular shapes exhibit flatter space distributions with a hump in the right tail.

Figure 2: Spatial distribution of buildings and transactions in city of Rotterdam

Notes: The gray areas indicate all buildings from Rotterdam's 3D city map. The black areas represent the final sample of single family homes for which transaction data is available in 2006-2013. Solid lines mark the official neighborhood boundaries. The majority of residential transactions can be found in the residential neighborhoods in Rotterdam proper in the east, while the west is dominated by harbor, infrastructure, warehouses and industrial properties.

Figure 3: Shape distributions of single family homes in sample (Rotterdam, 2006-2013)



Notes: The shape distributions of all 6,129 buildings in the sample exhibit substantial heterogeneity, indicating a large diversity in the shapes of the single-family homes in Rotterdam. At the same time, the darker areas in the figure show a clustering around typical distributions – despite all uniqueness, building exteriors appear to be variations of a limited number of typical architectural forms.

Table 1: Distribution of Shape similarity S across all Rotterdam buildings

	Min	*1st Quantile*	*Median*	*Mean*	*3rd Quantile*	*Max*	*SD*
All buildings	0.21	0.69	0.77	0.76	0.84	1.00	0.11
Cube-shape buildings	0.60	0.94	0.96	0.95	0.97	1.00	0.03

Notes: The pairwise shape similarity measure *S* is calculated for all combinations of 185,914 buildings in Rotterdam. The distribution of similarity values clearly differs from the distribution for cube-shaped buildings, which display higher levels of similarity. Overall, 1,229 buildings are classified as having a cube shape: they consist of exactly 4 walls, a roof and ground plate which are all squares and of similar size. The difference in means between non-cube and cube-shape buildings is statistically significant (t-value = 3,828).

Table 2: Mean values for hedonic attributes and distribution of house types across shape clusters

Cluster	*Count*	*Mean*								*% House type*				
		Price (in '000)	*Int. space (m^2)*	*Lot (m^2)*	*Volume (m^3)*	*points Dist. Betw.*	*Num. rooms*	*Num. floors*	*Year of constr.*	*Detached*	*Semi-detaced*	*Corner*	*Linked-detached*	*Terraced*
1	319	285	148	128	428	7.7	5.2	3.0	1963	0.3%	0.3%	0.6%	2.2%	96.6%
2	581	303	151	127	439	7.7	5.5	3.2	1956	0.2%	0.2%	0.9%	0.7%	98.1%
3	406	286	131	208	392	7.3	4.6	2.5	1960	7.1%	15.8%	21.2%	5.9%	50.0%
4	860	322	140	275	403	6.6	5.1	2.8	1959	7.8%	19.3%	50.5%	2.3%	20.1%
5	491	248	123	165	351	6.4	5.1	3.0	1961	0.0%	0.2%	1.0%	0.8%	98.0%
6	800	301	141	181	406	6.8	5.3	3.1	1961	0.1%	7.0%	42.5%	1.9%	48.5%
7	716	259	133	151	378	7.1	4.8	3.0	1967	0.3%	1.0%	11.7%	2.5%	84.5%
8	319	194	101	179	281	6.4	4.1	2.1	1975	0.0%	1.3%	6.6%	0.3%	91.8%
9	785	284	137	146	396	6.9	5.4	3.2	1958	0.0%	0.1%	0.3%	0.5%	99.1%
10	849	365	153	273	450	7.0	5.1	2.8	1968	11.7%	24.4%	38.9%	6.8%	18.3%
Total	**6,126**	**294**	**138**	**192**	**400**	**7.0**	**5.1**	**2.9**	**1962**	**3.3%**	**8.3%**	**21.4%**	**2.5%**	**64.54%**
		One-way analysis of equal means, F-value								Chi-squared test of independence				
		89	96	62	96	159	64	145	22	X^2 =3333.4, df = 36				

Notes: Properties are grouped based on their shape distributions using the k-means clustering algorithm (k=10). The values for core hedonics differ across shape clusters. The equality of means can be rejected, with all F values exceeding 22 in one-way analysis (num. df = 9, denom. df = 2,163). Also, housing types are not equally distributed across clusters (X^2 =3333.4, df = 36).

Table 3: Spatial correlation in building shapes

Shape cluster	*Same-shape statistic*	*Expectation*	*Variance*	*P-Value*
1	42.48	8.48	1.18	0.00
2	99.19	28.19	3.60	0.00
3	60.71	13.75	1.86	0.00
4	131.95	61.79	7.21	0.00
5	85.88	20.12	2.64	0.00
6	101.44	53.46	6.36	0.00
7	119.51	42.82	5.24	0.00
8	57.81	8.48	1.18	0.00
9	153.00	51.48	6.15	0.00
10	132.11	60.22	7.05	0.00

Notes: Joint count tests under nonfree sampling (Cliff & Ord, 1981) suggest that buildings of similar shapes tend to be close to each other. The odds of observing buildings from identical shape clusters within 100 m off each other are 2.8 times higher than expected under the assumption of random spatial distributions. The same-shape statistics are statistically highly significant.

Table 4: Automatic vs. human classification

		Automatic classification (WS)	
		Different	*Similar*
Classification by survey respondents	*Different*	193	56
	Similar	9	116

Notes: Overall, 374 pairs of buildings have been classified by human subjects as either being similar or different. The corresponding values in similarity matrix WS show that the automatic shape comparison leads to classifications that are, on average, similar to classifications by humans. $X^2 = 162.81$, df = 1, p-value < 0.001.

Table 5: Regression coefficient estimates

	I				II				III				IV			
Variable	*Coeff.*	*SE*	*P-Val.*		*Coeff.*	*SE*	*P-Val.*		*Coeff.*	*SE*	*P-Val.*		*Coeff.*	*SE*	*P-Val.*	
Const.	12.510	0.015	0.000	***	12.631	0.016	0.000	***	8.399	0.069	0.000	***	8.475	0.072	0.000	***
Year of sale (vs. 2006)																
2007	0.013	0.010	0.211		0.019	0.010	0.047	*	0.018	0.008	0.019	**	0.019	0.008	0.016	**
2008	0.047	0.011	0.000	***	0.049	0.010	0.000	***	0.044	0.008	0.000	***	0.046	0.008	0.000	***
2009	-0.003	0.011	0.766		0.003	0.011	0.748		0.006	0.009	0.461		0.008	0.009	0.357	
2010	-0.005	0.012	0.668		0.000	0.011	0.993		0.008	0.009	0.382		0.009	0.009	0.311	
2011	-0.014	0.012	0.257		-0.012	0.011	0.282		0.006	0.009	0.525		0.006	0.009	0.505	
2012	-0.082	0.012	0.000	***	-0.079	0.011	0.000	***	-0.050	0.009	0.000	***	-0.050	0.009	0.000	***
2013	-0.122	0.016	0.000	***	-0.131	0.016	0.000	***	-0.102	0.013	0.000	***	-0.106	0.013	0.000	***
Type (vs. detached)																
Corner									-0.207	0.016	0.000	***	-0.212	0.016	0.000	***
Terraced									-0.269	0.016	0.000	***	-0.274	0.016	0.000	***
Semi-det.									-0.088	0.017	0.000	***	-0.095	0.017	0.000	***
Linked-det.									-0.216	0.021	0.000	***	-0.226	0.021	0.000	***
ln(int. space m^2)									0.741	0.015	0.000	***	0.725	0.016	0.000	***
ln(lot size m^2)									0.039	0.003	0.000	***	0.039	0.003	0.000	***
ln(Volume)									0.247	0.036	0.000	***	0.252	0.038	0.000	***
Year of construction (vs. before 1906)																
1906-1930									-0.013	0.016	0.413		-0.014	0.015	0.364	
1931-1944									0.018	0.017	0.303		0.006	0.017	0.738	
1945-1959									0.025	0.019	0.192		0.018	0.020	0.365	
1960-1970									-0.054	0.021	0.010	**	-0.055	0.021	0.010	**
1971-1980									-0.090	0.021	0.000	***	-0.080	0.021	0.000	***
1981-1990									-0.071	0.019	0.000	***	-0.063	0.019	0.001	***
1991-2000									0.094	0.019	0.000	***	0.086	0.019	0.000	***
Yoc ≥ 2001									0.102	0.020	0.000	***	0.095	0.021	0.000	***
Yoc unknown									0.007	0.187	0.972		-0.056	0.186	0.763	
λ_W spat.	0.805		0.000	***	0.699		0.000	***	0.637		0.000	***	0.609		0.000	***
λ_{WS} shape					0.286		0.000	***					0.120		0.030	**
R^2	0.716				0.740				0.838				0.840			
Adj. R^2	0.716				0.740				0.837				0.839			

Notes: N=6,126.

Appendix 1

Do these buildings look similar?

If you stood in front of the two houses, would you think "they are pretty much the same" or "they are different"?

Missing walls are not a problem - just imagine how the building would look like if all walls were present.

Your name (voluntary): Your Name

Buildings look similar | Buildings look rather different

You can rotate each of the shapes to get a better view.