Issues in Developing Rent Hedonic Index

(Draft)

* Sample Selection and Data Preparation (60% +/-10%)
  + Define asking rent:
    - Which asking rent? Listed vs. leased?
      * If leased deals only, do we have a sample selection biases?
    - On which date? Listing date vs. sign date?
    - If using asking rents, most renewals are likely missing from the sample. Implications?
    - What if asking rent was continuously updated?
  + Define hedonic factors:
    - Research more than 1,000 tables in CoStar Enterprise database alone. Candidate factors may easily as many as a few hundreds or more.
    - Location and building characteristics
      * Location and building characteristics may be changing over time
      * Is Andy able to provide the access to more accurate/updated amenity database?
        + If so, how to match? Using a radius of 500 feet?
    - Lease characteristics
      * Actual terms, conditions, and provisions of a lease deal may be very different from those when asking rent was initially set. Which set of lease characteristics should we use in our model?
      * Leases may be of mixed uses, e.g., retail/office. What to do with them?
  + Data & Sample Size
    - Is the sample large enough for rent index (for a particular cluster/submarket). If not, how to handle the problem?
    - Need to further familiarize with and research into about 1,000 CoStar Enterprise tables
  + Data Cleaning
    - Univariate analysis and summary statistics:
      * Market/cluster/sub market
      * Property type/segment
      * Time
    - Data winsorization (for more robust model)
    - Dummy creation (time dummy, missing variable dummy, and characteristic dummy)
  + Need to understand market segments better
    - For example, …
    - How urban pedestrian retailing rental market may be different from that of a suburban super regional mall?
    - Life-style center vs. community shopping center
    - High rise vs. low rise office
    - So far, not familiar with warehouse market and its segmentation
* Model Development & Validation (40%+/-10%)
  + Selection of hedonic factors in our model (for about 3,000 property type and submarket combinations. We plan to cover three property types.
  + How to deal with reporting lag issue?
  + Model specification issues:
    - Selection of appropriate functional form: Semi-log vs. Log-Log vs. Box-Cox?
    - Testing stable coefficients:
      * Over time?
        + Academic studies reject this assumption
      * Cross-section?
        + If not, incorporate interaction terms?
    - Testing spatial effects:
      * Determine appropriate weight matrix
      * Assessment of spatial dependency (for a given selected weight matrix)
        + *Moran’s I and Geary’s C* tests
        + Spatial dependence may vary across space: test of G statistics
      * Determine the type of spatial dependency:
        + SAR, SEM, or Spatial Durbin?
        + Localled weighted?
      * Visually examine spatial distributions of errors