STA 371H Homework 10: Capstone Case Studies

Hard-copy write-ups due Friday, May 6, 2016, by 5PM
Turn in to my office (CBA 6.478) or my mailbox in CBA 5.202

You have learned many important skills in this course. Now is your chance to turn those skills loose on two realistic case studies.

1) A hospital audit: assessing radiologist performance

The case

The data in "brca.csv" consist of 987 screening mammograms[1] administered at a hospital in Seattle, Washington. Five radiologists, each of whom frequently read mammograms, were selected at random from those at the hospital. For each radiologist, roughly 200 of the mammograms each had read were selected at random. Each row of the data set corresponds to a single woman's mammogram. The radiologist who read it is identified by a three-number code (1-999).

For each patient, two outcomes are recorded. The first is an indicator of whether the patient was recalled by the radiologist for further diagnostic screening after the radiologist read the mammogram (1=Recalled for further diagnostic screening, 0=Not recalled). The second outcome is an indicator of whether there was an actual diagnosis of breast cancer within 12 months following the screening mammogram (1=Yes, 0=No). Ideally, the radiologist should be: (1) minimizing false negatives, i.e. recalling the patients who do end up getting cancer, so that they can be treated as early as possible; and (2) also minimizing false positives, i.e. not recalling the patients who do not end up getting cancer, so that they are not alarmed unnecessarily. Of course, this ideal not attainable. Mammography is inexact, and sometimes there will be mistakes.

The data

In addition to the cancer and recall outcomes, several risk factors for breast cancer identified in previous studies are provided in the data set. Referent values for a "typical female" are indicated by asterisks:

age: 40-49*, 50-59, 60-69, 70 and older
family history of breast cancer: 0=No*, 1=Yes
history of breast biopsy/surgery: 0=No*, 1=Yes

[1] `https://www.nlm.nih.gov/medlineplus/mammography.html`

breast cancer symptoms: 0=No*, 1=Yes
menopause/hormone-therapy status: Pre-menopausal, Post-menopausal & no hormone replacement therapy (HT), Post-menopausal & HT*, Post-menopausal & unknown HT
previous mammogram: 0=No*, 1=Yes
breast density classification: 1=Almost entirely fatty, 2=Scattered fibroglandular tissue*, 3=Heterogenously dense, 4=Extremely dense

Audit goals

The goal of this case study is to examine the performance of the radiologists. This kind of statistical audit is a crucial link in the chain of modern evidence-based hospital practice. Specifically, your audit should address two questions.

(A) Are some radiologists more clinically conservative than others in recalling patients, holding patient risk factors equal?

Some advice: imagine two radiologists who see the mammogram of a single patient, who has a specific set of risk factors. If radiologist A has a higher probability of recalling that patient than radiologist B, we'd say that radiologist A is more conservative (because they have a lower threshold for wanting to double-check the patient's results). So if all five radiologists saw the same set of patients, we'd easily find out whether some radiologists are more conservative than others.

The problem is that the radiologists don't see the same patients. So we can't just look at raw recall rates—some radiologists might have seen patients whose clinical situation mandated more conservatism in the first place. Can you build a regression model that addresses this problem, i.e. that holds risk factors constant in assessing whether some radiologists are more conservative than others in recalling patients?

(B) When the radiologists at this hospital interpret a mammogram to make a decision on whether to recall the patient, does the data suggest that they should be weighing some clinical risk factors more heavily than they currently are?

Again, some advice: let's focus on family history as a specific risk factor (a similar line of reasoning applies to any risk factor). Consider two different regression models: Model A, which regresses a patient's cancer outcome on the radiologist's recall decision; and Model B, which regresses a patient's cancer outcome on the radiologist's recall decision AND the patient's family history. If that the radiologist were appropriately accounting for a patient's family history of breast cancer in interpreting the mammogram and deciding whether to recall the patient for further screening, would you expect that Model B would be any better than Model A at predicting cancer status? Why or why not? If it turns out that Model B is significantly better than Model A, what does that say about the radiologist's process for making a recall decision?

2) Green certification in commercial real estate

In this case study, you will conduct a systematic analysis of the economic value of environmentally friendly buildings, as measured in the marketplace for commercial rental properties (e.g. office buildings). Although the specific decision problem I ask you to consider is hypothetical, both the issue and the data set are real.

The case

Over the past decade, both investors and the general public have paid increasingly close attention to the benefits of environmentally conscious buildings. There are both ethical and economic forces at work here.

To quote a recent report entitled "Energy efficiency and real estate: Opportunities for investors" created by Mercer, an investment-consulting firm:

> Investing in energy efficiency has two intertwined virtues that make it particularly attractive in a world with a changing climate and a destabilized economy: It cuts global-warming greenhouse gas emissions and saves money by reducing energy consumption. Given that the built environment accounts for 39 percent of total energy use in the US and 38 percent of total indirect CO2 emissions, real estate investment represents one of the most effective avenues for implementing energy efficiency.

This only scratches the surface. In commercial real estate, issues of eco-friendliness are intimately tied up with ordinary decisions about how to allocate capital. Every new project involves negotiating a trade-off between costs incurred and benefits realized over the lifetime of the building. In this context, the decision to invest in eco-friendly buildings could pay off in at least four ways.

1. Every building has the obvious list of recurring costs: water, climate control, lighting, waste disposal, and so forth. Almost by definition, these costs are lower in green buildings.

2. Green buildings are often associated with better indoor environments—the kind that are full of sunlight, natural materials, and various other humane touches. Such environments, in turn, might result in higher employee productivity and lower absenteeism, and might therefore be more coveted by potential tenants. The financial impact of this factor, however, is rather hard to quantify *ex ante*; you cannot simply ask an engineer in the same way that you could ask a question such as, "How much are these solar panels likely to save on the power bill?"

3. Green buildings make for good PR. They send a signal about social responsibility and ecological awareness, and might therefore command a premium from potential tenants who want their customers to associate them with these values. It is widely believed that a good corporate image may enable a firm to charge premium prices, to hire better talent, and to attract socially conscious investors.

4. Finally, sustainable buildings might have longer economically valuable lives. For one thing, they are expected to last longer, in a direct physical sense. (One of the core concepts of the green-building movement is "life-cycle analysis," which accounts for the high front-end environmental impact of acquiring materials and constructing a new building in the first place.) Moreover, green buildings may also be less susceptible to market risk—in particular, the risk that energy prices will spike, driving away tenants into the arms of bolder, greener investors.

Of course, much of this is mere conjecture. At the end of the day, tenants may or may not be willing to pay a premium for rental space in green buildings. We can only find out by carefully examining data on the commercial real-estate market.

The data

The file `greenbuildings.csv` contains data on 7,894 commercial rental properties from across the United States. Of these, 685 properties have been awarded either LEED or EnergyStar certification as a green building. You can easily find out more about these rating systems on the web—e.g. `www.usgbc.org`. The basic idea is that a commercial property can receive a green certification if its energy efficiency, carbon footprint, site selection, and building materials meet certain environmental benchmarks, as certified by outside engineers.

A group of real estate economists constructed the data in the following way. Of the 1,360 green-certified buildings listed as of December 2007 on the LEED or EnergyStar websites, current information about building characteristics and monthly rents were available for 685 of them. This information came from the

CoStar Group (`www.costar.com`). In order to provide a control population, each of these 685 buildings was matched to a cluster of nearby commercial buildings in the CoStar database. Each small cluster contains one green-certified building, and all non-rated buildings within a quarter-mile radius of the certified building. On average, each of the 685 clusters contains roughly 12 buildings, for a total of 7,894 data points.

The columns of the data set are coded as follows:

CS.PropertyID: the building's unique identifier in the CoStar database.

cluster: an identifier for the building cluster, with each cluster containing one green-certified building and at least one other non-green-certified building within a quarter-mile radius of the cluster center.

size: the total square footage of available rental space in the building.

empl.gr: the year-on-year growth rate in employment in the building's geographic region.

Rent: the rent charged to tenants in the building, in dollars per square foot per calendar year.

leasing.rate: a measure of occupancy; the fraction of the building's available space currently under lease.

stories: the height of the building in stories.

age: the age of the building in years.

renovated: whether the building has undergone substantial renovations during its lifetime.

class.a, class.b: indicators for two classes of building quality (the third is Class C). These are relative classifications within a specific market. Class A buildings are generally the highest-quality properties in a given market. Class B buildings are a notch down, but still of reasonable quality. Class C buildings are the least desirable properties in a given market.

green.rating: an indicator for whether the building is either LEED- or EnergyStar-certified.

LEED, Energystar: indicators for the two specific kinds of green certifications.

net: an indicator as to whether the rent is quoted on a "net contract" basis. Tenants with net-rental contracts pay their own utility costs, which are otherwise included in the quoted rental price.

amenities: an indicator of whether at least one of the following amenities is available on-site: bank, convenience store, dry cleaner, restaurant, retail shops, fitness center.

cd.total.07: number of cooling degree days in the building's region in 2007. A degree day is a measure of demand for energy; higher values mean greater demand. Cooling degree days are measured relative to a baseline outdoor temperature, below which a building needs no cooling.

hd.total07: number of heating degree days in the building's region in 2007. Heating degree days are also measured relative to a baseline outdoor temperature, above which a building needs no heating.

total.dd.07: the total number of degree days (either heating or cooling) in the building's region in 2007.

Precipitation: annual precipitation in inches in the building's geographic region.

Gas.Costs: a measure of how much natural gas costs in the building's geographic region.

Electricity.Costs: a measure of how much electricity costs in the building's geographic region.

cluster.rent: a measure of average rent per square-foot per calendar year in the building's local market.

You will notice several systematic differences between the green-certified buildings and their non-certified counterparts. To cite just three examples, the green buildings tend to be newer, bigger, and of a higher quality. In your initial exploration of the data, you will probably find many other differences, too!

Goals

An Austin real-estate developer has hired you to assess the possible economic impact of "going green" in her latest project: a new 15-story mixed-use building on East Cesar Chavez, just across I-35 from downtown. The property would have restaurants and retail stores on the lower floors, and office buildings on the upper floors. The total planned area would be 250,000 square feet of rental space. Two other things worth mentioning are: (1) in terms of its climate, economy, and energy prices, Austin is like Cluster 969; and (2) the real-estate developer that hired you typically runs her buildings on a "net-rental" basis, and wants to do the same here.

The issue at hand is: will investing in a green building be worth it, from an economic perspective? If

so, how many years might it take until the extra up-front investment in green modifications turns a net profit, accounting for opportunity cost: that is, the fact that this money could have been invested elsewhere instead? There are several sources of uncertainty in this calculation. One source, obviously, is the role of green certification in producing greater future rental income, but other sources include:

(1) The baseline construction costs, assuming no green modifications. The developer estimates that these costs will come in at $50 million, or about $200 per square foot, not counting any modifications necessary to achieve green certification. Furthermore, she expects that they will fall between $40 and $60 million with probability 0.95.

(2) The additional costs necessary to achieve green certification. In conjunction with an engineer, the developer estimates that these modifications will add something like a 7% premium to the final construction costs, and that this premium will end up falling between 4% and 10% with probability 0.95.

(3) The real rate of return that the developer could earn by taking the extra money intended for green modifications, and investing it elsewhere. Your client respects your business acumen, and therefore wants to get your input on realistic alternatives here.

Remember that the return on an investment like this is a function not just of the rents, but also of the occupancy rate. (Commercial developers sometimes measure the effective rental rate as the rate per square foot times the occupancy rate.)

You are to examine the available data on green buildings; use your skills in statistcal modeling and decision analysis to form an opinion as to whether pursuing green certification for this project seems financially sensible; and prepare a report summarizing your methods and conclusions. You may assume that the developer has $70 million of capital for the project, and the question is how to allocate this capital. I recommend that you attack this problem by modeling the risk-return profile, over time, of two possible scenarios:

- build a normal building, and invest the balance of the available capital somewhere else, or
- build a green building, and invest the balance of the available capital somewhere else.

Under the second scenario, of course, the "balance of the available capital" will be a bit higher.

Your client is statistically literate, but she will be skeptical if all you do is present the results of an analysis without justifying the approach you have taken. She will therefore want you to explain clearly what you have done (using pictures where relevant), and to give an honest assessment of how much confidence one should place in your conclusions.

You may ignore the effect of inflation on future rental prices; thus implicitly you will be modeling real returns, rather than nominal returns, and assuming that future rental prices will be adjusted for inflation appropriately. You can also ignore other accounting issues like depreciation, taxes, NPV, and so forth; your client has other consultants for that! Just model the cash streams.

Some notes further notes. (1) If you feel like you need to make some extra assumptions somewhere, state what they are and justify them briefly. (2) If you feel like you don't know how to handle some aspect of the problem, take your best shot, and explain what you have done/would like to do, and why. (3) You don't need to assume a particular utility function for your client. Instead, assume that she'd like to understand both the expected payoffs and volatility of payoffs under the competing scenarios, and let her decide for herself using her own utility function. (4) Please be concise. In the real world, when someone commisions a reportl like this, they want the headline conclusions, together with the main evidence in favor of those conclusions, summarized up-front in a few pages. Then there should be appendices for the (important) technical stuff. It's good to practice using this format now. (5) This report doesn't need to be long, just clear and competent.