

Welcome,

The following is the fifth in a series of newsletters designed to foster a healthy dialog. I hope you enjoy it and encourage your feedback and discussion. The full piece is available on the web:



<http://www.caswell.org/newsletters/20090915.asp>

Take a look on <http://www.caswell.org> for prior articles, access to resources and information on what I can do for you. If you know someone who would benefit from this piece, feel free to forward it.

Please let me know your thoughts on this piece and other subjects you would like explored.

Regards,

Ward S Caswell
ward@caswell.org
(617) 304-2689

Making Sense of Messy Data

Are you an expert in commercial real estate? Malcolm Gladwell's Blink, the Power of Thinking Without Thinking suggests we are all experts in myriad subjects and simply lack the nomenclature to communicate with others. By way of example he describes our ability to recognize faces instantly and accurately. It's something we can all do very well, but few of us are able to describe a face sufficiently for another person to recognize an individual. Forensic artists are able to communicate within their field using terms defined by fellow practitioners, allowing them to precisely describe a face. In a similar way, most of us can recognize a high quality, well located office tower in



comparison to a lower quality, more poorly placed property. Along those lines, we are able to differentiate the value of two properties from our own knowledge of the basic forces influencing pricing. But to accurately and precisely define the value of a property, you need market data. The difficulty arises when you review the data available for commercial real estate (CRE) and discover inconsistencies between sources. How do you know which sources are accurate? In this article I will review the sources, issues, and solutions for data. As CRE defaults mount, threatening the larger economy with a double dip recession, a better understanding of the data will help you gauge the opportunities and make more informed decisions.

Commercial real estate is not a transparent market place. There are no benchmarks for rental rates, income, demand, or even supply. Different indices exist, but each has flaws in methodology and without the ability for the end user to recreate the data, they are indices, not a benchmarks, and therefore not open to the level of scrutiny necessary to improve reliability. To get started in creating your own set of data, take a close look at what is available. The brokerage houses produce some excellent quarterly and annual reports as well as white papers and opinion pieces. Pulling together reports for a few markets provides a wealth of detailed data in just a few minutes. The trouble begins when you compare the reports. While the general trends are usually similar, the specific values for things like vacancy rates, net absorption, and rents differ widely. Each firm has its own set of properties included in their surveys and use different methods to calculate variables. Most firms have gone away from including extensive descriptions of their methodologies, making it impossible for the reader to clarify the differences.

The methodology page is the least read but most important section.

A well placed call to the firms' research departments will usually provide more detailed data, clarifications of the methodologies, and sometimes even a handy sheet describing the differences between sources within a market.

Many people attempt to create cross market comparisons using data from individual market reports. This can seriously misrepresent the situation when levels are used across differing methodologies. For example, comparing vacancy rates from two markets may be misleading if one market's report includes single tenant, owner occupied properties, while the other does not. The first market's vacancy may appear to be lower as it includes properties that are generally not as well surveyed and also tend to have

Proper Prior Planning

I have been building a boat in my basement. It's a lot like golf in that it sounds like fun, but is an exercise in frustration. The main difference is that in the end, I'll have a boat. To make it float, a lot of parts need to come together perfectly. Little tricks, properly executed, save countless hours, minimize frustration, and produce beautiful results. They are usually simple, involving things like tracing a line beyond the edge of the piece, cutting wide on outside edges and tight on inside edges, and remembering to account for the thickness of the wood when laying out the ribs. I also have to remember not secure the pieces permanently together, until I move it from the basement to the garage.

When starting your analysis of commercial real estate, or any venture, it's critical to think it through before you start investing time, money, credibility, or anything you value. One technique is to visualize a process from beginning to end, talking through what will happen, and what is expected. This simple process, used by the Blue Angels before every flight, helps avoid mishaps. Give it a shot. You might look silly doing it, but it beats forgetting to prepare a critical component of your plan.

a lower natural vacancy rate. Even using the same source across geographies will not provide consistency as very few firms are able to enforce standards in all markets as each market may have its own accepted styles.



Many firms produce national and global reports by compiling data from different markets without adjustments for the differing methods. You can spot them easily by comparing the local market reports, to the national reports. If they are the same, you should be suspicious. Some of the larger firms attempt to produce reports using a single set of standards across all markets. One example is the CBRE U.S. Office Vacancy and Industrial Availability Indices. These reports use well enforced standards to provide a true cross market comparison. While they only release a narrow set of variables at the market level, they are quite valuable.

Users experience difficulties finding or creating long time series. Most available reports release only current period data in numeric form with graphs showing a limited history. These are designed to show off the power of the firms' internal data, available to their brokers and clients, without actually releasing it to the broader

audience. Occasionally you can tap into detailed data by request, but expect to spend some time. Don't be upset if your request is denied. After all, these firms spend millions collecting and analyzing the data for their clients' competitive advantage. You can try to create a longer time series by obtaining older editions of the reports and piecing together the data. But be warned.

Check carefully for changes in methodology with each edition of a report.

The firms make changes to the reports which break the free time series. For Example, the CBRE Indices switched the building sets used several times over the thirty year history of the report, most recently in the 2009Q1 editions. Formulae using a time series of vacancy from that report need to account for the shift. CBRE did an excellent job in alerting users to the change. Others are not so forthcoming. That's why you need to take the time to read the reports completely, along with any release notes, and compare the graphs to prior editions to see if there are any unexplained shifts in levels. If the numerical observations obtained from a prior edition of a report don't seem to line up with the graphed numbers in a more recent edition, that's a clue to a shift in methodologies.

Commercial real estate history changes. What was once true can change over time. This is not due to politically motivated conspiracies, or even neglect, but due to the nature of the raw data. Most statistics are derived from a set of properties, sampled on a regular basis, to produce a meaningful trend. Sometimes properties are added to the sample to keep up with changes in the structure of the market. Sometimes, these changes are substantial as in the addition of a large geographical area in a growing market. In such a case, the history of those properties must be feathered into the statistics, requiring a revision of the prior aggregations. While this is cumbersome for users of the data to incorporate, it is necessary in ever growing markets. The best sources of data will account for the changes clearly and sometimes even produce parallel sets of statistics for a few editions to allow the users to adapt to the revisions. You need to check yourself to detect many of the changes. A few simple tests in your collected data will expose some of the most common failings. One of my favorites is this:

- 1) Create a time series of total inventory and construction completions for a market.
- 2) Calculate the change in inventory for each period and compare it to construction completions. It should match.
- 3) Differences in any period may indicate a serious shift and render the data highly suspicious.

Another common problem with time series data is the inclusion of increasing numbers of markets in national statistics. As the brokerage firms grew in size, they added the new markets to their statistics. This means that the definition of the nation changed over time. As a result, reported levels can not be used in variables such as construction completions, inventory, net absorption, and vacant space. It also means that ratios such as vacancy rates may be skewed as markets like Atlanta, with systemically high vacancy rates, are added. Worse are markets like Los Angeles which changed dramatically over the decades and have huge volumes of inventory. Unless you know which markets were added and when, it is impossible to repair the trend lines. Comparing erroneous time series to independent variables such as population, employment, or gross domestic product, leads to incorrect conclusions.

Databases and the Internet	
<p>Many decades ago the computer age promised paperless offices. Most people's desks would indicate that promise is still unfulfilled. Still, we do access and use a lot more information than we did fifty years ago, and most of us use a lot less paper. One of the most powerful computing tools at our disposal is the relational database. Tables, holding fields of data for millions of records, allow us to quickly filter, sort, and aggregate statistics. Computer models use these statistics to generate forecasts and scenarios, output in graphs, charts, reports, and yes, even paper. Compiling data into these databases requires careful entry and consistent collection methodologies for the statistics and models to be valid.</p> <p>Now the internet age is fully upon us. We can look up information on every conceivable subject from anywhere. Search engines like Google use sophisticated algorithms to deliver what we want, even if we are poor at describing it. But search engines are designed to provide us content to read and enjoy. How do we turn those results into data fit for a database? Services such as Fetch, Wolfram Alpha and Seaglex do an excellent job there, but don't think your work is over at that point. You still need to spend time training these tools, and be very careful on how you use the results. You still need to understand what you are putting into your models. That said, the power of databases and the internet come together in these tools, and you won't get a paper cut flipping through old publications.</p>	

XML - Creating Order Out of Chaos

Relational databases have rigid structures. Indexes allow the computer to rapidly filter and sort the data by limiting the information it processes. Database designers work hard to make sure that the right fields are indexed to return results quickly for common searches. They also make sure that data are grouped correctly into tables to reduce the storage space needed by eliminating duplication. This process of normalizing the data prepares it to work well with the computers' style.

But does that style work well with messy data? What if you have differing amounts of information for each record? What if some records include information on details that others do not? In a database, you have to create a field for each type of information you might want to store, and leave those fields blank when you don't have it. But if those fields are numeric, how do you differentiate between a zero, and a blank? Many databases fail to account for this, and as a result statistics produced from them are wrong. Also, users who see a field exists, and search for data on it, may get the wrong impression when the results returned include a limited set of records. Experienced database designers work with users to eliminate fields that won't be consistently populated with data.

This is where XML solves the problem. XML stands for Extensible Markup Language. It is a relative of HTML, the basic language of the World Wide Web. The key is the extensible part. In XML, if you need an extra field sometimes, you just add it. You only populate the data with fields you have. Missing data leaves no gaps. Each record can possess a different quantity of fields. While this method of storing information requires a lot more computer space and processing power, the flexibility is much better suited for messy data.

Finally, there are simple errors that you can make in compiling data from different sources. Pay special attention to the units of measure, and scrutinize the reports as they commonly include mistakes in this area. Other common issues include:

- California markets usually report rents on a monthly basis while the rest of the U.S. reports annually.
- The components included in the rent vary widely between gross, net, and hybrids of the two.
- Almost all reported rents are asking rents which tend to differ from actual negotiated rents and those differences shift through the cycles.
- Inclusion or exclusion of sublet space can affect rent and vacancy levels.
- The shifting of large properties between single and multi-tenancy can skew the data.
- The definition of vacancy can mean anything from physically empty, to not currently competitively marketed for lease.
- Exchange rates for currency conversions are reported as of a certain date. Comparisons using different conversion dates need to adjust for the rate changes.
- Aggregating rounded observations increases the rounding error.
- Averages for variables like rent should be weighted by total available space rather than a simple average.
- Watch for posted revisions to reports.

```
<?xml version="1.0"?>
  <Record>
    <Name>Bill McNasty</Name>
    <Company>Acme Inc.</Company>
    <Title>Associate</Title>
    <Birthday>1966-05-15</Birthday>
  </Record>
  <Record>
    <Name>Suzie Sunshine</Name>
    <Company>Sprockets Co.</Company>
    <Title>Vice President</Title>
  </Record>
</xml>
```

In the simple XML example above, one row has a birthday field (in red) while the other does not.

There are as many mistakes to make in reporting data as there are methods to compute it. Frequent turnover in the staff hired to do the job contribute to the problem.

You need to make adjustments when combining data from different sources.

These adjustments can produce plausible data series, or result in even more errors, so be careful. A checklist is recommended.

The best solution for anyone truly interested in understanding the markets is to subscribe to long term, cross market data from a source such as PPR, TWR, REIS, RCA, or others. Each takes the time to cleanse the underlying issues and produce outputs that are usable in your own models. Many also provide econometric forecasts that tie the CRE variables to macro-economic forecasts from Moody's Economy.com or Global Insight. Some users complain that the data do not represent what they see on the ground in a particular market, but these differences are usually accounted for by the processes taken to assimilate each market into a consistent set of practices needed for accurate cross market, long term comparisons. The biggest hurdle for most users is the high cost for accessing these sources.

Another approach is to access the property level data directly. You could create your own survey of individual properties, contacting building owners and leasing agents regularly for updated data on availability as well as keeping up with new construction and zoning changes with data from First American CoreLogic or directly from the counties. Prepare to spend about twenty minutes per property per quarter. Or, you could subscribe to CoStar, Xceligent, or other property level data providers. This too is expensive but depending upon the size of the area you want to cover, may be less expensive than doing the survey work yourself. Here though you will run into problems in trying to export the data to work in your own databases and spreadsheets. Some of the sources contractually and practically prohibit exporting and storing the data and limit the derivative works you can create from it. Additionally, few of the property data sources have covered many markets prior to 2000. Since 2000 we have experienced a full economic cycle. Forecast models are more reliable when inputs span more cycles.

Messy data requires experience to manage into reliable answers. One of my favorite sources for economic data is the Blue Chip series which compiles economic forecasts from different sources and creates a consolidated, consensus view. This provides a less biased forecast. Since economic forecasts can be somewhat of a self-fulfilling prophecy, getting this type of answer is the best method of sensing what the market sees. A similar approach to CRE data is equally useful, even though it takes a lot of work and know-how to complete. Luckily for you, I have done the work and have the inside experience on how the data are collected, prepared, and released. Thus, I can save you time and produce a better result. Let's talk about what you want to better understand. I can help you get there a lot quicker and avoid costly mistakes.

What Really Happened Here?

This newsletter is more than it appears. It is a mix of the following skills:

Research - Writing in an accurate, informative and compelling manner.

Marketing – Getting the message in front of the right people – You! Finding topics of interest, and gauging that interest to increase response.

I.T. – The distribution email was generated from a program I wrote that embeds unique links into each email while still sending it through Microsoft Outlook so I can find it in my Sent Items. The embedded links help me track when you viewed the email. When you continued to the web to view the article, it tracked that too, linking it to the originating email. These databases, programs, emails, and web pages are all integrated around campaigns. I wrote each component using simple tools to avoid constraints and costs of third party tools as well as wasted time dealing with version upgrades.

Project Management – Getting it all done in a reasonable amount of time meant making choices around content, quality, timeliness, tool sets, speed, and more. Incorporating the efforts of proofreaders, editors, permissions from sources, and other demands on time requires strong project management experience.

Strategy – You are looking at it. I am an experienced professional with success in multiple disciplines. The best way to explain what I can do is to show you. In searching for work, I apply strategies to networking and self promotion. The point is not the content, but the approach. Wouldn't you want someone who puts this much thought and care into everything they do, working for you?

So how can I help you? I am available to work full time or as a consultant. Let me help you take your business to the next level. Please call today to discuss your ideas.

Regards,



Ward S. Caswell
ward@caswell.org
(617) 304-2689