



**StatCities** Trento

**La vendemmia statistica**

Linee operative e  
prospettive di riforma  
del sistema statistico  
nazionale a livello  
locale

**14 e 15 settembre 2017**

Palazzo Geremia  
via Belenzani, 20  
TRENTO



## La Provincia di Trento e la nuova riforma del catasto

**Gianfranco Cerea, Roberto Gabriele, Giuseppe Espa**

Dipartimento di Economia e Management, Università di Trento

## Perche' una nuova stima dei valori catastali

1. Valori del 1939 basati su una realtà economica, territoriale e sociale superate;
2. Aggiornamenti inadeguati e applicati in modo difforme sul territorio
3. Definizione delle rendite con criteri e riferimenti inadeguati
4. Necessità di disporre di valori «realistici» per poter applicare un prelievo locale equo, basato sul principio del beneficio:
  - Per la tassazione ricorrente;
  - Per il concorso al finanziamento di specifiche opere pubbliche

## Il valore delle case

1. Le abitazioni sono un “bene” con caratteristiche particolari
  - Durevole nel lungo termine
  - Eterogeneo ( ogni casa è diversa da un'altra)
  - Fisso nello spazio
2. Al contempo bene di consumo e bene di investimento
3. Costo molto elevato
4. Processo produttivo “lungo” per l'immissione di nuove abitazioni
5. Variazioni comunque modeste dello stock nel breve termine
6. Scambi che riguardano una frazione piccola dello stock (3-5%)

## Il valore delle case

- ▶ Con tali presupposti le indicazioni che vengono dal “mercato” non possono essere considerate direttamente rappresentative del valore delle abitazioni o del prezzo a cui le stesse possono essere cedute.
- ▶ In altri termini: *conoscere i m<sup>2</sup> e l'ubicazione non basta*

## Il valore delle case

- Secondo il modello di Lancaster (1966) non sono i beni ad entrare nella funzione di utilità di un consumatore ed a determinarne il livello, bensì sono le caratteristiche dei beni stessi a determinare l'utilità del consumatore.
- Questi tipi di modelli assumono dunque che beni eterogenei possano essere descritti sulla base di specifiche caratteristiche il cui valore implicito (prezzi edonici) è ottenibile partendo dai prezzi osservati dei beni

## Le possibili determinanti del valore della casa

### *Caratteristiche specifiche*

- 1) superficie (definita in modo omogeneo ed univoco),
- 2) grandezza del terreno in cui sorge l'edificio,
- 3) età,
- 4) numero di stanze da letto,
- 5) numero di bagni,
- 6) garage.
- 7) balconi e terrazze,
- 8) classe energetica;
- 9) riscaldamento centralizzato/autonomo,
- 10) piano di ubicazione,
- 11) numero piani/unità immobiliari edificio,
- 12) ascensore.

# Le possibili determinanti del valore della casa

## ► **Le caratteristiche del contesto**

- 1) *le caratteristiche sociodemografiche della popolazione del territorio in cui si trova l'immobile;*
- 2) *la concentrazione e il livello di servizi presenti nel territorio (sia pubblici che privati): es i servizi scolastici, trasporti e sistemi di trasporto; effetti negativi dovuti alla presenza di grandi infrastrutture di trasporto (aeroporti, ferrovia, autostrada), centri commerciali ed esercizi commerciali. impianti sportivi; per le aree turistiche l'accessibilità agli impianti sciistici e di elementi di attrazione.*
- 3) *la qualità dell'ambiente naturale del territorio*

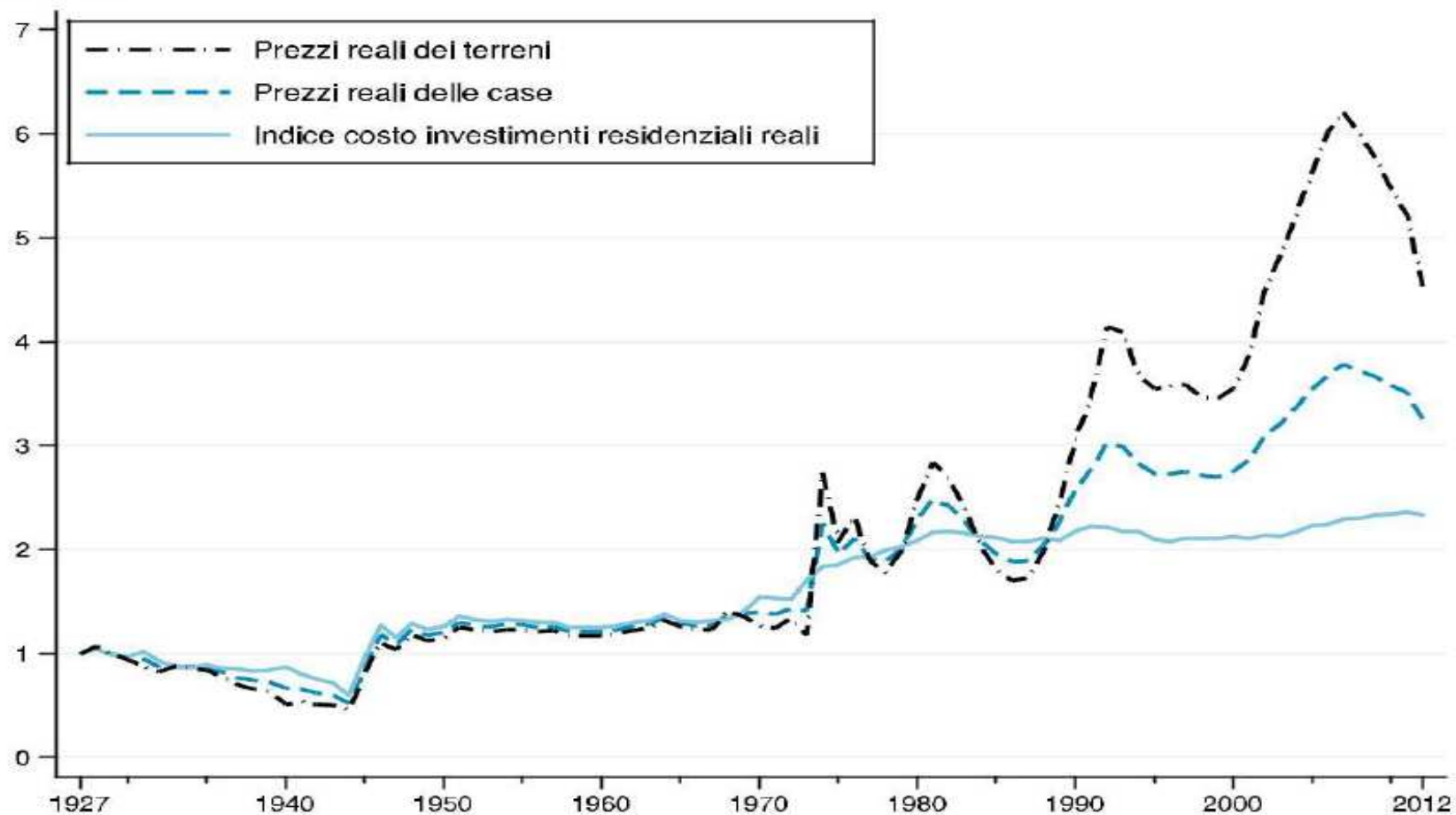
## Le possibili determinanti del valore della casa

### ► **Il contesto macroeconomico:**

1. *Tassi di interesse sui mutui;*
2. *Accesso al credito;*
3. *Prospettive dell'economia;*
4. *Andamento dei prezzi*



## Il mattone non tradisce mai !?



## Cosa stimare?

- VALORE DI MERCATO versus VALORE NORMALE

## Le fonti informative

Alcune domande:

- Quali sono le **richieste in termini di dati** per un simile progetto?
- Chi sono i **soggetti** da coinvolgere?
- Che tipo di **conoscenze** si devono impiegare per trattare i dati?
- Di quali **strumenti** c'è bisogno (server, programmi, ecc.)?

## Dati necessari

- Informazioni su **singoli immobili**:
  - Elementi Strutturali
- Informazioni sul **contesto socioeconomico e aspetti demografici** nel quale è in essere l'immobile
- **Ubicazione** dell'immobile (caratteristiche geografiche)

*Ma soprattutto:*

- Informazioni su **compravendite**:
  - Prezzi

## Dati necessari II

Che **arco di tempo** scegliere per l'analisi?

► Trade off:

► Finestra temporale più lunga, più informazioni

*ma...*

► Problemi di comparabilità (evoluzione del contesto di riferimento)

## Il contesto della ricerca

- **Periodo di analisi:** 2002-2016
- **Contesto:** Provincia di Trento
- **Obiettivo:** Individuare un metodo “oggettivo” per valutare i valori degli immobili in Provincia di Trento

## La “risoluzione” dei dati

- Livello desiderabile:
  - “Micro” dati su immobili:
    - Più fine possibile (caratteristiche degli immobili)
- Dati geografici:
  - Scelta del livello di risoluzione ottimale
  - Scelta della zonizzazione del territorio
    - Coerente con le definizioni amministrative
    - Meaningful per le analisi e le stime da effettuare

## La “risoluzione” dei dati II

- ▶ La questione dei **Valori di mercato** (prezzi) come elemento centrale:
  - ▶ Quali sono i **prezzi** da considerare:
    - ▶ I prezzi al metro quadrato
  - ▶ Da dove **recuperare l'informazione**?
    - ▶ Le compravendite
  - ▶ Come **correggere eventuali anomalie**:
    - ▶ Metodi empirici (inputazione dati integrando le fonti)
    - ▶ Metodi statistici



## Gestire i problemi

- In questo contesto appare cruciale prevedere protocolli “robusti” per gestire aspetti critici

In particolare:

Gestione delle:

- **Inconsistenze** delle informazioni
- **Ridondanze** (se disponibili più fonti quali scegliere)
- **incompletezze** (dati mancanti)

## Soggetti coinvolti

- **Soggetti istituzionali:**
  - Catasto, Comuni, Agenzia delle Entrate
  - ISPAT
- **Soggetti privati:**
  - Imprese partner del progetto (diverse expertise)
- **Università:**
  - Conoscenze tecniche di trattamento di **Big Data**
  - Conoscenze di **metodi statistico – econometrici**
  - Conoscenze di **teoria economica** (posizionamento del lavoro)

## Passi dello studio

1. Mappatura dei dati necessari
2. Mappatura dei soggetti proprietari del dato
3. Analisi delle diverse fonti di dati
4. Omogeneità delle basi di dati (dialogo tra diverse banche dati)
5. Verifica dei "termini di utilizzo" delle diverse informazioni (ad esempio: protezione privacy)
6. Costruzione di un database per usi statistici
7. Analisi economico-statistica

## La stima dei prezzi edonici

La stima di una funzione di prezzo edonica richiede:

- Un campione di prezzi di vendita in più istanti
- Informazioni sulle caratteristiche delle singole unità vendute

$$P = f(X) + \varepsilon$$

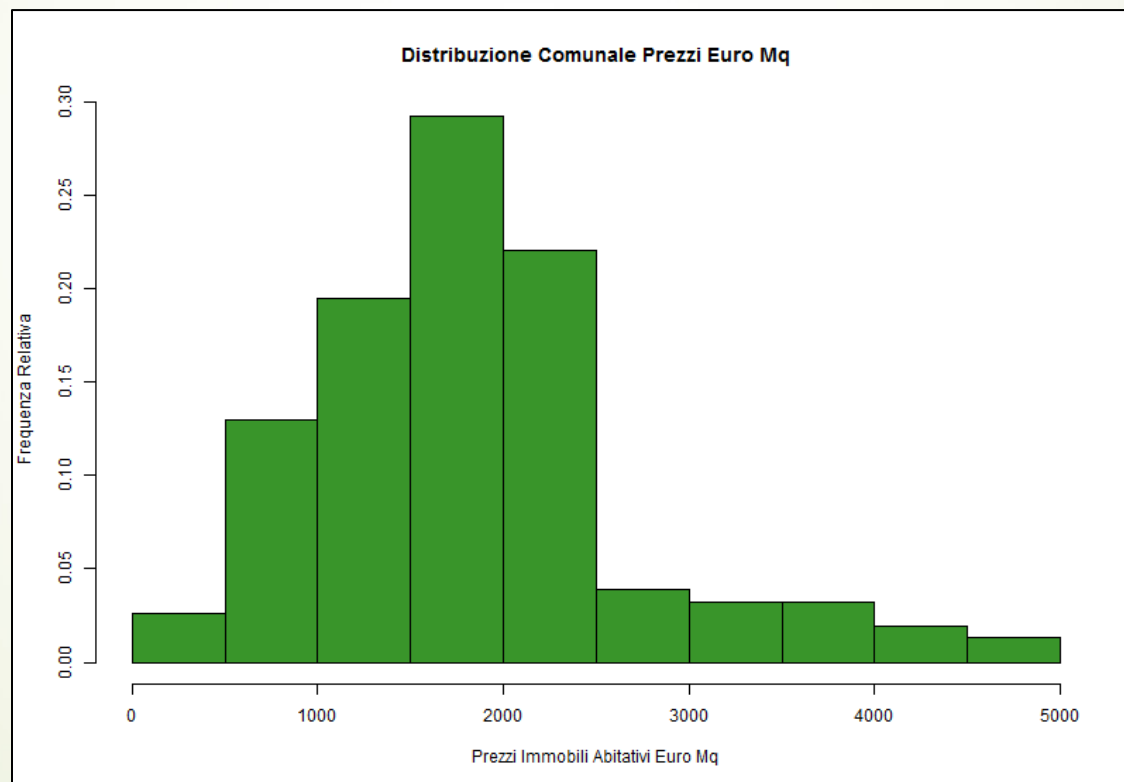
Regressione

$$P = f(S, N, G) + \varepsilon$$

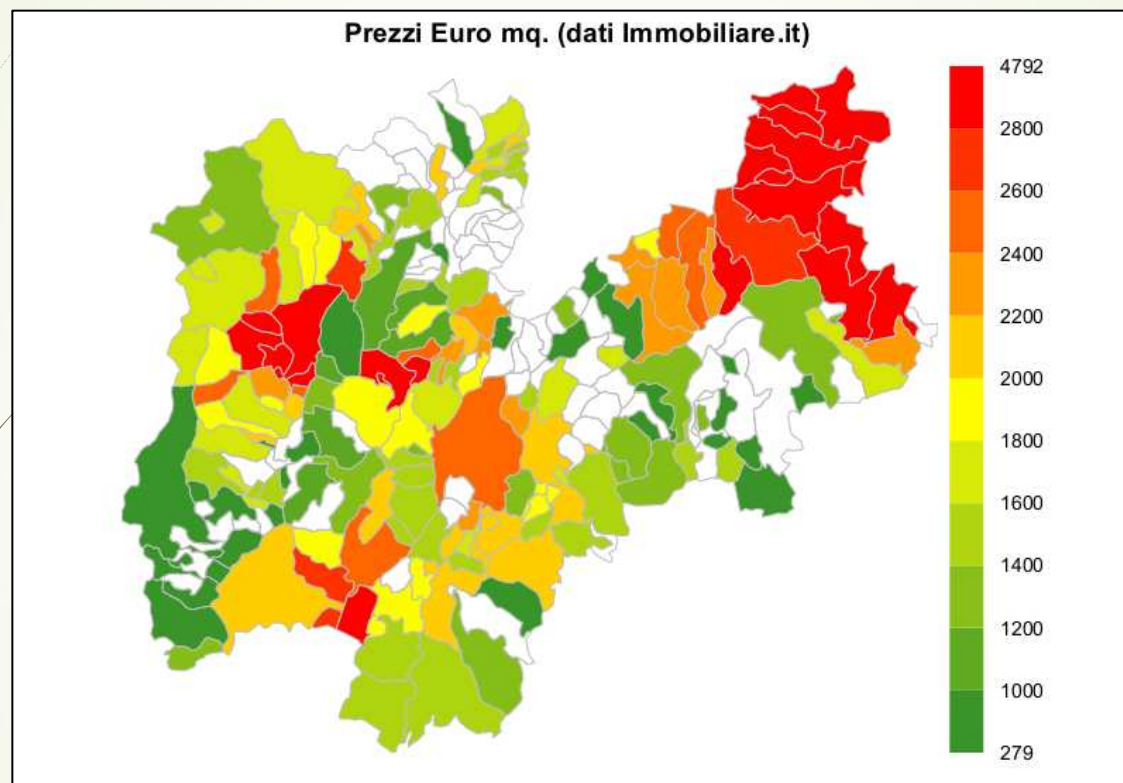
- $S$  = caratteristiche strutturali dell'immobile (numero di stanze, dimensione, qualità della costruzione, etc.);
- $N$  = caratteristiche dell'area di ubicazione (demografiche, socio-economiche, ambientali);
- $G$  = caratteristiche geografiche (distanza dal centro città, accessibilità ai servizi, contiguità rispetto a determinate aree, etc.)

## Un esercizio su dati aggregati: determinanti a livello comunale

- $P$  = prezzo medio comunale a m<sup>2</sup> degli immobili residenziali (marzo 2017; fonte: immobiliare.it)



## Variabili e modello



- Le  $X$  sono variabili dal Censimento della popolazione e delle abitazioni 2011 integrate con dati Istat non censuari.
- La  $f$  è un albero decisionale (CART; Breiman *et al.*, 1984)

**Variabili censuarie**

1. Sezione - Popolazione
  - 1.1. Densità abitanti (residenti e non) per Km<sup>2</sup>
  - 1.2. Percentuale popolazione residente oltre 85 anni
  - 1.3. Stranieri residenti
2. Sezione - Famiglie
  - 2.1. Famiglie per numero di componenti
  - 2.2. Famiglie in alloggio per titolo di godimento
3. Sezione - Alloggi ed edifici
  - 3.1. Superficie per occupante delle abitazioni occupate da persone residenti (valori medi)
  - 3.2. Abitazioni (occupate da residenti) per disponibilità di gabinetti
  - 3.3. Abitazioni (occupate da residenti) per disponibilità di impianto doccia e/o vasca da bagno
  - 3.4. Abitazioni (occupate da residenti) per disponibilità di servizi
  - 3.5. Abitazioni (occupate da residenti) per numero di stanze
  - 3.6. Edifici per stato d'uso
  - 3.7. Edifici residenziali per epoca di costruzione
  - 3.8. Edifici residenziali per numero interni
  - 3.9. Edifici residenziali per numero di piani fuori terra
  - 3.10. Edifici residenziali per tipo di materiale
4. Sezione - Istruzione e formazione
  - 4.1. Titolo di studio della popolazione residente di 6 anni e più
5. Sezione - Condizione professionale o non professionale
  - 5.1. Condizione professionale o non professionale della popolazione residente
6. Sezione - Caratteristiche dell'attività lavorativa
  - 6.1. Indicatori relativi al lavoro
  - 6.2. Occupati per sezioni di attività economica
7. Sezione - Pendolarismo
  - 7.1. Spostamento giornaliero della popolazione residente e motivo di spostamento
8. Sezione - Tipo di località abitata
  - 8.1. Popolazione residente e famiglie per tipo di località abitata
  - 8.2. Abitazioni per tipo di località abitata
  - 8.3. Edifici per tipo di località abitata

**Variabili non censuarie**

9. Sezione - Turismo
  - 9.1. Capacità degli esercizi ricettivi e movimento dei clienti negli esercizi ricettivi (alberghi e strutture simili, alloggi per vacanze e altre strutture per brevi soggiorni, aree di campeggio e aree attrezzate per camere roulotte)
10. Sezione - Struttura Imprese
  - 10.1. Unità locali e addetti per settori economici (Ateco 3 cifre)

## Elementi del CART

Gli elementi chiave di un CART sono un *set* di regole per:

- i) splittare ogni nodo dell'albero;
- ii) stabilire quando un albero è completo;
- iii) assegnare una etichetta (od un valore previsto nel caso della regressione) ad ogni nodo terminale

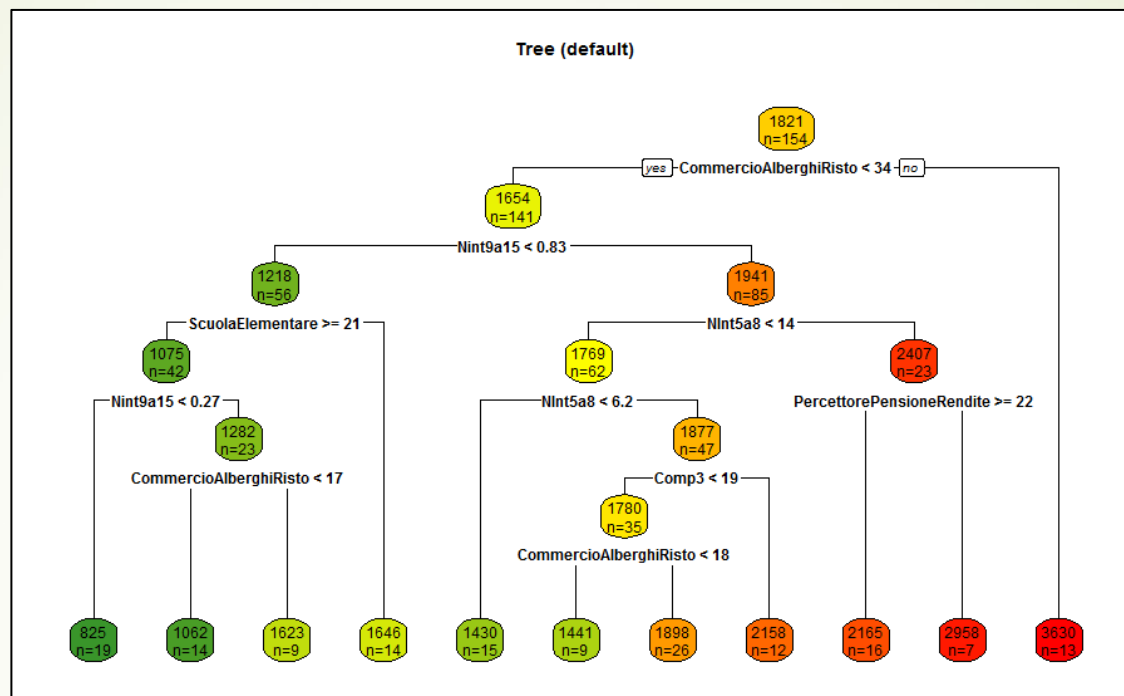
In particolare per i), si sceglie lo *split* che massimizza:

$$SS_T - (SS_L + SS_R).$$

Nel nostro esercizio:

- 10-fold cross-validation;
- parametro di complessità:  $\alpha = 0.01$



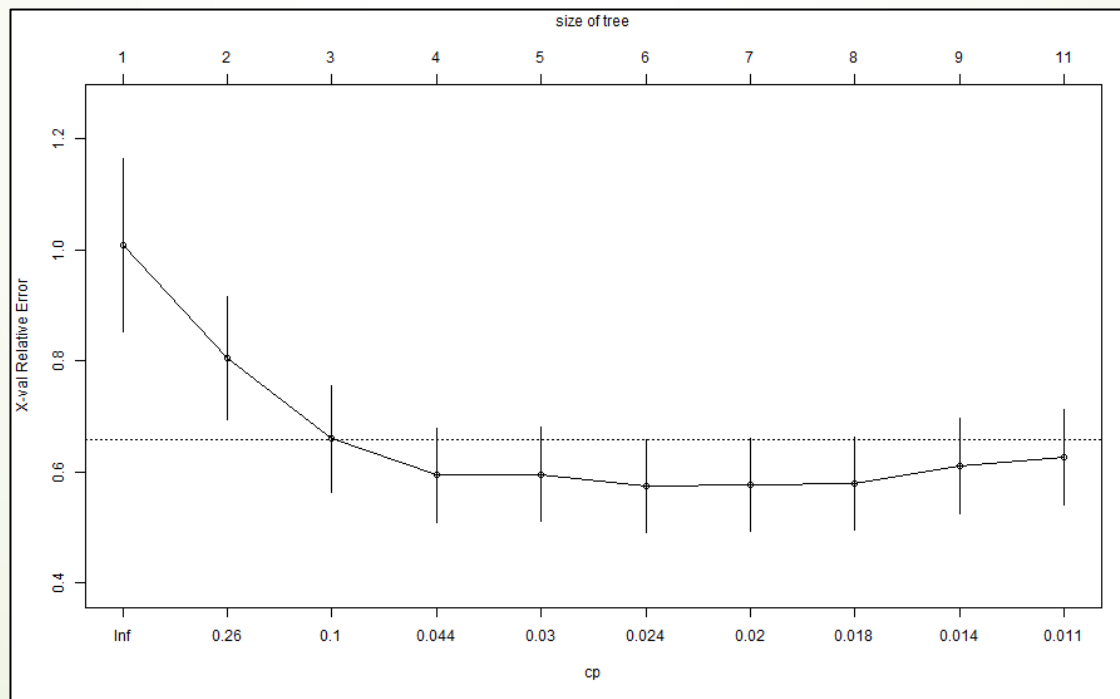


N.B.:  $154 < 217!$

(copertura parziale del territorio della fonte o  
assenza di compravendite)

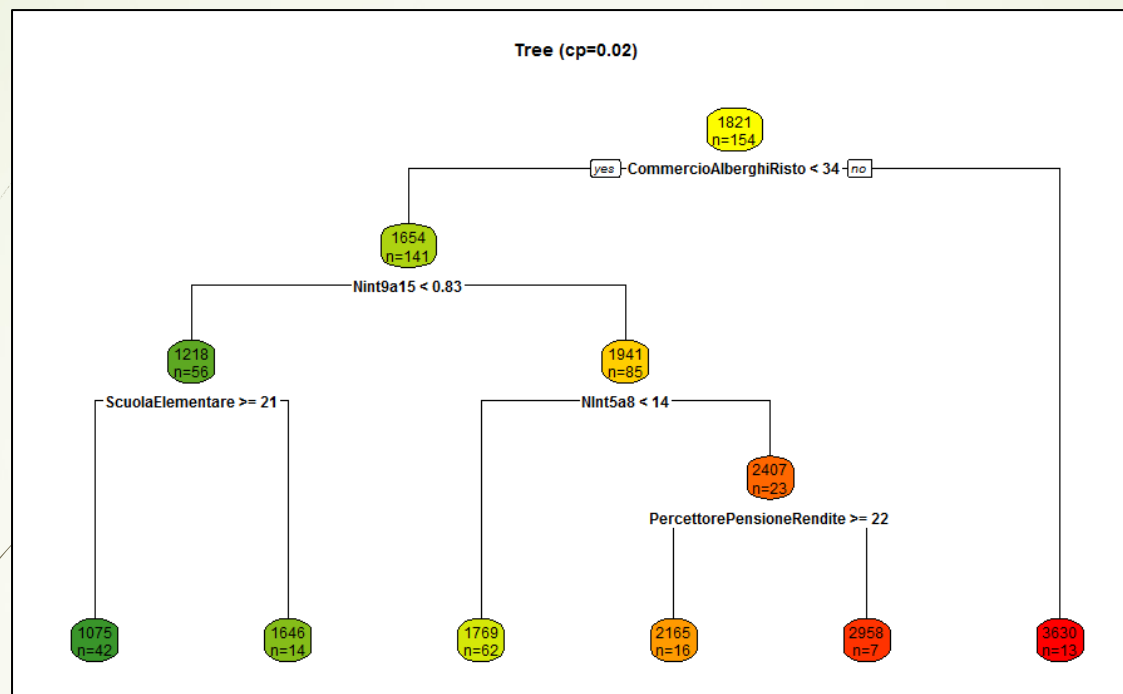
Nodi terminali con numerosità troppo ridotta: potiamo  
l'albero ma non in modo arbitrario (usiamo il criterio ii).

| CP         | nsplit | rel error | xerror    | xstd       |
|------------|--------|-----------|-----------|------------|
| 0,42709713 | 0      | 1         | 1,0073486 | 0,15572154 |
| 0,16235673 | 1      | 0,5729029 | 0,8051111 | 0,11155335 |
| 0,06276601 | 2      | 0,4105461 | 0,6601289 | 0,09564807 |
| 0,03143981 | 3      | 0,3477801 | 0,595092  | 0,08464003 |
| 0,02813164 | 4      | 0,3163403 | 0,5956611 | 0,08475294 |
| 0,02088856 | 5      | 0,2882087 | 0,5740753 | 0,08331931 |
| 0,01990208 | 6      | 0,2673201 | 0,5778689 | 0,08316509 |
| 0,01588702 | 7      | 0,247418  | 0,5788051 | 0,08357011 |
| 0,01226833 | 8      | 0,231531  | 0,6112781 | 0,08506049 |
| 0,01       | 10     | 0,2069944 | 0,6266094 | 0,0850117  |



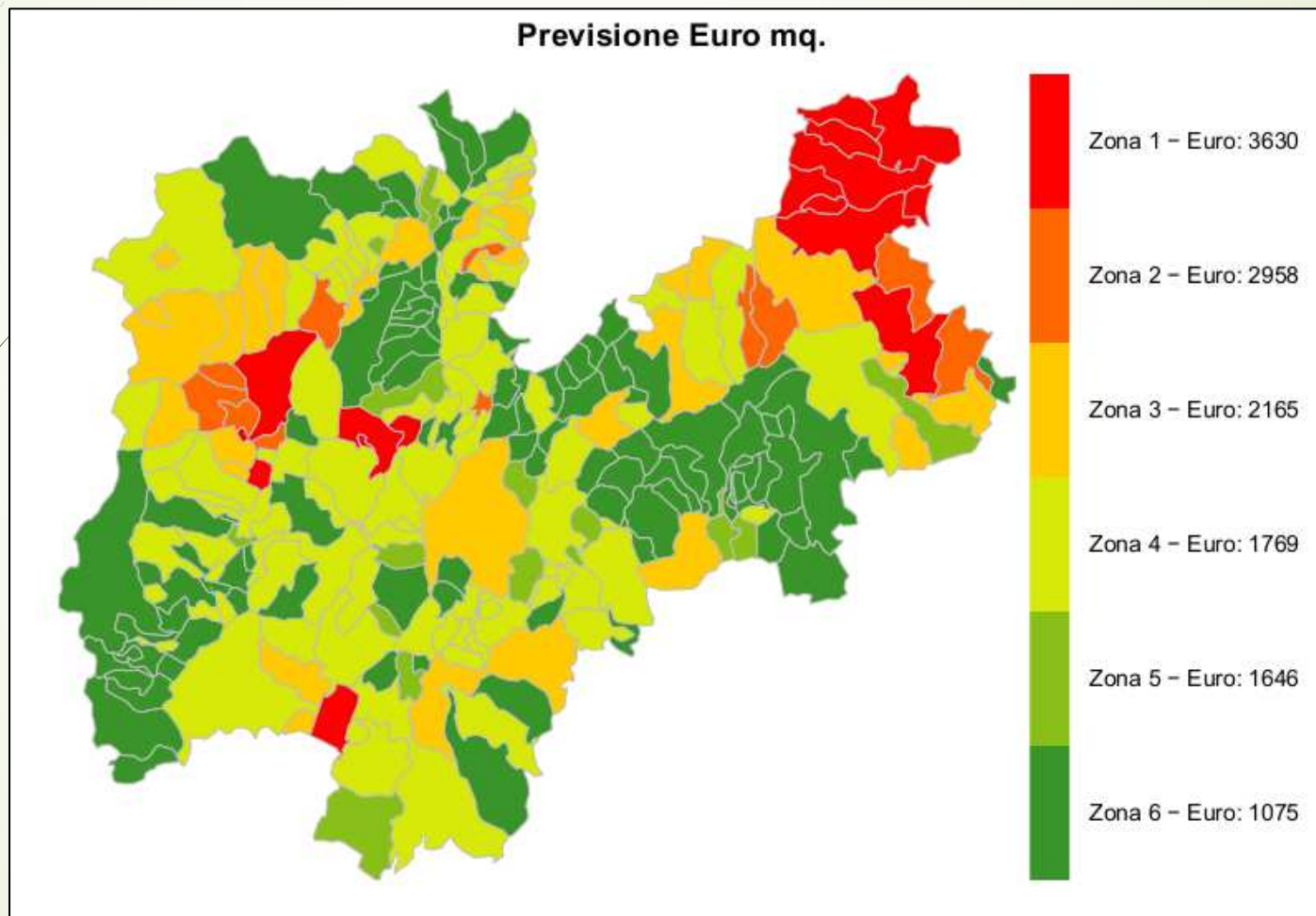
# Un albero “parsimonioso”

27

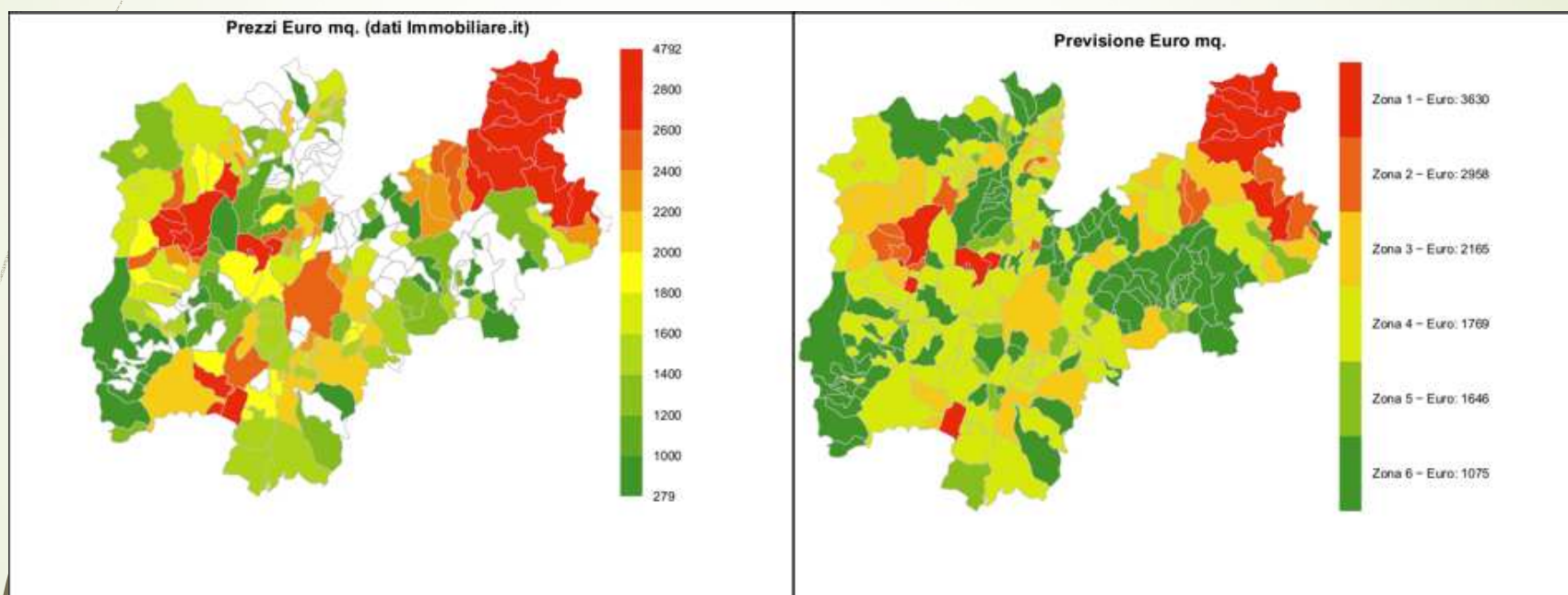


- Il primo *split* potrebbe identificare i comuni a vocazione turistica.
- Il secondo *split* potrebbe separare i comuni “rurali” dai non “rurali”.
- Il terzo *split* potrebbe catturare un effetto reddito degli abitanti residenti.
- Nelle zone “non-turistiche” e “non-rurali”, Il quarto *split* potrebbe separare i paesi dalle città.
- Il quinto *split* agisce sui comuni “non-turistici”, “non-rurali” e “di maggiori dimensioni” e usa la % di abitanti residenti che sono percettori di pensioni e rendite.

## Risultati: Una mappa previsionale



# Mappa reale e previsione

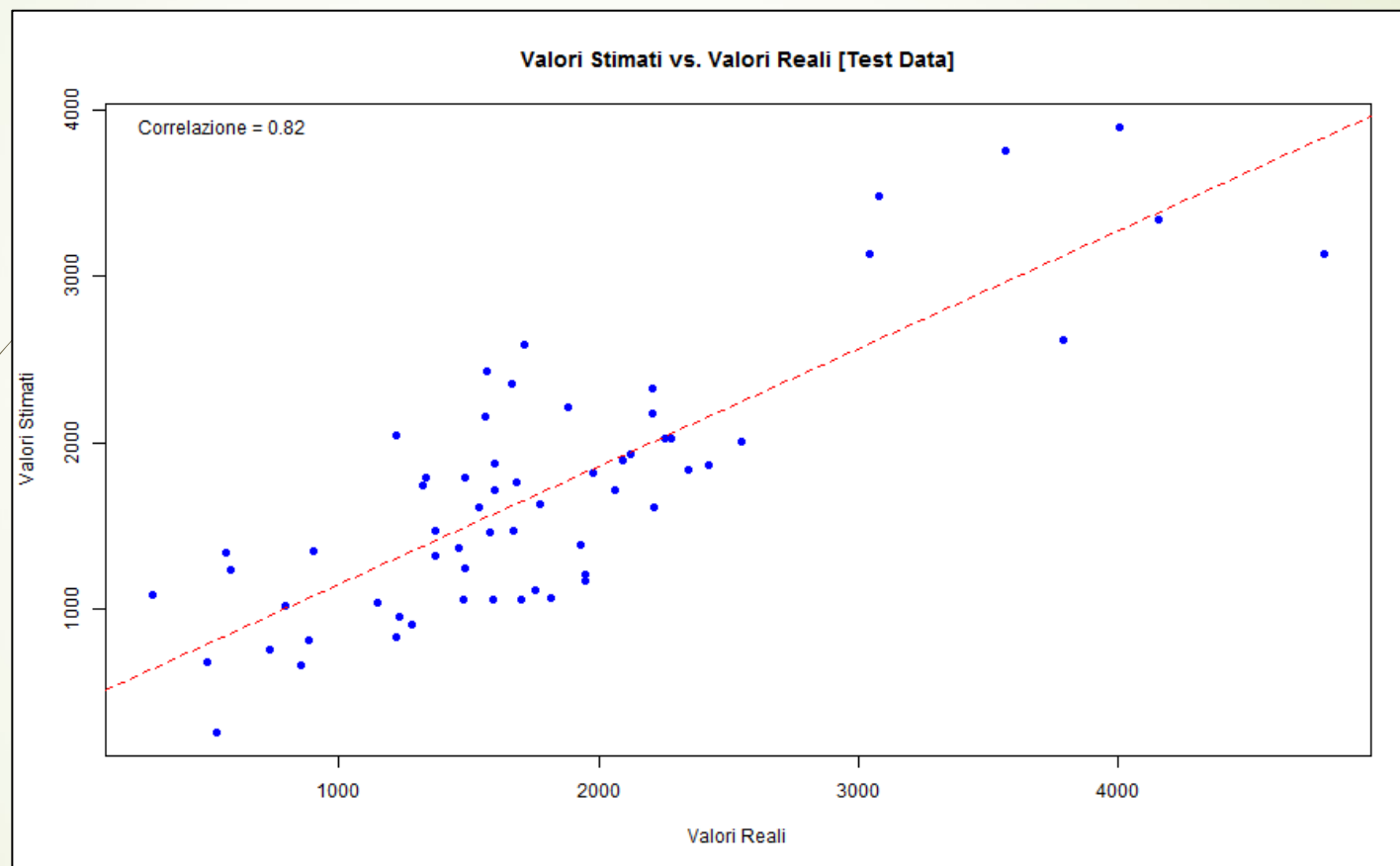


## CART: selezione di predittori per la modellistica tradizionale

- $X_1$  = % comunale della popolazione residente impiegata nel settore Alberghiero e del Commercio;
- $X_2$  = % comunale di edifici con un numero di interni abitativi da 9 a 15;
- $X_3$  = % comunale di edifici con un numero di interni abitativi da 5 a 8;
- $X_4$  = % comunale di residenti il cui titolo di studio più alto è quello di scuola elementare;
- $X_5$  = % comunale di residenti che sono percettori di pensioni e rendite.

$$P = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

# La bontà delle previsioni



## Uno sguardo avanti: uso dei micro dati

La stima dei prezzi edonici delle singole unità abitative basata su un modello del tipo

$$\log P_i = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_k X_{k,i} + \varepsilon_i$$

consente di tener conto di  $S$ ,  $N$  e  $G$ .

Ma la variabilità dei prezzi all'interno della provincia riflette solamente le differenze strutturali nelle abitazioni e nei servizi disponibili, oppure è anche il risultato di *differenze nel valore di mercato delle caratteristiche delle abitazioni?*



## La Geographically Weighted Regression

*Es.: un bagno aggiuntivo incrementa il valore dell'abitazione nella stessa misura a Trento e in un paese di montagna?*

► Nel secondo caso sarebbe troppo restrittivo imporre un “valore”  $\beta_h$  della caratteristica  $X_h$  unico per tutta la provincia

L'approccio *GWR (Geographically Weighted Regression)*

► Il problema può essere affrontato mediante una regressione in cui i prezzi  $\beta_h$  delle caratteristiche delle abitazioni varino a seconda dell'ubicazione delle abitazioni stesse:

$$\log P_i = \beta_0 + \beta_{1,i} X_{1,i} + \dots + \beta_{k,i} X_{k,i} + \varepsilon_i$$

## Caratteristiche del GWR

- In questo modo è possibile cogliere l'eterogeneità geografica dei prezzi delle caratteristiche delle abitazioni.
- I GWR consentono di ottenere le stime dei prezzi  $\beta_{h,i}$  per ogni abitazione cogliendo l'eventuale eterogeneità geografica *senza dover fare assunzioni sulla sua forma e sulla sua origine*.
- Quando alcune fonti di eterogeneità geografica non possono essere misurate o sono ignote, i GWR sono in grado tenerne conto con un generale miglioramento della capacità predittiva delle stime.

## Aspetti tecnici del GWR

Il **metodo GWR** prevede che:

- ▶ si stimino tante regressioni quante sono le unità nel campione;
- ▶ in ogni regressione si prenda a riferimento un'abitazione;
- ▶ nella stima di ogni regressione si ponderino le unità del campione in base alla loro distanza dall'abitazione di riferimento: maggiore è la distanza, minore è il peso dell'unità;
- ▶ un parametro (che può essere fissato o stimato) che regola la ponderazione delle unità determini il grado di variabilità geografica complessiva dei coefficienti  $\beta_{h,i}$ .

## Potenzialità del GWR

Sulla base dei risultati è possibile:

- ▶ **interpolare i prezzi**  $\beta_{h,i}$  delle caratteristiche delle abitazioni per stimare i prezzi di abitazioni non incluse nel campione;
- ▶ **aggregare i prezzi**  $\beta_h$  per aree omogenee di varia estensione (comuni, comuni catastali, particelle edificabili);
- ▶ **disegnare mappe** della variazione spaziale dei prezzi  $\beta_h$ .

## Bibliografia

- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984) *Classification and Regression Trees*, Wadsworth International Group, Belmont, California.
- Brunsdon C., Fotheringham A.S., Charlton M. (1996) Geographically weighted regression: a method for exploring spatial non-stationarity, *Geographical Analysis*, 28, 281–298.
- Huang, B., Wu B., Barry M. (2010) Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices, *International Journal of Geographical Information Science*, 24.3, 383– 401.
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, 74(2), 132-157.
- Stanca L. (2008) Le determinanti dei prezzi delle abitazioni: aspetti microeconomici, Working Paper Series, Department of economics, University of Milan – Bicocca, N. 143.