Advanced Space-Time Analysis: Constructing a Real Estate Price Index

Darren K. Hayunga[*] and Alexander Kolovos[†]

November 27, 2015

ABSTRACT

Real estate data exhibit autocorrelation and heterogeneity across both space and time. The literature is beginning to advance methods that account for the four components. This article accordingly introduces the Bayesian Maximum Entropy (BME) method to real estate analysis. In addition to controlling for spatiotemporal autocorrelation and heterogeneity, BME allows for probabilistic and missing data and does not require a known distribution as assumed by linear and log-likelihood techniques. We apply BME to a dataset of house prices, which illustrates the use of the technique for developing a price index on a small geographical area.

[*] Department of Insurance, Legal Studies, and Real Estate, University of Georgia, Athens, Georgia, USA. E-mail: hayunga@uga.edu

[†] SpaceTimeWorks, LLC, San Diego, California, USA. E-mail: alexander.kolovos@spacetimeworks.com

1. Introduction

Over the last two decades, the spatial literature investigating real property assets has exhibited a growing interest in incorporating the temporal vector. Most papers in the real estate literature use a form of the general spatial linear model with a spatiotemporal proximity or weight matrix as the mechanism to model time and separation distance (see Appendix A for introduction to the general spatial linear model). [1,2] The techniques are extensions of the standard spatial weight matrix in specifications like the spatial autoregressive model (SAR) or the spatial error model (SEM).[3] Modeling both space and time is welcomed innovation given the development of better georeferenced datasets that include longer times series of variables of interest.

But recent literature by Pinkse and Slade (2010), McMillen (2012) and Gibbons and Overman (2012) effectively argue that the standard spatial linear models have significant issues, which are not mitigating by incorporating the temporal component in a weight matrix. McMillen (2012) explains how the use of a spatially lagged dependent variable is a form of linear smoothing and it would be surprising not to find a spatial

---

[1] Papers that use the spatiotemporal weight matrix include the first studies by Can and Megbolugbe (1997) and Pace, et al. (1998, 2000) as well as extensions and applications by Tu, Yu, and Sun (2004), Sun, Tu, and Yu (2005), LeSage and Pace (2009), Smith and Wu (2009), Nappi-Choulet and Maury (2009), Alberto et al. (2010), and Huang, Wu, and Barry (2010). Nappi-Choulet and Maury (2011) is a notable extension that considers autocorrelation and heterogeneity across both space and time.

[2] Another strand of literature considers space and time in dynamic spatial panel data models See Elhorst (2012) for a review of this literature. These models are not typically helpful for real property studies because datasets are generally not longitudinal thus multiple selling prices of an individual property are not collected over a period of time. A repeat sales dataset could build a panel but the literature seems reluctant to employ repeat sales datasets outside of price index construction. Lastly, Dube and Legros (2013a, 2013b) are careful to model real estate prices as a pooled dataset, also using the spatiotemporal weight matrix.

[3] The notable exceptions are Gelfand et al. (1998, 2003).

relation using the standard spatial linear models. Gibbons and Overman (2012) note that the spatial econometric methods are not well suited for causal analysis. Pinkse and Slade (2010) describe the many issues with the spatial autoregressive model, which include the implausible normality assumption, endogeneity, and nonlinearity. They also note that modeling of the entire dependence structure of a spatial dataset is improbable using linear methods.

To address these issues and others, this article introduces the Bayesian Maximum Entropy (BME) method. Over the past fifteen years, BME has been employed in the physical sciences as a multidisciplinary technique examining studies sponsored by the National Institutes of Health and National Science Foundation. BME applications include ambient air pollution, arsenic mitigation, air pollutants, and the risk of cognitive disorders in the elderly. We consider its application to housing data.

BME offers many benefits that are not found in regression-based models. The first is that BME does not require any assumption regarding the shape of the underlying probability distribution and, accordingly, does away with the normality assumption. Further, the distribution–whether Gaussian or not–is often unknown to the econometrician; otherwise there is not a need for a spatial weight matrix. Common estimation procedures like log-likelihood functions still require that the researcher knows what distribution to apply to the data. But as McMillian (2012) and Gibbons and Overman (2012) point out, uncertainty about functional forms and lack of information on the true spatial weights

means alternative strategies are more appropriate. The paradox is that the use of the general spatial model is to explain the model structure.[4]

Another notable benefit of BME is that it can incorporate all possible relevant informative sources about a phenomenon without being limited to observed data. Such secondary data can be soft intervals or data provided as probabilistic distributions that cannot otherwise be used by other methods. A further useful result of this feature is that BME can reach beyond the estimation of determinant factors in hedonic models or use information from hedonic mode as an input.

We demonstrate the employment of secondary data by using tax assessment values to supplement transaction prices in building a housing price index. Assessment values provide much more coverage as they generally are available for every property within a geographical area of interest. However, assessment values measure transaction prices with error. The BME method can properly account for the uncertainty. And the granularity can be at any level, such as a submarket or neighborhood level. The limitation is only set by the available data.

A third advantage to BME is mitigation of the Modifiable Areal Unit Problem (MAUP) cited in the regional economics and the economic geography literature (Openshaw and Taylor 1979). The MAUP is a statistical bias due to the shape, size, and position of the areal unit when point-based measures of spatial phenomena are aggregated into clusters. Some current methods assume a particular cluster shape of the data but the

---

[4] The GMM estimator of Kelejian and Prucha (1999) also does not require any distribution assumption. However, they still use the spatial weight matrix, which assumes that the spatial weight matrix is known. The standard use of the spatial weight matrix is to estimate the marginal effects of exogenous explanatory variables. Also, the explanatory variables are assumed to be exogenous and fixed across time.

literature is quickly showing that geographical functions are not limited to boundaries. BME mitigates the MAUP because it dynamically measures separation of distance and time.

Another benefit is in the geostatistical nature of the BME method that enables us to model both macroscopic and microscopic data. Cressie (1993) and Pace and Gilley (1997) recognize that spatial patterns can be large-scale and small-scale. That is, certain effects are macroscopic that manifest as mean surface trends in space-time, while others are microscopic, which can be modeled as residuals that sit on top of mean trends. The dual aspects are certainly true for real estate, which functions in both a market that has, for instance, mortgage rates set nationally as well as a local employment market.

The logical structure of this study is as follows. We briefly describe the BME method and stochastic framework in the next two sections. We detail the data and empirical design of our study in Section 4. The results of our study are in Section 5, which includes comparison to ordinary kriging. Section 6 provides summary discussion.

2. A Brief Overview of BME

BME integrates multi-sourced types of information in three consecutive stages known as the structural (or prior) stage, the specificatory stage, and the posterior (or integration) stage. In the prior stage, one considers all relevant available general knowledge bases G-KB which comprise theoretical and empirical expressions of the $X(\boldsymbol{p})$ characteristics. An example of G-KB is a surface trend, denoted by the S-TRF expectation $m_x(\boldsymbol{p}) = \overline{X(\boldsymbol{p})}$, which represents systematic spatiotemporal patterns at scales larger than the study scale. Another example of G-KB is expressing spatiotemporal variability between attribute points via the 2-point statistic of covariance between pairs of attribute

points; for a pair of transaction price space-time points $\boldsymbol{p}$ and $\boldsymbol{p}'$ their covariance is

$c_x(\boldsymbol{p}, \boldsymbol{p}') = \overline{[X(\boldsymbol{p}) - m_x(\boldsymbol{p})][X(\boldsymbol{p}') - m_x(\boldsymbol{p}')]}$. The surface trend and covariance

function express the first- and second-order statistical moments of an S-TRF. The BME

framework provides solid foundation to account for knowledge of higher order moments

and multipoint statistics, too (Christakos, 2000). A different example of *G*-KB can be a

physical law or an empirical model that is applicable to the phenomenon under study

(Kolovos, 2002; Christakos *et al*., 2004; Kolovos *et al*., 2013). On the basis of the *G*-KB,

the BME prior stage results in a map of probability density functions (PDFs) that quantify

the distributions f_G of the prior probabilities. By designating an individual S-TRF

realization as χ, the prior PDF is defined by

$$f_G(\chi, \boldsymbol{p})d\chi = Prob[\chi \leq X(\boldsymbol{p}) \leq \chi + d\chi]$$

A concise description of the prior stage is given by the first fundamental BME equation:

$$\int d\chi (g - \bar{g})e^{\boldsymbol{\mu}^T g} = 0 , \tag{1}$$

where g is a vector of N_c g_α-functions, $\alpha = 1, 2, \dots, N_c$, that represents the *G*-KB, and $\boldsymbol{\mu}$

is a vector of μ_α-coefficients that depends on the space-time coordinates and associates

with g. These coefficients express the relative significance of each g_α-function, and the

prior PDF f_G is fully defined by solving Equation (1) with respect to μ_α, $\alpha = 1, 2, \dots, N$.

The specificatory stage is an assessment of the observed data, $\boldsymbol{\chi}_{data}$. The $\boldsymbol{\chi}_{data}$

values are recorded at a given set of m space-time points $\boldsymbol{p}_i$, $i = 1, \dots, m$, and they

constitute the specificatory knowledge bases (*S*-KB). The *S*-KB might consist of hard data

$\boldsymbol{\chi}_{hard}$, soft data $\boldsymbol{\chi}_{soft}$, or a combination of both as in the present study.

In the final BME stage of integration, the S-KB updates the G-KB. Blending the two types of knowledge bases yields the total given information $K = G \cup S$ about the stochastic process. The second fundamental BME equation illustrates the blending as

$$\int d\chi \xi_s e^{\mu^T g} - A f_K(\boldsymbol{p}) = 0, \qquad (2)$$

where ξ_s is an operator that represents the S-KB, A is a normalization parameter, and f_K is the posterior PDF at each spatiotemporal point $\boldsymbol{p}$ where prediction is sought. The subscript in f_K indicates that the posterior PDF is computed on the basis of the total information K. f_K is computed through Equation (2) and provides a complete stochastic description of the variable(s) of interest at a series of specified prediction locations $\boldsymbol{p}$. A variety of predicted attribute measures can be extracted from f_K to match the goals of the intended study, such as the distribution mean or mode. Equation (2) shows that BME uses a nonlinear predictor, which in the absence of soft data reduces to the kriging predictor (Christakos, 2000). In our housing transaction prices study that follows, we predict the BME mean and compare it to the corresponding ordinary kriging prediction.

3. Housing Prices Stochastic Framework

We use a stochastic representation of the housing transaction prices evolution in space S and time T, which enables us to account for uncertainty that can stem from ontological (or aleatoric) causes and epistemic limitations. In our analysis, the transaction prices are represented as a spatiotemporal random field (S-TRF; Christakos, 1991) $X(\boldsymbol{p})$, where $\boldsymbol{p} = (\boldsymbol{s}, t)$, $\boldsymbol{s} = (s_1, s_2)$ is the spatial location vector with coordinates s_1 and s_2, and t is the time variable. Price uncertainty manifests as an ensemble of all possible realizations of $X(\boldsymbol{p})$ values across space-time points $\boldsymbol{p}$. The S-TRF assigns to each of these realizations a probability that depends on $\boldsymbol{p}$, and is fully defined when the distribution of

the transaction price random variable $x(\boldsymbol{p})$ is known at all locations $\boldsymbol{p}$ in the spatiotemporal continuum $\mathcal{E} = S \times T$ (e.g., Christakos, 1992; Christakos and Hristopulos, 1998). In the S-TRF context, inter-point distances in $\mathcal{E}$ are defined on the basis of a joint space-time metric $d\boldsymbol{p}$. For a spatial distance $|d\boldsymbol{s}|$ and a time interval dt, the segment length $|d\boldsymbol{p}|$ between two spatiotemporal locations $\boldsymbol{p}_1$ and $\boldsymbol{p}_2 = \boldsymbol{p}_1 + d\boldsymbol{p}$ on the Euclidian plane is

$$|d\boldsymbol{p}| = \sqrt{d\boldsymbol{s}^2 + (v\,dt)^2},$$

where v is a spatiotemporal metric coefficient (Christakos et al., 2000). The kriging and the BME methodologies are used to predict unknown values $\boldsymbol{\chi}_k$ of the transaction prices S-TRF at k selected space-time locations.

4. Empirical Design

We model house transaction prices from a portion of Tarrant County in Texas. The study area is northeast of the Fort Worth central business district and measures 4,600 m by 9,000 m. Our sample period extends from January 2009 to December 2012. The observed transaction prices are the hard data of our study. We consider a temporal resolution of 1 month, hence all transactions that occur within a given calendar month are assigned as data attributed to this month.

Apart from the hard data, we also use tax assessed values that are publicly available for almost every property in the US. Our test method accounts for the fact that most market participants do not have access to proprietary data from local realtor databases but can obtain tax valuations from local county assessor offices. To obtain the soft data observations used in the subsequent BME analysis, we regress transaction prices against the assessed values and other property characteristics that are public information in the tax

assessor's database. Appendix B provides variable definitions and Table 1 presents the parameter estimates and heteroscedasticity-consistent errors. The model statistics indicate a strong fit, primarily due to the close relationship between the tax-assessed values relative to transaction prices. Figure 1 shows the observed prices compared to the predicted values.

The tax assessed estimates are not hard data because the model produces values with nontrivial uncertainty. We use the uncertainty in the estimated tax assessed values to define an additional knowledge base of interval soft data that are used by BME to complement the hard data set and improve prediction. Specifically, after fitting the model, we compute the upper and lower 95 percent confidence interval for an individual estimation. As compared to the 95 percent interval for the expected mean value of the dependent variable, the individual estimation interval is more conservative because it includes both the variance of the error as well as the variance of the parameter estimates. We then compute the mean of the both the lower and upper confidence interval and use it to specify the price model uncertainty as a percentage of the model estimates values. For each model-estimated value this percentage determines the lower and upper bound of this value. The bounds constitute an interval soft datum where the transaction price variable can take any value within this range.

Tarrant County makes new assessments values publicly available around the end of each calendar year. Our study considers realistically that a given year's worth of tax assessed values are only available to BME in the January of the following year. Hence, in our analysis between 2009 and 2012 we use soft data that reflect tax assessment records from years 2008 through 2011. Practically, this reflects in our study as having soft data available only in the month of January every year. This is a sensible modeling perspective

for the current analysis, because it reflects our understanding that the same data are also being made available to the entire housing market at the same time and no earlier. In turn, one can plausibly expect transaction prices in the following months to be affected and adjusted on the basis of the newly released information every January. Therefore, by considering the soft data of an entire year to be associated with the following January, our model accounts for this real-life constraint about tax assessed data availability in the process that governs housing transaction prices.

In our spatial area of interest, a total of 1,901 homes have both transaction and assessed prices. Figure 2 shows this area in context within the entire Tarrant County, Texas. The study area is highlighted by the aggregate hard data sample of transaction prices that appears as a collection of points in a rectangular towards the northeastern part of the county. Our study examines three different scenarios:

Case A: The sample includes all transaction prices within our initial focus area plus an extensive number of tax assessment values. So that our results generalize to the more typical market, we initially restrict transaction prices to a low of $50,000 and a high of nearly $1 million. The data sample for this case includes all transactions without regard for property specifics.

Case B: Using the same spatial grid as in Case A, we account for market stratification to perform a more sensible analysis in a market sample that exhibits less variability. Specifically, a home with 2 bedrooms, 1 bathroom, and no garage is in a different market stratum than a home with 4 bedrooms, 3 bathrooms, and a 3-car garage. In Case B we refine the sample to observations in the range between the 25th and 90th percentile of transaction prices. Homes in the Case B sample consequently

have 2-3 bathrooms, 3-4 bedrooms, 1-2 garage spaces, and 1,600 – 2,600 square feet, and transaction prices range from $100,000 to $250,000.

Case C: Goodman (1977) defines neighborhoods as "small urban areas with... a common set of socioeconomic effects". Beyond the boundaries of neighborhoods and subdivisions, homes can vary considerably. Given some of the new urbanism development in the marketplace, there can exist material variability even within a subdivision or neighborhood. Case C investigates transaction price at the neighborhood scale by using the same focused market segment observations as Case B; the difference between Cases B and C is that the latter concentrates on a smaller scale where the grid is twice as dense as in Cases A and B.

In spatiotemporal prediction, the existence of surface trends can obscure underlying spatial and temporal correlations in the current scale of analysis (Chilès and Delfiner, 1999). Specific to our data, the notion of surface trend refers to a base price component of house values as well as macroscopic economic conditions. We initially remove the trend components and use the subsequent residual prices for the prediction process. To obtain the residuals, we employ a moving window exponential filter that leaves transaction price fluctuations around 0. This filter uses ranges that extend spatially to approximately a quarter of the size of the spatial domain and temporally to roughly 12 months to account for price evolution in the market. For illustration, the descriptive statistics tables in each case report the part of the total variability in the data which can be linked to the surface trend; this is quantified by the percentage $V_{tr} = (1 - \frac{V_R}{V_T})$ where V_T is the total variance prior to trend removal, and V_R is the residual variance for each of the examined cases in our analysis. We retain the spatiotemporal trend values until the end of the prediction

11

process and then restore them to obtain the actual transaction price at all space-time locations.

The residuals are used to yield empirical estimates of the spatiotemporal covariance function that reveals the underlying structure of the transaction price residual S-TRF. A covariance model is then fitted to the empirical estimate via the efficient Bound Optimization By Quadratic Approximation method (Powell, 2009). The fitted model provides our assessment of the underlying correlation structure to use at the prediction stage. Covariance models must adhere to permissibility conditions (Christakos, 1992) and can be simple functions in single or nested structures, or even more elaborate algebraic constructs derived from general principles or relevant physical-based models (Kolovos et al., 2004).

For the prediction analysis we combine the modeled spatiotemporal covariance with the transaction price data. Both BME and ordinary kriging consider the same G-KB that comprises of the transaction price surface mean trend and covariances. However, only BME can account for soft data rigorously, when available. Kriging has been shown to handle uncertain information inefficiently (e.g., Savelyeva *et al*., 2010), and to this end our kriging analysis uses only the hard data observations. Kriging yields predicted transaction price residuals, whereas BME produces their posterior PDF. For mapping purposes, we illustrate the BME mean. We restore the formerly removed surface trends onto the predicted transaction price residuals to create transaction price maps across the study domain for every month in the study period. The outcome is a complete series of predicted residential transaction price values on predefined grid nodes. Animation of the complete

series of monthly predicted transaction prices over a four-year period is available upon request.

We compare the performance of BME and kriging methods by computing validation statistics for 2012, which is the most recent year in the study period. Typically, predicted values depend on neighboring data. In the time domain, temporal neighbors can contribute from both past and future instances. To make our validation test realistic in the context of the housing market, where prices depend only on past instances, prediction relies only on past temporal neighbors. We assess performance by means of three measures: (i) mean absolute error (MAE) defined as $\frac{1}{N}\sum_{i=1}^{N}|\chi_i - \widehat{\chi_i}|$, (ii) mean relative error (MRE) defined as $\frac{1}{N}\sum_{i=1}^{N}\frac{\chi_i - \widehat{\chi_i}}{\chi_i}$, and (iii) absolute relative error (ARE) defined as $\left|\frac{\chi_i - \widehat{\chi_i}}{\chi_i}\right|$. In the preceding expressions, N predicted values $\widehat{\chi_i}$ are compared to the corresponding observed values χ_i. In addition, we provide visual assessment of the methods via plots of the relative error count at the validation locations throughout 2012.

5. Analysis and Results

5.1. Case A: Larger Data Sample with Increased Price Variability

The Case A sample uses all of the transactions within the study time period without regard for property size, quality, or amenities. The study area covers a rectangular domain where Easting ranges within [715,150 m, 719,750 m] and Northing ranges within [2,127,400 m, 2,136,400 m] in the Texas north-central State Plane coordinate system. In this focus area there are 2,420 transactions from 2009 to 2012. The hard data for the kriging analysis range from 25 to 83 observations per month. According to the empirical rules for geostatistical prediction by Chilès and Delfiner (1999), this sample size is adequate for

prediction when considering the existence of temporal neighbors, too. Table 2 presents descriptive statistics for Case A. In addition to the larger price variability, the present sample prices are unevenly distributed and exhibit strong positive skewness. For example, only a fraction of the houses have prices close to $1 million, as illustrated in the left histogram of Figure 3. For this reason, the analysis in Case A is performed in the log transaction price scale to approximate normality (Banerjee et al., 2004). For both BME and kriging, we predict log transaction prices on a rectangular spatial grid of 47×91 nodes across 48 months, which is a spatiotemporal grid of 205,296 nodes. The nodes are equidistant in the horizontal and vertical directions at 100 m, which is approximately the distance of a city block.

The soft data consists of tax assessed values for 66,908 homes. The regression model for this larger sample specifies an interval of $\pm22.5\%$ uncertainty. As detailed in the previous section, this percentage is applied to each model estimate to construct the Case A soft data. Given the voluminous soft data, we preserved computational resources by considering 40% of the sample. We randomly select 26,226 of the original assessed soft data along with the 2,420 hard data and we call this sample BME40. Table 2 provides the descriptive statistics for the BME40 data. For illustration, Figure 4 shows the locations of the BME40 soft and hard data in January 2010. The figure shows vividly how the soft data outline the housing subdivisions and neighborhoods, in addition to the dwarfing effect the soft data have on the hard data. Given that the BME40 sample still has over 25,000 soft data, we also analyze a subset of just 20%, or 14,772 soft data, and refer to this sample as BME20. The purpose of BME20 is to examine potential effects of selecting significantly fewer soft data in the present BME analysis.

Applying BME and kriging to the respective samples yields the covariance models in Table 3 and the plots in Figure 5. The BME40 covariance plot displays a longer lasting temporal correlation compared to kriging. This effect can be clearly attributed to the additional data used by BME, which help reveal more detailed information about the underlying correlation in space-time than the comparatively limited knowledge resources of kriging. The BME covariance model is expected to have some noise due to the uncertainty in the included soft data, in addition to the uncertainty in the empirical estimation process. Our study follows the existing literature paradigm and does not examine empirical covariance uncertainty. The BME20 covariance model, not shown here, has a comparable plot to BME40 and similar considerations apply.

In general, the prediction results from the two methodologies produce relatively similar maps of housing values and similar standard errors across the 48 months of the study period from 2009 to 2012. A closer look shows that the BME maps provide more spatial structure detail than kriging, particularly in areas with lower hard data density. This meets with our prior based upon more input information from the soft data. To illustrate, Figure 6 provides April and May 2009 results. We note the multiple levels of color and greater detail in Figure 6A using BME, compared to the corresponding kriging prediction plots in Figure 6B. The BME method also appears to pick up some "hot spots" in the center zone of the spatial domain in Figure 6A. These suggest local non-stationarities in the price function over the spatial domain.

Table 4 provides numerical validation of the three samples. Panel A details the MAE and MRE errors measures. The results are mixed across the months in 2012. In January, the BME error is clearly influenced by the uncertainty in the soft data base,

15

whereas kriging remains unaffected. One might expect the additional benefit of using soft data to markedly reduce the error metrics, yet the large variability of transaction prices makes it challenging for both techniques to provide consistently good prediction accuracy throughout the validation period. Although BME and kriging might each perform better in different months, their errors exhibit rather small fluctuations and both techniques appear to be overall equally good performers.

Panel B of Table 4 reports the percentage of validation points within ARE categories. Aggregating the two categories that total 10% error or less, we find that BME40 performs best with 58.38% of all validation points versus 54.49% for BME20 and 50.03% for kriging. This finding suggests an approximate 8% benefit in predicting accuracy with each incremental usage of additional soft data on top of the hard data. At the same time, this benefit comes at a nontrivial computational cost. Processing times were found to be almost proportional to the size of the data sample. Specifically, the BME20 data set is roughly 10 times larger than the kriging one, and BME20 analysis took about 10 times longer to complete prediction than its kriging counterpart. Subsequently, by doubling the amount of soft data, BME40 took about twice as much time to perform prediction than BME20. Figure 7 plots the count of validation points at each relative error level. Thinner and taller curves centered at zero designate better performers, because they indicate that more validation relative errors are around zero. The emerging shapes for the different techniques are largely similar. There is a rather increased error bias of kriging towards the positive errors around zero, and this explains the aforementioned accuracy edge of BME. Figure 7 also illustrates the incremental improvement in the performance of BME20 over kriging, and BME40 over BME20. The plot suggests these are measurable, if rather

16

unimpressive, performance improvements in the present Case A. Given the kriging lower computational burden detailed earlier, the kriging prediction accuracy can be considered comparatively adequate and competitive to BME. We next consider a more practical example.

5.2. Case B: Data Sample with Moderate Price Variability

The Case B sample has 718 hard data observations across the period 2009-2012. The study area, time span, and prediction grid in Case B are identical to those of Case A. However, the reduced data variability compared to the preceding Case A led the regression model to reduced mean estimation error that corresponds to an uncertainty interval of $\pm 15\%$. This percentage was applied to each model estimate to construct the Case B soft data, as detailed in Section 3 earlier.

Narrowing the house type enables us to perform a focused analysis and also produces results that are of more practical use. The sample in Case B offers fewer observed transaction prices than Case A. Consequently, kriging uses between 6 and 29 hard data observations each month. The sample also has 24,294 soft intervals. Table 5 provides the descriptive statistics and Table 6 details the Case B covariance models. The BME model is similar to the Case A covariance model, which meets with our priors since both cases operate in the same space-time domain and there is again an abundance of soft data for the covariance analysis. Further justification is that one might reasonably expect the underlying data generating mechanisms that correlate the observed values to be quite similar in these two cases. The kriging covariance model is a relatively simple single nesting in space and in time.

Figure 8 details the Case B predicted transaction prices using BME and kriging at selected time instances. Due to the reduced variability in transaction prices, log transformation is not necessary in Case B. BME prediction leads to maps of similarly detailed texture as in Case A. In contrast, the kriging results are coarser and feature a smoother prediction surface where the data locations are rather more distinguishable throughout the domain than in the BME maps. Although kriging can carry out predictions across the domain, the results demonstrate that hard data alone produce lower quality inference compared to BME. Prediction errors reflects this finding, with BME prediction PDFs standard deviation ranging up to $14,427, and the corresponding kriging prediction error reaching up to $18,963. One might reasonably expect the soft data uncertainty to contribute to higher BME prediction error. However, the performance suggests that the relatively low soft data uncertainty level of 15% in Case B works synergistically with the abundance of soft data to result in lower prediction error levels for BME.

Table 7 and Figure 9 detail further performance information and error metrics. In Panel A of Table 7, the MAEs and MREs are generally balanced, if slightly overall lower for BME. Panel B of Table 7 provides the error levels as a percentage of validation points. BME achieves AREs that are equal to or less than 10% error for 83.72% of the total validation points. The corresponding kriging score is 66.69%, or about at 4/5 of the BME performance. Figure 9 illustrates the validation error distribution. In Case B, BME's increased accuracy is more apparent through the contrast between the thinner and taller BME curve against a rather flatter and broader kriging curve. In addition, the plot suggests that kriging exhibits a slight negative prediction bias, i.e., its prediction tends to exceed the observed value.

In a realistic scenario where market participants can focus on a price index for their particular house type, the Case B results demonstrate the primary conclusion that available public data greatly enhance the analysis and are necessary to make valid inferences in the absence of hard data.

5.3. Case C: Neighborhood Scale Data Sample

In Case C we investigate a smaller spatial domain within a square spatial field of size 2,500 m × 2,500 m. The spatial grid consists of 51 × 51 nodes and spans across 48 months, thus giving a total of 124,848 nodes in space-time. Compared to Cases A and B, the grid is twice as dense with separation distances of 50 m between nodes in both of the Easting and Northing directions.

By targeting our analysis at the same sample used in Case B, we find that only 111 properties transact within the smaller focus area over the 4-year study period. These data are insufficient to conduct a kriging analysis. In contrast, tax assessment data exist for 4,193 homes in the Case C area, which is abundant information for BME prediction. Table 8 provides the descriptive statistics of the hard and soft data.

Using again year 2012, Table 9 details the error metrics for Case C. Since kriging is unavailable for comparison, we contrast Case C with the larger-area Case B findings. The denser grid in Case C means that prediction is downscaled compared to Case B. The high density of input data enables this downscaling by preserving the same BME prediction characteristics in both Cases B and C. Therefore in principle, a similarity in the results of both cases is rather expected. We observe a reduction in MAEs for four months and a comparatively larger increase in the February, June and July MAEs. There are no statistics for the months April and November when no observed data are available. The percentage

of validation points at or below 10% error in Panel B of Table 9 is 84.85%, which is practically equivalent to the corresponding percentage 83.72% of Case B. Figure 10 portrays maps of BME mean predicted transaction values for Case C at selected time instances.

Overall, the map and statistics being similar to Case B are encouraging results because the Case C resolution is quite high for the real estate research context. This means that BME enables conducting valid space-time analysis at the neighborhood level by using a data set that nearly entirely comprises of soft data, as opposed to kriging that has insufficient input in this scale to perform any analysis.

6. Discussion

Space and time form jointly the space-time continuum rather than existing as separate dimensions. In the scientific analysis of attributes like real estate prices that exist in both vectors, it is imperative to correctly account for the spatiotemporal geometry. Otherwise, separate consideration of the spatial and temporal components can lead to insufficient representation and specious interpretation of space-time phenomena. Whereas the majority of the existing real estate literature has made no advances in dynamic modeling of both the space and time vectors, we use geostatistical methodologies to implement the above perspective and introduce pure spatiotemporal analysis to that field.

Our analysis examines a series of cases. Namely, a general case with increased data variability across the spectrum of transaction prices and two other cases of higher practical interest that include centering on a focused market category and analyzing prices at a more detailed neighborhood level. Our findings demonstrate housing transaction prices prediction in the space-time continuum across the spatial domains of interest and at

a series of temporal instances. In addition, our results highlight BME as a valuable tool for determining housing price indices on the basis of its capability to integrate rigorously soft data that contain nontrivial levels of uncertainty. Compared to classical geostatistical prediction, BME prediction was found to perform slightly more accurately in environments of increased price variability. In such cases, classical geostatistical prediction might be a preferable alternative because processing reduced information carries a lower computational burden compared to BME analysis. However, BME displays noticeably more accurate performance in price conditions with less variability. Finally, we illustrate that BME can also exhibit the same accurate predictive behavior at neighborhood-level analysis in the presence of almost entirely soft information.

We believe that our work makes a useful contribution to the communities of spatial and spatiotemporal statistics and the real estate researchers alike. Our work extends for the first time recent and established space-time prediction methodologies into the real estate domain of application, and showcases additional case-specific benefits from the advanced features of BME analysis. Namely, the present work demonstrates the strength of spatiotemporal prediction as a tool for assessing house transaction prices in specific practical instances of interest to real estate experts, as summarized above. We also illustrate how geostatistics can help understand the spatiotemporal transaction price variability and its characteristics by translating the analysis outcome into informative maps. Transaction price maps reveal the local market behavior as it evolves across space and time, and can be valuable tools for assessment, analysis, planning and management for both real estate researchers and laypeople in housing markets.

Appendix A

Much of the spatiotemporial literature in real estate uses the general spatial model. The specification is

$$y = \rho W_1 y + XB + u$$

$$u = \theta W_2 u + \epsilon \qquad (A1)$$

$$\epsilon \sim N(0, \sigma^2 I_n),$$

where y is an $n \, x \, 1$ vector of dependent variables and X is an $n \, x \, k$ matrix of explanatory variables. W_1 and W_2 are $n \, x \, n$ nonnegative spatial weight matrices.

The weight matrices describe the spatial arrangement of the sample observations. The diagonal elements are set to zero since no spatial unit is a neighbor to itself. Procedures to determine the elements of the weight matrix vary, one simple method is to define $w_{i,j} = max\left\{1 - \frac{d_{i,j}}{d_{max}}, 0\right\}$ where $d_{i,j}$ is the separation distance converted from the latitude and longitude for each observation. The weight matrix is then row normalized bounding the eigenvalues between -1 and 1.[5] When $W_1 = 0$ and W_2 captures the spatial relation, the specification is the familiar SEM in which the vector of residuals u addresses spatial dependence or autocorrelation. When $W_2 = 0$ and the spatial relationships form W_1 the specification is the typical SAR, which assumes the spatial autocorrelation is a spatial lag of the dependent variable.[6] ρ is a scalar measuring the average influence of neighboring

---

[5] Lee (2004), LeSage and Pace (2009), and Elhorst (2012) discuss more advanced topics such as stationarity, the need for characteristics roots to lie in the unit circle, and the fact that complex eigenvalues can exist. Tu, Yu, and Sun (2004) illustrate a different method to determine the matrix elements using a modified distance-decay weighting scheme as in McMillian (1996).

[6] Alternative specifications of the general spatial model include the spatial Durbin model and spatial lags on the explanatory variables. The spatial Durbin model assumes a spatial structure in the dependent variable as

observations of the dependent variable on the regressand. Note that ρ is not dynamic and imposes a single model structure *a priori* on the data. The same is true for θ.[7] That one parameter can account for the entire spatial dependence structure is a considerable expectation.

McMillian (2012) notes that it is not surprising that ρ is often significant. Instead of $W_1 y$ having a casual effect on y due to nearby dependent-variable levels, $W_1 y$ is a set of predicted values of the regressand and thus just a form of linear smoothing. Accordingly, the spatial econometrician should expect ρ to be significant. Of course the major issue is that this model misspecification produces significant estimates of ρ and θ when the true values are zero.

In addition to ρ and θ imposing a single model structure and potentially being spurious, most prior real estate applications make the assumption that the regression coefficients X in Equation (A1) are constant across the region of study. However, this is not correct if the sample exhibits spatial or temporal heterogeneity. Spatial heterogeneity occurs when there is structural instability or nonstationarity of economic relationships over the study region.[8] The problem is further compounded by the quality characteristics of

---

well as the explanatory variables but no spatial structure in the residuals, $y = \rho W_1 y + XB + W_1 X \gamma + u$ and $W_2 = 0$. An alternative is to drop the assumption of a spatial lag on the dependent variable and instead assume that the regressand is a function of spatial lags of the explanatory variable only i.e., $y = XB + W_1 X \gamma + u$ and $W_2 = 0$. Since these evolve from the general spatial model, the issues we discussion pertain equally to these specifications.

[7] Unless W_1 or W_2 are markedly different, which is suspicious in its own right, estimation of both ρ and θ will cause an identification problem and thus is not done.

[8] This is not spatial heterogeneity due to heteroscedasticity, when spatial units exhibit different error variance, which can be addressed Bayesian heteroscedasticity robustness procedure as in Gweke (1993) and Tu, Yu, and Sun (2004).

properties change over time. For instance, Munneke and Slade (2001) find intertemporal variation in the regression coefficients and that the conventional time dummy approach does not represent the pure-price effect of time.

Appendix B

Our study uses a regression model to estimate tax assessed transaction values that regresses transaction prices against the tax assessed values as well as property characteristics that are public information in the tax assessor's database. This regression model is an ordinary least squares specification that controls for fixed effects across five realtor-defined areas. The dependent variable is the observed transaction prices. The independent variables come from the public records database and include measures of home size, age, and quality. The following are the variable definitions:

Sales prices: The dependent variable and the observed prices at which the property recently transacted.

Appraised values: The prices the local tax assessor's office values the properties.

# of bedrooms: The number of bedrooms in each home; a measure of size and quality.

# of bathrooms: The number of bathrooms in each home; a measure of size and quality.

Home age: The age of the home.

Living area: The amount of living area in the home; a measure of size and quality.

Garage capacity: The number of garage spaces, which can indicate size but typically denotes more of a quality measure.

Areas 126–130: Five subareas within our sample space, which are defined and named by the local Board of Realtors.

References

Bannerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC, Boca Raton, FL. 2015, New Edition.

Can. A. and I. Megbolugbe. 1997. Spatial Dependence and House Price Construction. *Journal of Real Estate Finance and Economics* 14: 203–222.

Case, B., Clapp, J., Dubin, R., Rodriguez ,M., 2004. Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models. The Journal of Real Estate Finance and Economics. 29(2), 167–191.

Chilès, J.P., Delfiner, P., 1999. Geostatistics-Modeling Spatial Uncertainty. John Wiley & Sons. New York, NY.

Christakos, G., 1991. On Certain Classes of Spatiotemporal Random Fields With Application to Space-Time Data Processing. IEEE-Trans on Systems, Man, and Cybernetics. 21(4), 861–875.

Christakos, G., 1992. Random Field Models In Earth Sciences. Academic Press. San Diego, CA; 2005. New edition, Dover Publ., Inc. Mineola, NY.

Christakos, G., 2000. Modern Spatiotemporal Geostatistics. Oxford Univ Press. New York, NY; 2012. New edition. Dover Publ., Inc. Mineola, NY.

Christakos, G., 2010. Integrative Problem-Solving in a Time of Decadence. Springer-Verlag. New York, NY.

Christakos, G., 2014. Stochastic Medical Reasoning and Environmental Health Exposure. Imperial College Press.

Christakos, G., Hristopulos, D.T., 1998. Spatiotemporal Environmental Health Modelling: A Tractatus Stochasticus. Kluwer Academic Publishers. Boston, MA.

Christakos, G., Bogaert, P., Serre, M.L., 2002. Temporal GIS. With CD-ROM. Springer-Verlag, New York, NY.

Christakos, G., Hristopulos, D.T., Bogaert, P., 2000. On the Physical Geometry Concept at the Basis of Space/Time Geostatistical Hydrology. Advances in Water Resources. 23(8), 799–810.

Christakos, G., Kolovos, A., Serre, M.L., Vukovich, F., 2004. Total Ozone Mapping by Integrating Data Bases From Remote Sensing Instruments and Empirical Models. IEEE Transactions on Geoscience and Remote Sensing. 42(5), 991–1008.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*. John Wiley and Sons: New York, NY.

Dubé, J. and D. Legros. 2013a. Dealing with Spatial Data Pooled Over Time in Statistical Models. *Letters in Spatial and Resource Sciences* 6: 1–18.

Dubé, J. and D. Legros. 2013b.A Spatio-Temporal Measure of Spatial Dependence: An Example Using Real Estate Data. *Papers in Regional Science* 92(1): 19–30.

Elhorst, J.P. 2012. Dynamic Spatial Panels: Models, Methods, and Inferences. *Journal of Geographical Systems* 14: 5–28

Gelfand, A.E., Ghosh, S.K., Knight, J.R., Sirmans, C.F., 1998. Spatio-Temporal Modeling of Residential Sales Data. *Journal Business & Economics Statistics* 16(3), 312–321.

Gelfand, A.E., Kim, J.-J., Sirmans, C.F., Banerjee, S., 2003. Spatial Modeling with Spatially Varying Coefficient Processes. *Journal of the American Statistical Association* 98(462), 387–396.

Gibbons, S., Overman, H.G., 2012. Mostly Pointless Spatial Econometrics? *Journal of Regional Science* 52(2), 172–191.

Goodman, A., 1977. Comparison of Block Group and Census Tract Data in a Hedonic Housing Price Model. *Land Economics* 53, 483–487.

Gweke, J., 1993. Bayesian Treatment of the Independent Student-*t* Linear Model. *Journal of Applied Econometrics* 8, S19–S40.

Huang, B., B. Wu and M. Barry. 2010. Geographically and Temporally Weighted Regression for Modeling Spatio-temporal Variation in House Prices. *International Journal of Geographical Information Science* 24(3): 383–401.

Kelejian, H.H. and I.R. Prucha. 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40: 509–533.

Kolovos, A., Christakos, G., Serre, M.L., Miller, C.T., 2002. Computational Bayesian Maximum Entropy Solution of a Stochastic Advection-Reaction Equation in the Light of Site-Specific Information. *Water Resources Research* 38(12), 1318–1334.

Kolovos, A., Skupin, A., Jerrett, M., Christakos, G., 2010. Multi-Perspective Analysis and Spatiotemporal Mapping of Air Pollution Monitoring Data. *Environmental Science and Technology* 44(17), 6738–6744.

Kolovos, A., Christakos, G., Hristopulos, D.T., Serre, M.L., 2004. On Certain Classes Of Non-Separable Spatiotemporal Covariance Models. *Advances in Water Resources* 27, 815–830.

Kolovos, A., J.M. Angulo, K. Modis, G. Papantonopoulos, J-F. Wang, and G. Christakos. 2013. Model-Driven Development of Covariances for Spatiotemporal Environmental Health Assessment. *Environmental Monitoring and Assessment* 185(1): 815–831.

Lee, L-F. 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Econometric Models. *Econometrica* 72: 1899–1926.

LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall/CRC. Boca Raton, FL.

McMillen, D.P. 1996. One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach. *Journal of Urban Economics* 40(1): 100–124.

McMillen, D.P., 2012. Perspectives on Spatial Econometrics: Linear Smoothing with Structured Models. *Journal of Regional Science* 52(2), 192–209.

Munneke, H.J. and B.A. Slade. 2001. A Metropolitan Transaction-Based Commercial Price Index: A Time-Varying Parameter Approach. *Real Estate Economics* 29(1): 55–84.

Nappi-Choulet, I. and T-P. Maury. 2009. A Spatiotemporal Autoregressive Price Index for the Paris Office Property Market. *Real Estate Economics* 37(2): 305–340.

Nappi-Choulet, I. and T-P. Maury. 2011. A Spatial and Temporal Autoregressive Local Estimation for the Paris Housing Market. *Journal of Regional Science* 51(4): 732–750.

Openshaw S. and P. Taylor. 1979. A Million or So Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In Wrigley N (ed.) *Statistical Applications in the Spatial Sciences* Pion: London, 127–144

Pace, R.K., R. Barry and C.F. Sirmans. 1998. Spatial Statistics and Real Estate. *Journal of Real Estate Finance and Economics* 17(1): 5–13.

Pace, R.K. and O. Gilley. 1997. Using the Spatial Configuration of the Data to Improve Estimation. *The Journal of Real Estate Finance and Economics* 14(3): 333–340.

Pace, R.K., R. Barry, O. Gilley and C.F. Sirmans. 2000. A Method for Spatial-Temporal Forecasting with an Application to Real Estate Prices. *International Journal of Forecasting* 16: 229–246.

Pinkse, J., Slade, M.E., 2010. The Future of Spatial Econometrics. *Journal of Regional Science* 50(1), 103–117.

Powell, M. J. D., 2009. The BOBYQA algorithm for bound constrained optimization without derivatives (Report). Department of Applied Mathematics and Theoretical Physics, Cambridge University. DAMTP 2009/NA06.

Savelyeva, E., Utkin, S., Kazakov, S, Demyanov, V., 2010. Modeling Spatial Uncertainty for Locally Uncertain Data. In: geoENV VII–Geostatistics for Environmental Applications: Quantitative Geology and Geostatistics. Eds.: Atkinson, P. M., and Lloyd, C. D. Springer, The Netherlands, pp. 295—306.

Smith, T.E. and P. Wu. 2009. A Spatio-temporal Model of Housing Prices Based on Individual Sales Transactions Over Time. *Journal of Geographical Systems* 11: 333–355.

Sun, H., Y. Tu and S-M. Yu. 2005. A Spatio-temporal Autoregressive Model for Multi-unit Residential Market Analysis. *Journal of Real Estate Finance and Economics* 31(2): 155–187.

Tu, Y., S-M. Yu and H. Sun. 2004. Transaction-Based Office Price Indexes: A Spatiotemporal Modeling Approach. *Real Estate Economics* 32(2): 297–328.

Table 1: Transaction Price model ($) using an ordinary least squares fixed-effect model

Variable	Parameter Estimate	p-value
Appraised value	1.04	0.00
Living area	-3.34	0.12
Number of bedrooms	-628.63	0.49
Number of bathrooms	858.65	0.48
Garage capacity	6,088.86	0.00
Home age	-112.07	0.02
Realtor-defined area 127	913.00	0.62
Realtor-defined area 128	-2,661.31	0.14
Realtor-defined area 129	-2,248.32	0.53
Realtor-defined area 130	-125.54	0.95
Intercept	-11,072	0.02
Number of observations	1,901	
F statistic	6,264.56	
Adjusted R^2	0.97	

Table 2: Case A descriptive statistics

	Full Sample for Kriging	BME40	BME20
Sample size	2,420	28,646	17,192
Minimum ($)	50,000	50,000	50,000
Maximum ($)	400,000	998,880	989,531
Mean ($)	117,832	118,669	118,742
Standard Deviation ($)	53,338	62,232	62,863
Skewness	1.79	5.69	5.58
Kurtosis	7.32	55.77	53.97
Median ($)	105,000	104,800	104,600
V_{tr} (%)	57,00	55.84	52,05

Table 3: Case A covariance models (r and t indicate spatial and temporal distances, respectively)

Analysis Technique	Nested Model	Spatial Structure	Temporal Structure	Sill c ($\$^2$)	Spatial Range ρ (m)	Temporal Range τ (months)
BME40	1	Nugget	Spherical	0.022		81.42
	2	Exponential	Spherical	0.034	75.01	255.00
	$C(r,\tau) = c_1 \left[1 - \dfrac{3|t|}{2\tau_1} + \dfrac{|t|^3}{2\tau_1{}^3} \right] + c_2 \exp\left(-\dfrac{r}{\rho_2} \right) \left[1 - \dfrac{3|t|}{2\tau_2} + \dfrac{|t|^3}{2\tau_2{}^3} \right]$						
BME20	1	Nugget	Spherical	0.019		48.71
	2	Exponential	Spherical	0.044	134.93	170.00
	$C(r,\tau) = c_1 \left[1 - \dfrac{3|t|}{2\tau_1} + \dfrac{|t|^3}{2\tau_1{}^3} \right] + c_2 \exp\left(-\dfrac{r}{\rho_2} \right) \left[1 - \dfrac{3|t|}{2\tau_2} + \dfrac{|t|^3}{2\tau_2{}^3} \right]$						
Kriging	1	Gaussian	Gaussian	0.026	74.89	11.85
	2	Exponential	Spherical	0.042	359.08	127.50
	$C(r,\tau) = c_1 \exp\left(-\dfrac{r^2}{\rho_1^2} \right) \exp\left(-\dfrac{t^2}{\tau_1^2} \right) + c_2 \exp\left(-\dfrac{r}{\rho_2} \right) \left[1 - \dfrac{3|t|}{2\tau_2} + \dfrac{|t|^3}{2\tau_2{}^3} \right]$						

Table 4: Case A error metrics

Panel A: Mean Absolute Errors and Mean Relative Errors

Month	MAE ($)			MRE (%)		
	BME40	BME20	Kriging	BME40	BME20	Kriging
January	34,116.56	35,775.70	33,776.71	-4.85	-4.69	-0.60
February	27,748.26	29,919.14	25,647.09	-2.80	-3.77	-3.40
March	18,281.60	22,607.56	22,685.04	-4.50	-2.92	-2.85
April	20,717.20	23,547.85	24,571.37	-1.19	-1.31	-3.36
May	20,898.61	21,403.51	21,806.01	-3.58	-1.80	-3.66
June	15,784.03	16,045.08	17,414.79	-6.07	-5.33	-4.29
July	18,603.97	17,111.56	21,358.81	1.75	1.18	2.03
August	18,081.09	19,421.99	20,057.93	-1.45	-1.50	-5.85
September	30,221.98	27,194.93	33,012.39	0.88	2.67	0.37
October	26,715.39	26,928.89	24,646.65	2.91	2.11	0.80
November	20,195.03	22,157.77	21,625.46	0.81	-1.47	-2.14
December	21,905.61	21,348.34	22,233.24	3.08	3.01	-1.60

Panel B: Percentage of Validation Points

ARE Categories	BME40	BME20	Kriging
Less than 5%	20.05	19.08	16.93
Between 5% and 10%	38.33	35.41	33.10
Greater than 100%	1.54	1.77	1.07

Table 5: Case B descriptive statistics

	BME	Kriging
Sample size	25,012	718
Minimum ($)	79,140	100,000
Maximum ($)	249,623	245,000
Mean ($)	135,278	141,424
Standard Deviation ($)	23,000	27,604
Skewness	1.02	1.14
Kurtosis	4.56	4.69
Median ($)	131,200	136,950
V_{tr} (%)	60.65	52.74

Table 6: Case B covariance models (r and t indicate spatial and temporal distances, respectively)

Analysis Technique	Nested Model	Spatial Structure	Temporal Structure	Sill c ($\$^2$)	Spatial Range ρ (m)	Temporal Range τ (months)
	1	Nugget	Spherical	1.0e8		99.30
BME	2	Exponential	Spherical	1.0e8	78.32	155.00
		$C(r,\tau) = c_1 \left[1 - \dfrac{3|t|}{2\tau_1} + \dfrac{|t|^3}{2\tau_1{}^3} \right] + c_2 \exp\left(-\dfrac{r}{\rho_2} \right) \left[1 - \dfrac{3|t|}{2\tau_2} + \dfrac{|t|^3}{2\tau_2{}^3} \right]$				
	1	Exponential	Spherical	3.6e8	180.02	15.17
Kriging		$C(r,\tau) = c_1 \exp\left(-\dfrac{r}{\rho} \right) \left[1 - \dfrac{3|t|}{2\tau} + \dfrac{|t|^3}{2\tau^3} \right]$				

Table 7: Case B error metrics

Panel A: Mean Absolute Errors and Mean Relative Errors

Month	MAE ($)		MRE (%)	
	BME	Kriging	BME	Kriging
January	17,108.86	19,113.75	-1.65	-7.43
February	8,844.30	7,205.73	1.61	-1.56
March	16,194.10	15,802.41	-0.44	-1.72
April	14,835.48	11,890.95	2.28	1.99
May	13,915.95	13,190.73	2.78	-0.39
June	9,040.79	14,586.61	-1.26	-4.87
July	11,924.45	15,621.73	3.72	-0.35
August	13,344.60	14,246.72	0.76	-0.85
September	17,224.21	14,824.04	2.69	-1.38
October	15,360.89	16,312.42	-0.77	-3.32
November	12,388.27	12,476.14	0.73	0.51
December	13,823.62	13,251.16	-0.59	-5.60

Panel B: Percentage of Validation Points

ARE Categories	BME	Kriging
Less than 5%	29.43	22.78
Between 5% and 10%	54.29	44.31
Greater than 100%	0.07	0

Table 8: Case C BME descriptive statistics

Sample size	4,304
Minimum ($)	100,000
Maximum ($)	249,800
Mean ($)	137,809
Standard Deviation ($)	26,676
Skewness	1.58
Kurtosis	5.85
Median ($)	130,350
V_{tr} (%)	66.20

Table 9: Case C error metrics

Panel A: Mean Absolute Errors and Mean Relative Errors		
Month	MAE ($)	MRE (%)
January	15,594.78	-13.92
February	31,215.24	-28.38
March	13,377.30	-10.65
April	N/A	N/A
May	15,769.40	4.00
June	13,808.65	-1.09
July	23,940.64	9.35
August	14,262.87	-2.71
September	15,935.84	-14.49
October	15,277.72	4.63
November	N/A	N/A
December	14,083.59	4.95

Panel B	
ARE Categories	Percentage of Validation Points
Less than 5%	27.27
Between 5% and 10%	57.58
Greater than 100%	0
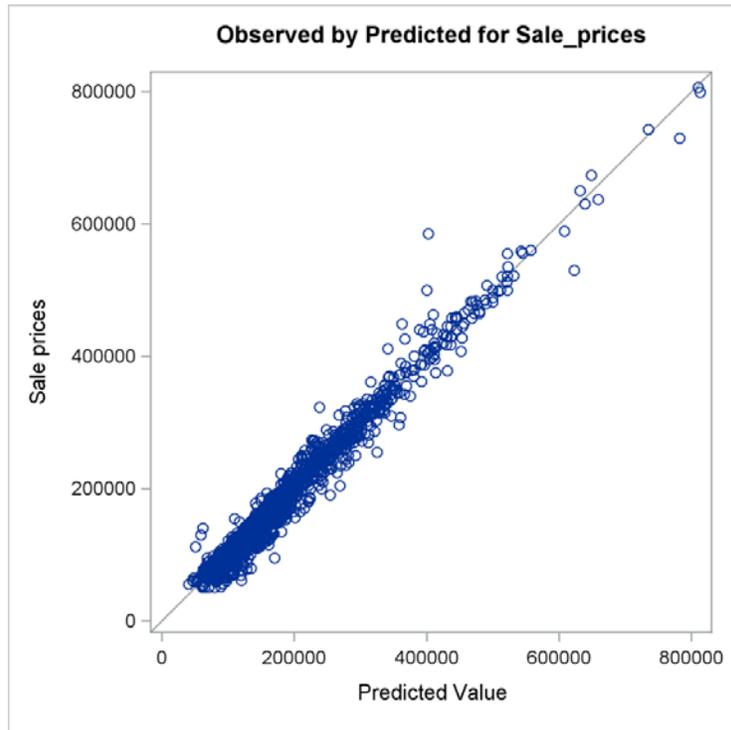
Figure 1: Predicted values of Transaction Prices (shown as variable Sale_prices) versus the observed values using the ordinary least squares model detailed in Table 1.
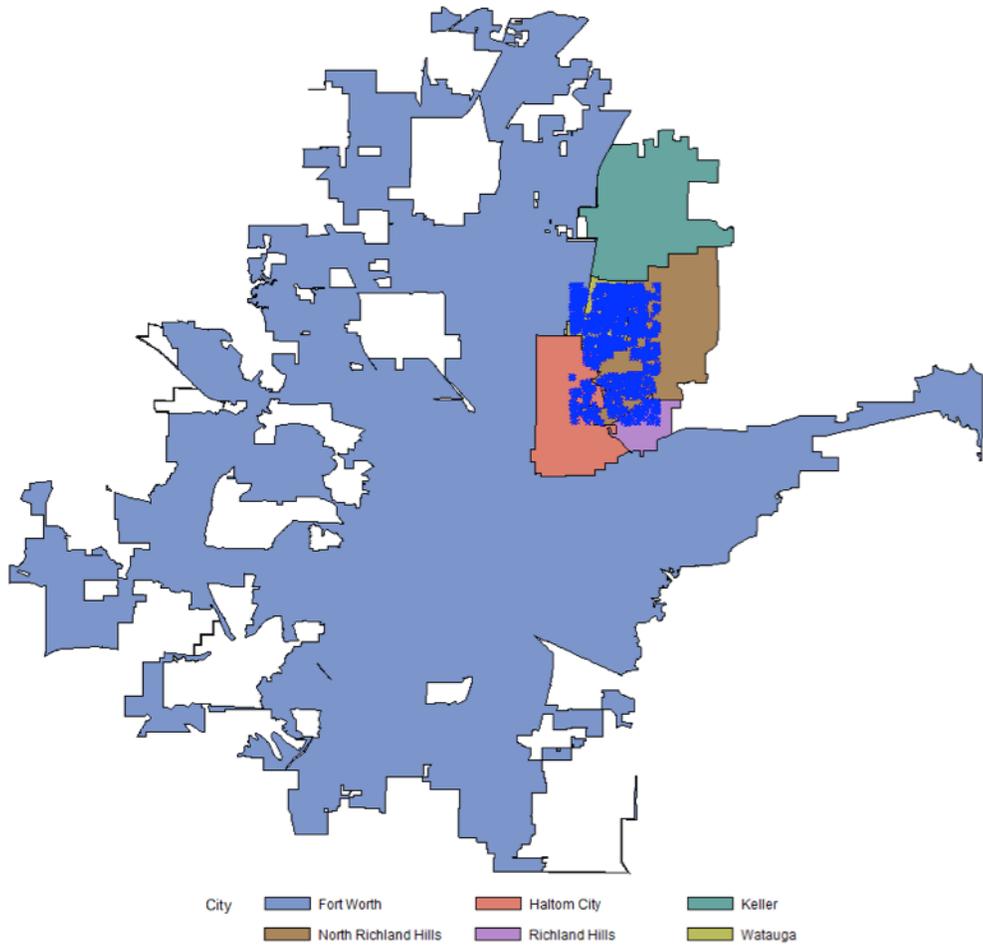
Figure 2: Domain of a sample of residential Transaction Prices in Tarrant County, Texas.
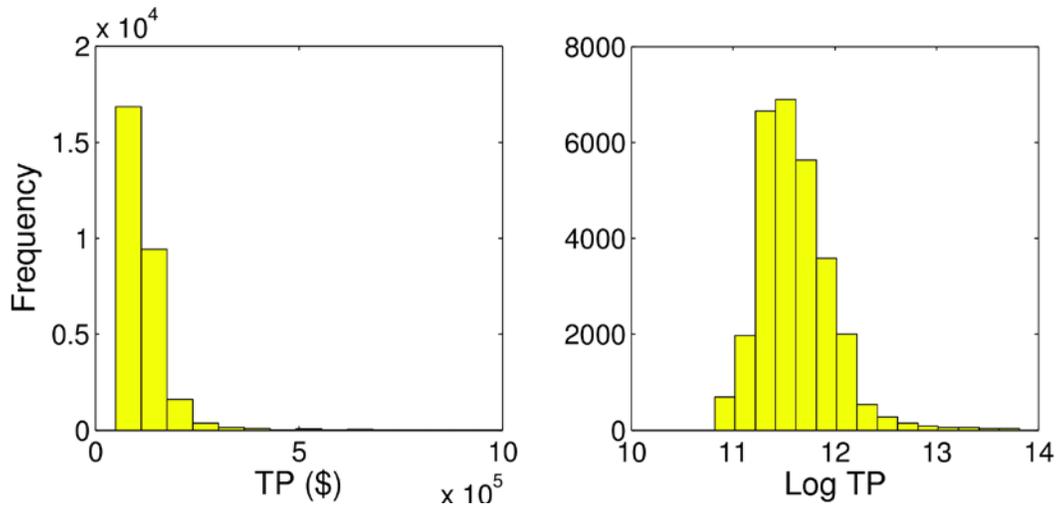
Figure 3: Histogram of observations for Case A in original (left) and log Transaction Price space (right).
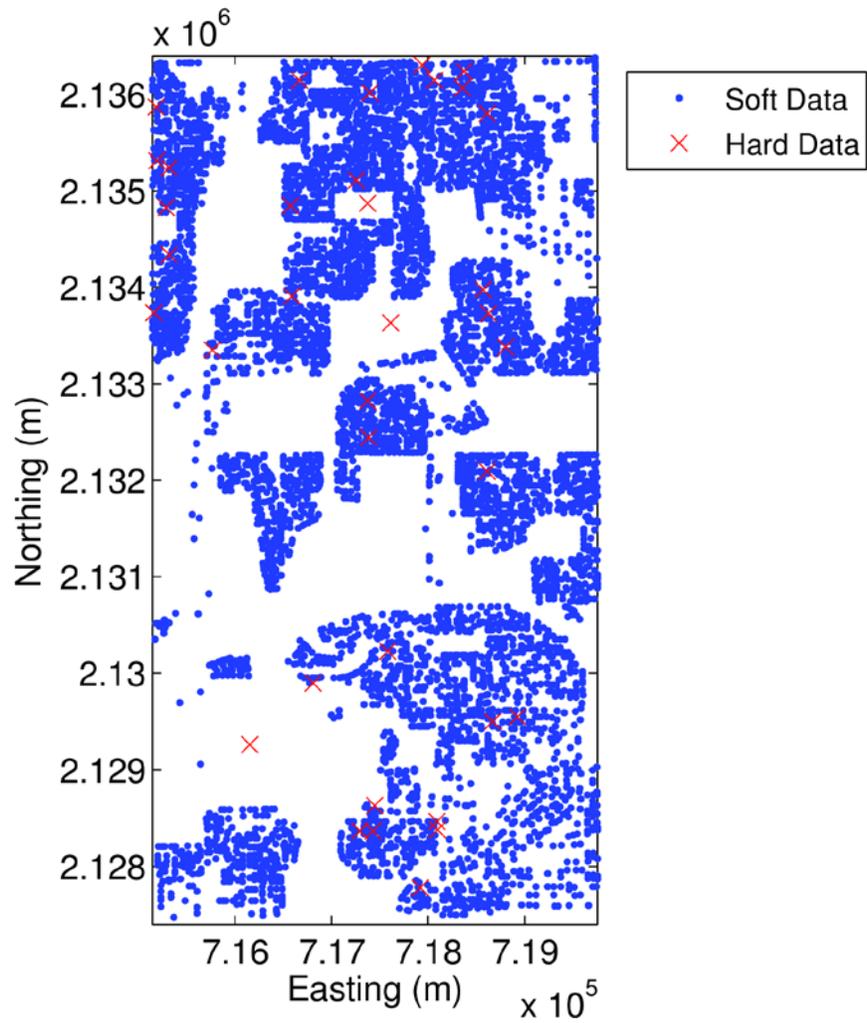
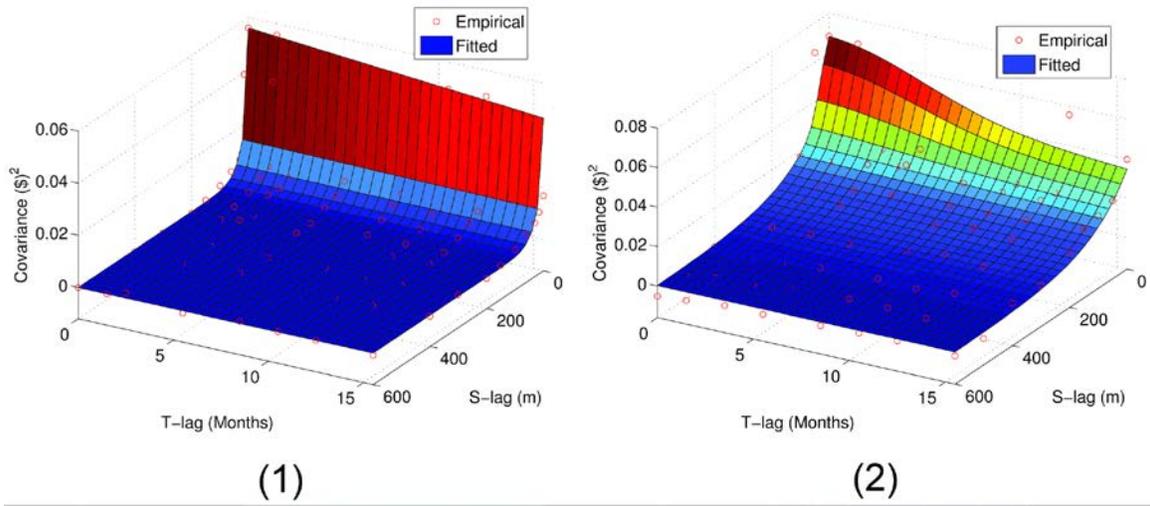Figure 4: Locations of tax assessed (soft data) values and Transaction Prices (hard data) in January 2010.

Figure 5: Case A empirical and fitted covariance plots for (1) BME40, and (2) Kriging. Covariance values are plotted against temporal and spatial lags.
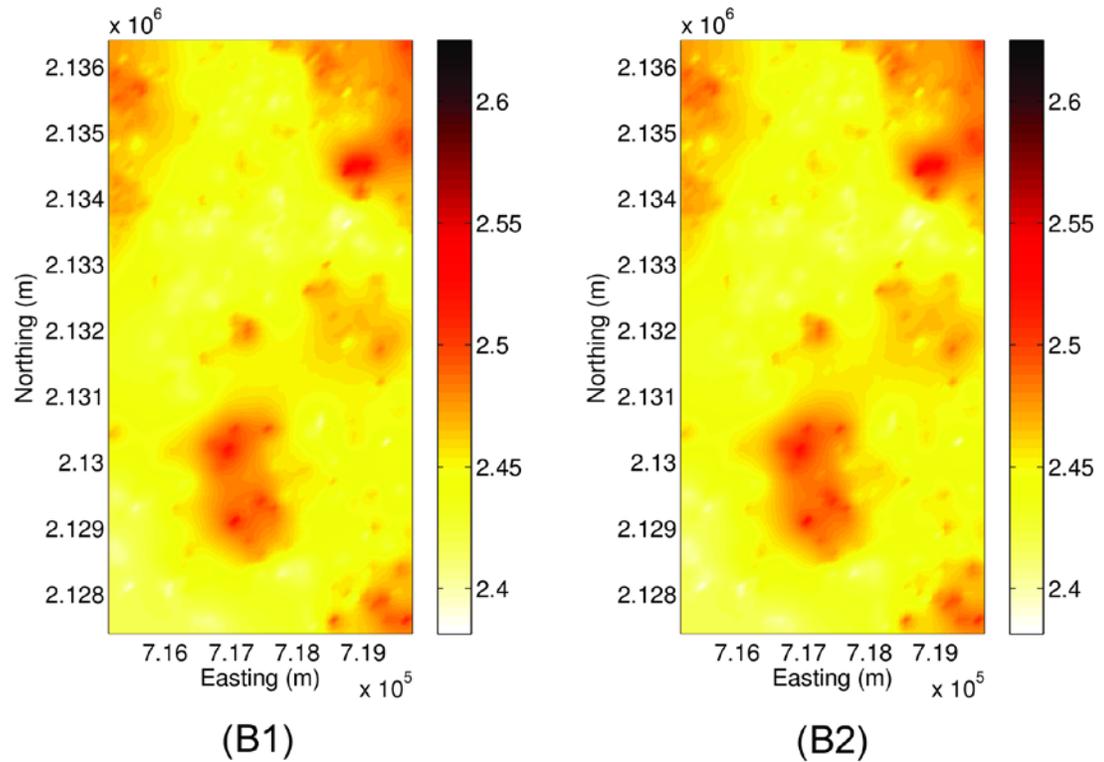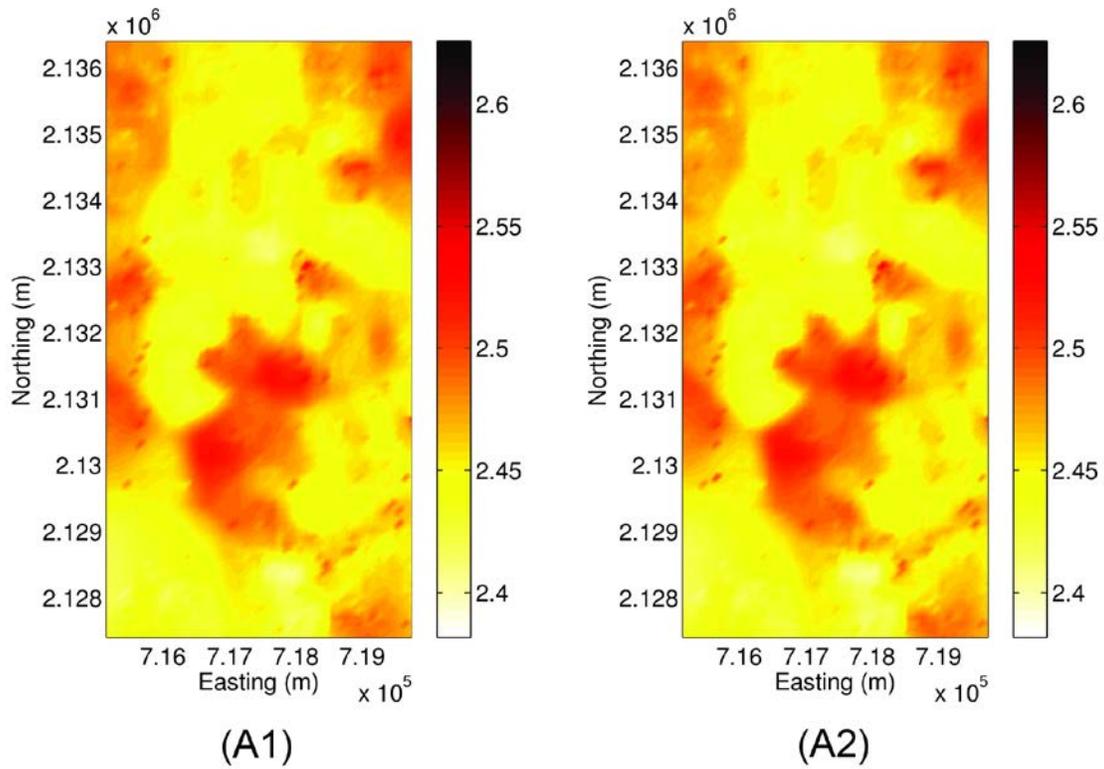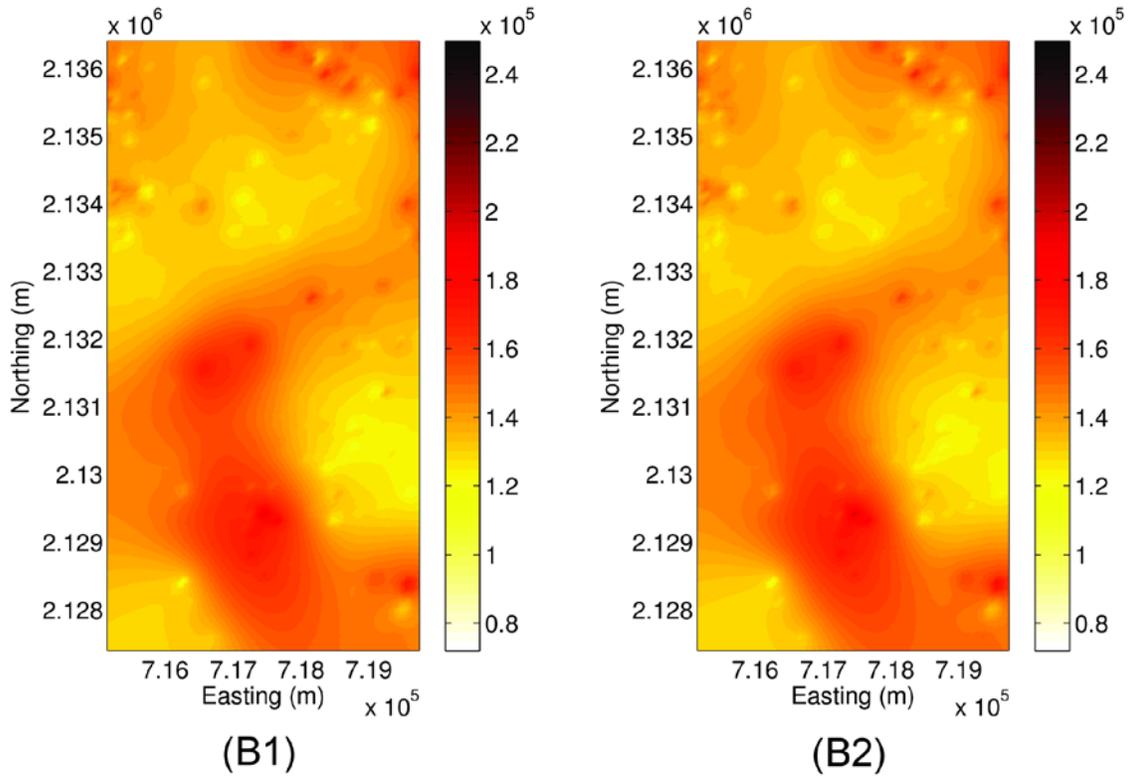
Figure 6: Predicted log Transaction Price at selected time instances for Case A: Figure (6A) shows BME40 prediction for (A1) April 2009, and (A2) May 2009. Figure (6B) shows kriging prediction for (B1) April 2009, and (B2) May 2009.
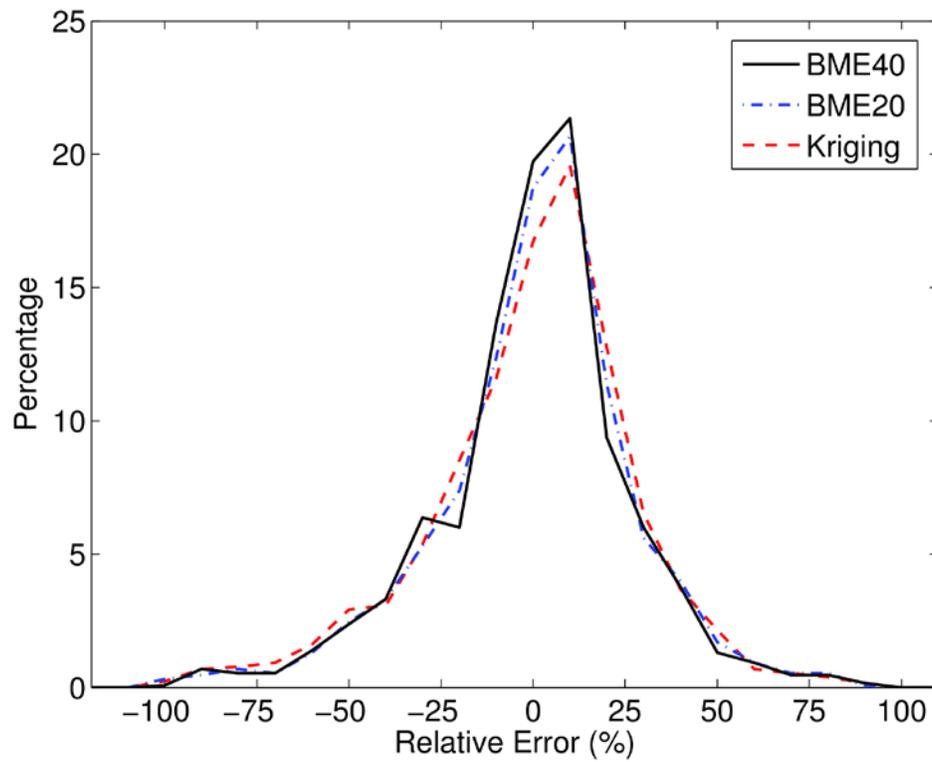
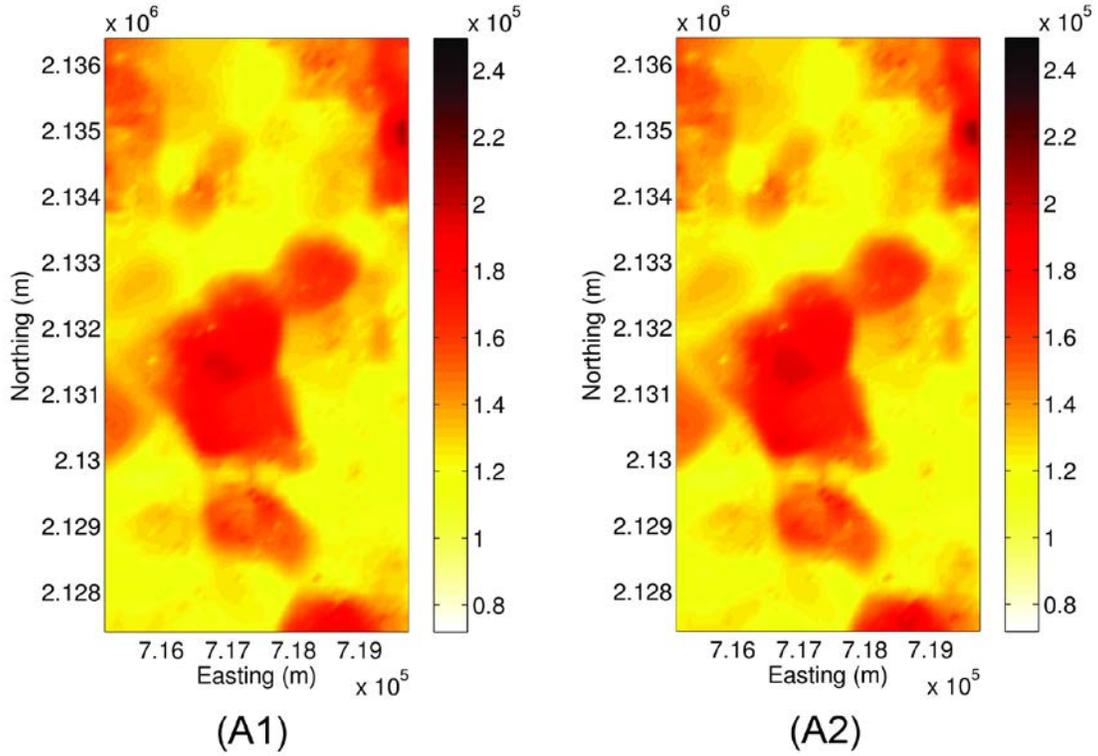Figure 7: Count of validation points in relative error bins using Case A approaches.

Figure 8: Predicted Transaction Price at selected time instances for Case B: Figure (8A) shows BME prediction for (A1) April 2009, and (A2) May 2009. Figure (8B) shows Kriging prediction for (B1) April 2009, and (B2) May 2009.
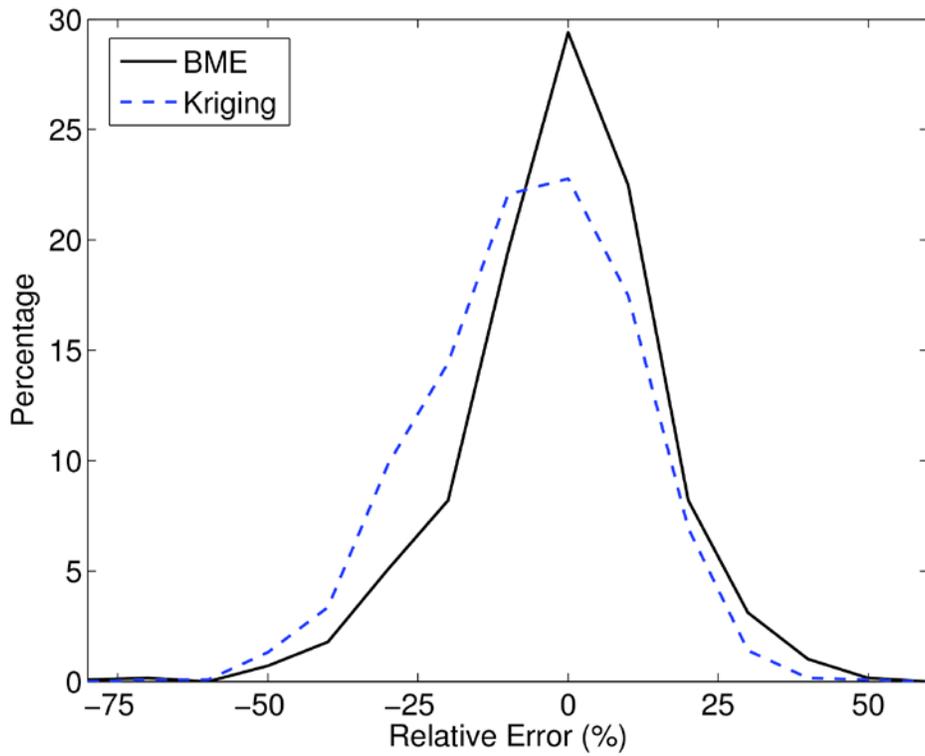
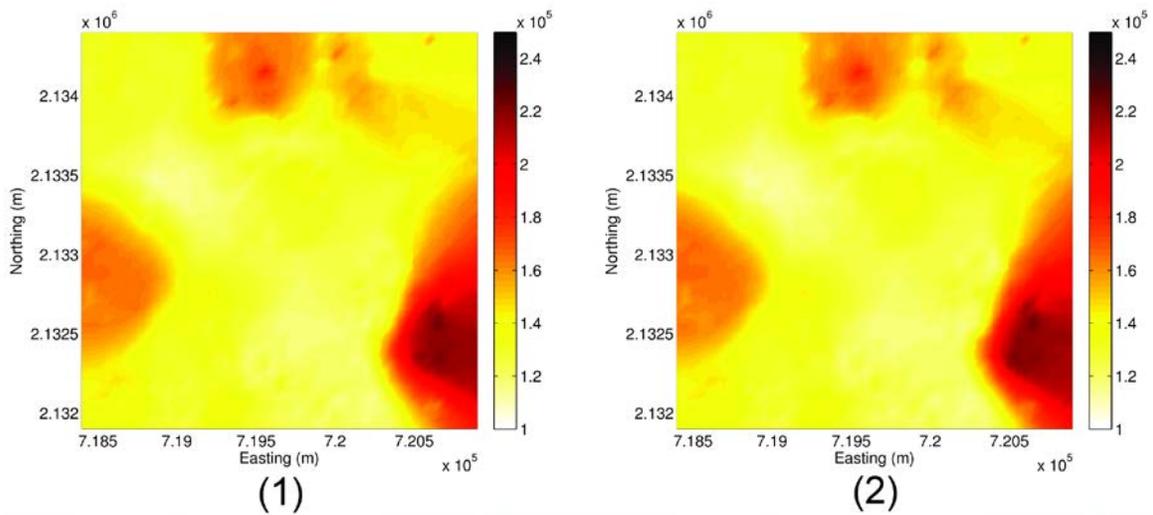Figure 9: Count of validation points in relative error bins using Case B approaches.

Figure 10: Predicted Transaction Price at selected time instances for Case C: Plots show BME prediction for (1) April 2009, and (2) May 2009.