Recent trends in the real estate market and its analysis
Narodowy Bank Polski & Warsaw School of Economics (SGH)

21st November, 2017

Assessing the accuracy of Automated Valuation Models (AVMs)

Professor George Matysiak
Krakow University of Economics

Agenda

- Residential Automated Valuation models - AVMs
- Examples of AVM valuation (appraisal) accuracy
- Challenges
- Ongoing work
  - analysing prices and AVM valuations
  - analysing errors

Importance of accurate residential property valuations – some examples

- Property purchase is normally an individual's biggest lifetime purchase and an error, say +/- 5%, in the valuation is a significant component.
- Mortgage providers assessing Loan to Value (LTV) ratios, i.e. the value of the property is collateral on a mortgage.
  - residential value as a proportion to income/earnings ratio
- Property values and attendant taxes.
- Borrowing (re-mortgaging)
  - contingent 'wealth effect'– macroeconomic impact resulting from spending.
- Loan impacts on banks' balance sheets
  - banks want to mitigate risk by not issuing mortgages in excess of the 'underlying value' of a house
  - accurate valuation of existing loan portfolios
- Regulators and Central Banks do not want a repeat of the impact of valuation inaccuracies in the boom years preceding a collapse in market values (implementation of macro-prudential policies/regulation so as to avoid financial stress in the economy).
- Accuracy of valuation based residential property indices.
- Mis-valuation implications for investment portfolio asset holdings/asset allocation.

AVMs

- AVMs have their origins in North America, the first commercial application being in 1981, and began to be developed in the UK in the 1990s.
- Despite traditional approaches being extensively employed in the valuation profession, there has been a significant growth in independent residential Automated Valuation Model (AVM) providers, who offer their services routinely on a fee-based basis, to both lenders and the fee-paying public.
- AVMs are widely used by lenders and institutional investors, largely for monitoring purposes, and are seen as complementary to traditional valuations. The widespread use of AVMs is now firmly established.
- These computer-assisted quantitative methods have some advantage in that they are systematic and fast, thereby reducing reliance on labour input in providing an end-to-end valuation.
- By removing the human element, it is claimed by some advocates, it also reduces inaccuracies due to reliance on human judgement. This is an unsubstantiated assertion.
- The overall attitude and degree of acceptance of such automated approaches to valuation varies.

An AVM prediction ?

What is an AVM 1?

- Although different underlying AVM models are employed by vendors, fundamental to the approach are *statistical, data mining and computing technicalities.*

- TEGoVA provide the following, Definition 2.1, in their European Valuation Standards EVIP 6:

'Automated Valuation Models (AVMs) can be defined as statistic-based computer programmes, which use property information (e.g. comparable sales and property characteristics etc.) to generate property-related values or suggested values.'

- The International Association of Assessing Officers, IAAO (2003), describes an AVM as

'a mathematically based computer software programme that produces an estimate of market value based on analysis of location, market conditions, and real estate characteristics from information collected. The distinguishing feature of an AVM is that it produces a market valuation through mathematical modelling. The credibility of an AVM is dependent on the data used and the skills of the modeller producing the AVM.'

What is an AVM 2?

- The following definition of an Automated Valuation Model is provided by the RICS AVM Standards Working Group:

  'Automated Valuation Models use one or more mathematical techniques to provide an estimate of value of a specified property at a specified date, accompanied by a measure of confidence in the accuracy of the result, without human intervention post-initiation.' (RICS 2013).

- A key component in the RICS definition is the qualification *'...accompanied by a measure of confidence in the accuracy of the result...'*.

→ All three definitions of an AVM exclude any appraiser involvement in arriving at a value.

Practitioner attitudes towards AVMs

- An international survey undertaken in 2008 on AVMs and the integration of AVMs within the valuation process provides some interesting findings. There were 473 valuer responses, representing both lending and valuation organisations, and described as senior professional members with ‘much experience of mortgage valuations’. The results of the survey include the following :
- 71% of the valuers agreed that AVMs were inadequate for loan valuations as a result of no physical inspection.
- 87% of the valuers agreed that physical valuations were more accurate than AVMs, as a result of local knowledge.
- 90% of valuers agreed that the ability to evaluate comparables was a major advantage over AVMs.

Information on AVMs

- Debate regarding the role and accuracy of AVM valuations is an ongoing topic of discussion. For there to be a meaningful debate, AVM vendors need to make available access to their models for independent testing and verification of the models' output and accuracy of the results.
- Whilst there are a large number of AVM vendors, the inner functioning of the models and details of their specification are not released, nor are 'accuracy' figures usually disclosed.
- Vendors claim they test their models regularly for accuracy, and some may have the figures independently assessed. However, this non-disclosure puts a constraint on the analysis which can be externally undertaken as regards the assessment of the reliability and accuracy of the models.
- Other than submitting information to rating agencies, European/UK AVM operators are unwilling to have their data/methodologies exposed to independent scrutiny. *Vendors argue that their accuracy figures need to be put into a wider perspective.*
- US AVM market is highly developed and 'accuracy' figures are available on websites.

Measuring AVMs accuracy 1

- A variety of ways to measure AVM valuation error. First, a *benchmark reference* needs to be established:
  - is the AVM forecast measured relative to a valuer's estimate of the property's price or,
  - is the AVM forecast measured relative to the market price achieved in the market?

- Fitch, for example, evaluate model accuracy based on surveyors' assessment of values.

- UK vendors do not provide any details.

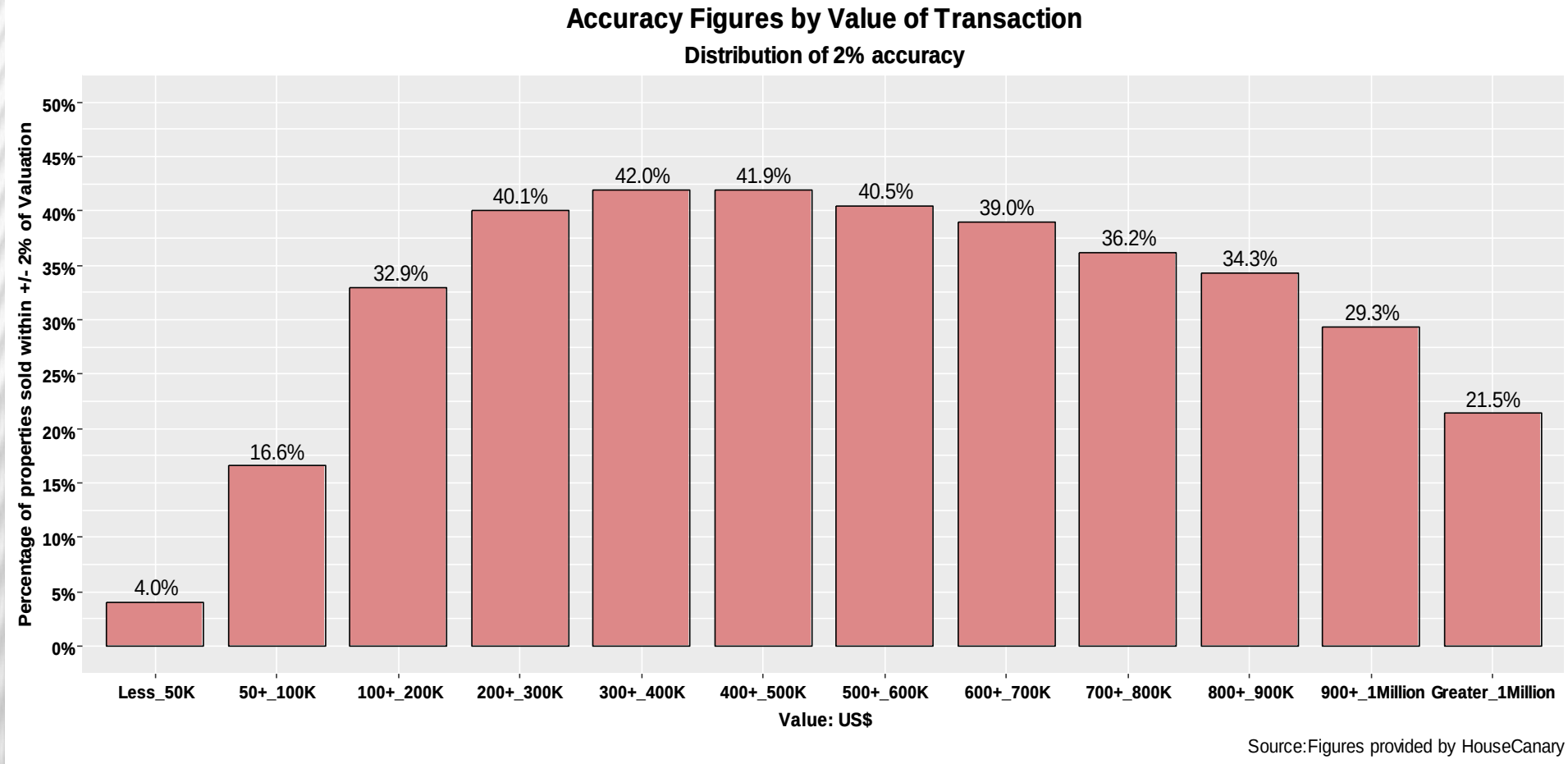
Measuring AVMs accuracy 2

- Several US vendors publish accuracy results based on achieved sales prices.

- How can accuracy be evaluated ? Several ways, for example:
  - can look at the percentage of AVMs falling within a specified range of error, for example within:
  +/- 5%, +/-10%, or +/-20% of the sales prices
  - the 'spread' around the 'average' of all errors (FSD)
  - the median (50% of values less/greater than the median)
  - average of the absolute errors (MAD/MAPE)

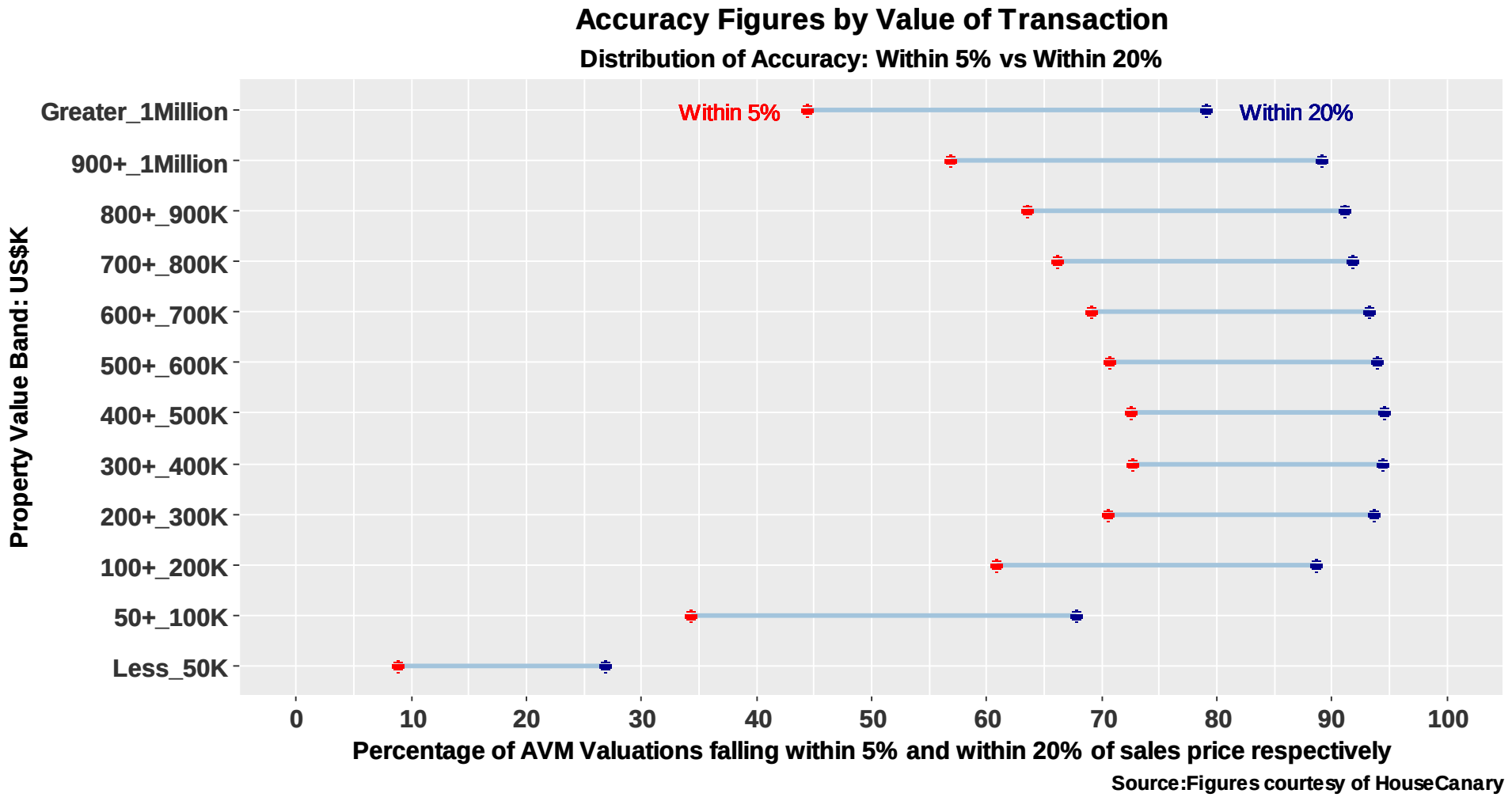
- AVM vendors typically qualify their valuation estimates by providing a *prediction range* with a specified degree of confidence.

Distribution of valuation accuracy within +/-2% of sales price - *HouseCanary*



Source:Figures provided by HouseCanary

Distribution of valuation accuracy - *HouseCanary*



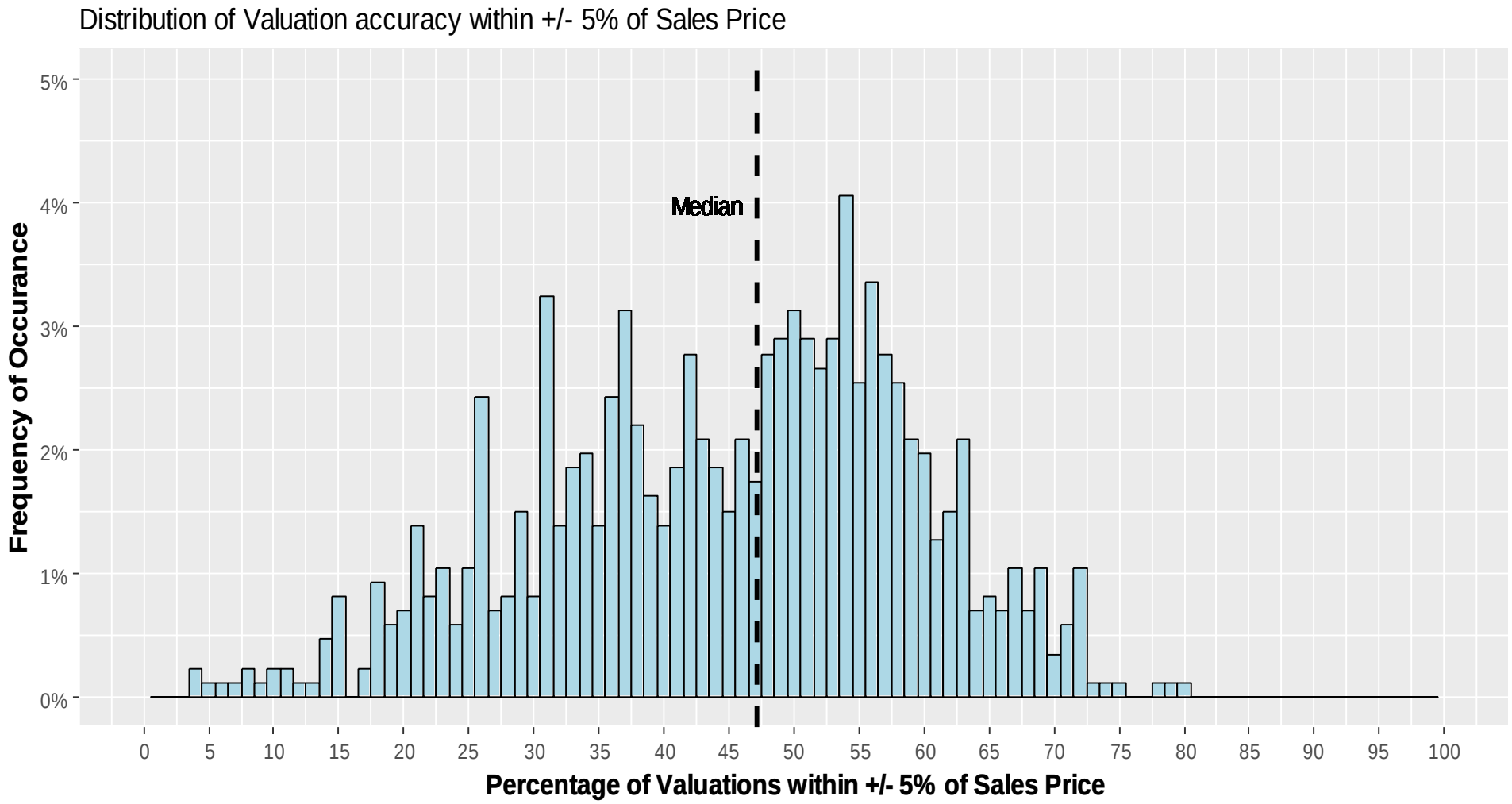
Features of *HouseCanary's* accuracy numbers

- The figures show the percentage of properties valued within +/- 5% of the sales price and within +/- 20% of the sales price, across 12 price bands.
- What stands out for the lower valued properties, less than US$ 50K, less than 10% of the valuations were within +/- 5% of the achieved sales prices, rising to 27% within +/- 20% of the sales price.
- For property values in the range US$ 100K – US$ 1 Million, the figure shows that in excess of 85% of values are within +/- 20% of the sales price, with 80% of properties valued in excess of US$ 1 Million falling within the +/- 20% band.
- If 10% is seen as an acceptable margin for error, for properties with values in excess of US$ 100k, some 75% of the valuations fall within +/- 10% of the sales price, and hence, one-quarter of AVM generated appraisals will have errors greater than +/- 10%.
- The *median absolute percentage error* across all properties (674 metropolitan statistical areas, MSAs, as at June 2017) is 5.6%, i.e. half of the errors nationwide were within 5.6% of the final selling price, and half exceed 5.6%.

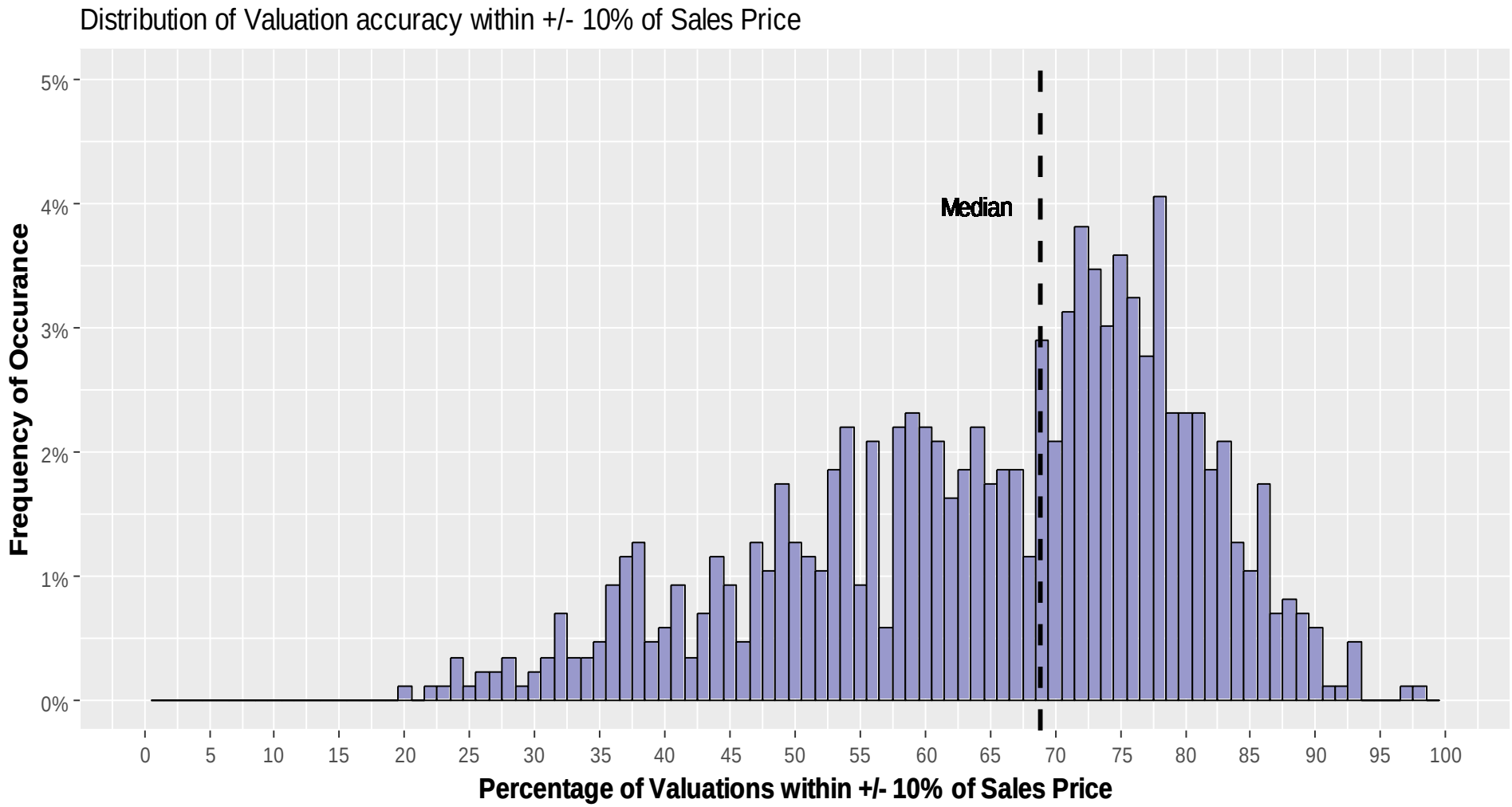
Accuracy numbers – *Zillow 1*

- Zillow claim to be the largest AVM provider in the US.
- The following three histograms show the distribution of Zillow's AVM accuracy rates for 864 US Counties.
- Different areas will have different valuation errors – aggregating data will miss this.
- The histograms show the distribution of AVM valuation accuracy *within each of the individual 864 Counties.*

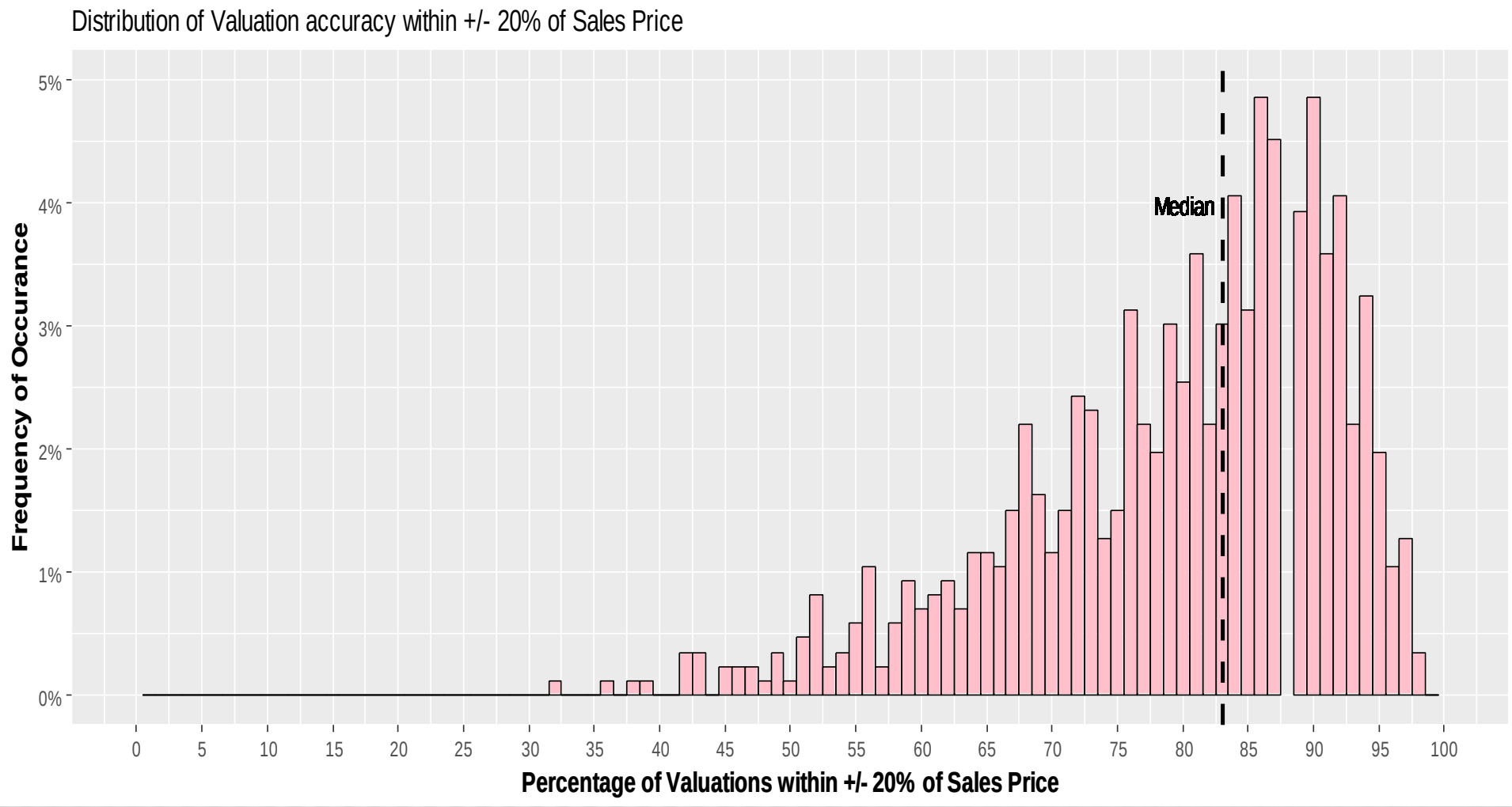
F1: Zillow's distribution of Valuation accuracy within +/-5% of sales price



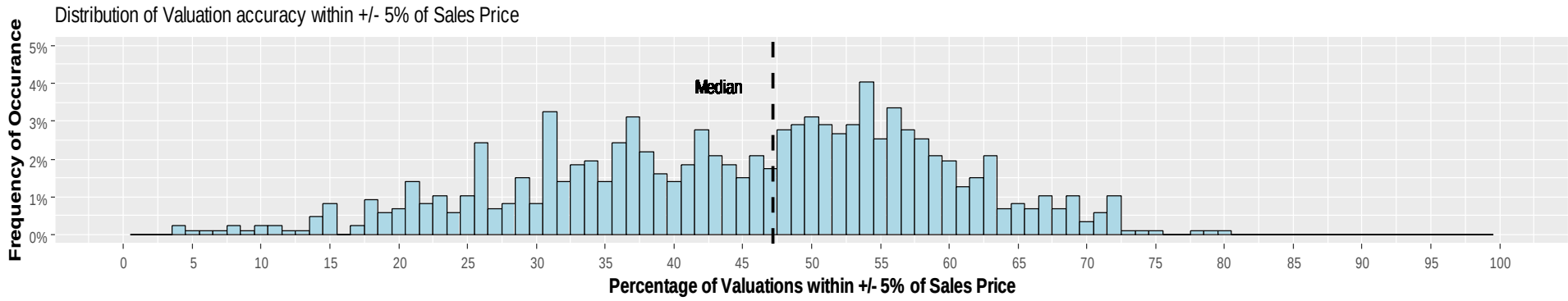
F2: Zillow's distribution of Valuation accuracy within +/-10% of sales price

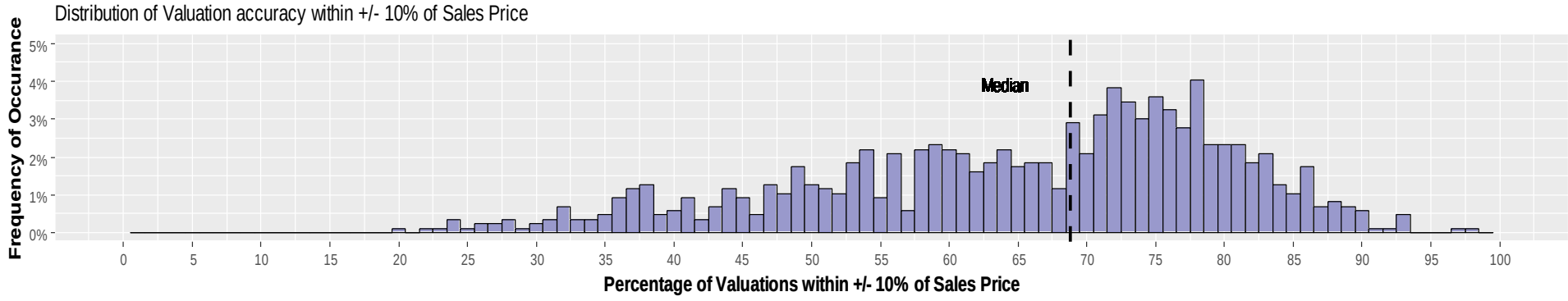


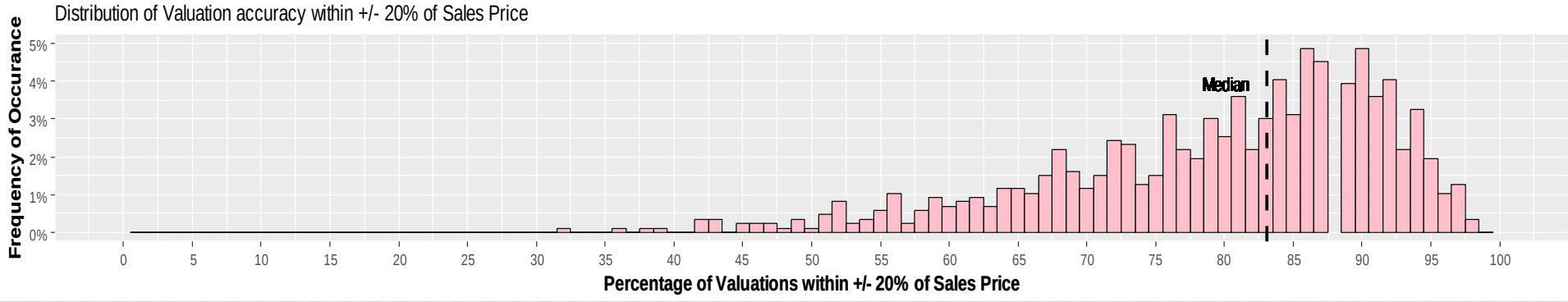
F3: Zillow's distribution of Valuation accuracy within +/-20% of sales price



Comparison of distributions







Accuracy numbers - *Zillow 2*

- The median level of *valuation error across the US* is 4.3% (as at June 2017), i.e. half of the errors nationwide were within 4.3% of the final selling price, and half had an error exceeding 4.3%.
- Figure F1: shows +/- 5% valuation accuracy rates; Figure F2: +/- 10% valuation accuracy rates and Figure F3: +/- 20% valuation accuracy rates. The Histograms provide a detailed visual insight across 864 US Counties.
- The average accuracy across all counties is superimposed in order to provide a reference point. On balance, it appears that some 50% of the valuations are **likely *to be outside*** the +/- 5% range of achieved sales price, which falls to almost 30% for the +/- 10% range and 17% for the +/- 20% range.
- *As shown in the histograms, given the skewed nature of the distributions, even at the wider range of +/- 20%, there exist a significant proportion of valuations in many locations which lie outside the specified ranges of accuracy.*

Features of *Zillow's* accuracy figures

- At the *individual County level,* the median ranged from some 2% to 32%, which represents a wide range of variation in accuracy across the different locations.
- Almost half (47%) of all valuations across all Counties were within +/- 5% of the sales price, half being on excess of +/- 5%.
  - One quarter of the Counties (215) had less than 35% of the valuations falling with +/- 5% of the sales price.
  - However, 10% of the Counties had less than 25% valuations within the 5% bracket. The lowest recorder accuracy was 3.5% and the highest recorded 80%.
- The average percentage of valuations across all Counties falling within +/- 10 % of the sales price is almost 70%. However, this can vary between 20% and 98%, depending on the County.
- On average, the percentage of valuations across all Counties falling within +/- 20% of the sales price is 83%. However, this can vary between 32% and 100%, depending on the County.

Margin for error

- What is an acceptable margin for the error level?

  - a long established concept developed in (UK) case law.

- Very little evidence for residential property markets.
- Market conditions will likely raise or lower the margin.
- Evidence from commercial real estate markets may provide an indication of an upper limit on inaccuracy.
- Examples:

  valuations compared against achieved market prices.

  some country comparisons: absolute average errors and the range.

Country	Average Absolute Error (%)	Range
Netherlands	9.1	6.5-13.0
Italy	10.3	6.3-13.8
Germany	11.7	9.1-15.6
UK	10.7	8.5-13.3
France	11.0	7.3-14.5
Sweden	12.5	8.7-19.0
Japan	12.0	7.4-17.2
Canada	11.2	7.8-14.1
USA	9.9	6.7-14.7

Netherlands, Germany, UK, France, Sweden, Canada, USA: 2000-2016
Japan: 2005-2016
Italy: 2006-2016

Source MSCI, 2017

Valuation under different conditions

- Is the margin for error unchanging i.e. constant? Will most likely vary under different conditions.
- There is a whole series of circumstances which would need to be taken into consideration when looking to assess what would be an acceptable margin for error in valuing residential properties, including the following:
  - different market environments, such as rising/falling prices
  - up market *versus* down market (asymmetric effect)
  - different size/value properties
  - quality of property
  - age of property
  - market liquidity e.g. dependent on the volume of transactions
  - geographic location/different neighbourhoods
  - type of property
- All of the above are likely to vary *by* country and *within* each country!
- Are the results likely to be sample specific?

Data considerations

- Current short comings of AVMs in the US are much more information based rather than methodology based.

  - current condition of most properties is largely unknown from the typical property level datasets being used today (especially for properties which have not been listed or sold for a long time).

  - secondary data deficiencies include home amenities that are important to consumers, but are not readily measurable in current property level datasets:

  Examples:

  1: does the home have a desirable open floor plan, 2: is there good natural light coming into the home, 3: is there good usable outdoor space on the lot.

- Above type of data is available today, but only for a very small portion of the entire housing stock, but prohibitively expensive.
- As the data becomes cheaper and more broadly available, the information it can provide will make it into valuation models and one can expect error rates will continue to decline.

Acknowledgement: Insight and observations credited to discussions with HouseCanary.

Data considerations: summary

- Improvement in AVM models will come with increased access to above types of data.
- Data quality improvements at the micro-data level required. The technologies are there but the costs currently inhibit across the board adoption.
- General assumption is that only marginal gains will result from new algorithms used within the final prediction model.

  - *a caveat*

    developing new algorithms used for non-traditional data feature extraction and/or faster processing times, will remain very important.

Acknowledgement: Insight and observations credited to discussions with HouseCanary.

Challenges 1

- The independent validation and standards of validation of European AVMs needs to be promoted more vigorously, otherwise the role of AVMs will continue to be contested. How best to proceed?

- In the absence of regulatory/enforceable controls, there has to be a commercial or reputational benefit to the vendor in order to make it worthwhile for them to provide information.

- Independent professional bodies qualified to scrutinise AVMs
  - setting a standard of best practice for AVM vendors
  - access to the underlying models
  - access to database(s) on which the models are calibrated/estimated/tested
  - access to AVM output under different market environments and any adjustment made to pure model generated forecasts
  - AVM accuracy/standardisation of published accuracy measures
  - clear definitions/standards for transparent ‘testing’ of AVMs procedures

Challenges 2

- Certification, implying positive publicity.
- Alternatively, more voluntary information from AVM vendors, but issues of impartiality arise.
- Collating users' AVM experience.
- Establishing more comprehensive micro-oriented databases.

Observations and questions

- The distribution of the accuracy figures of the US models, *across* both locations and *within* locations, appears to provide tolerable results within what could be considered acceptable levels of statistical confidence.
- However, a purely statistically derived or data-mined valuation (an AVM) risks being widely off the mark, lying well outside the +/- percentage ranges, as shown in the Zillow charts for example.
- Despite the high degree of accuracy reported by some US AVM vendors, there still remains a requirement for professional judgement to augment model-based valuations in arriving at a more broadly considered opinion of value.
- Requirement for more discussion about what is a fitting framework for assessing and evaluating AVMs. Effective independent validation of AVMs is hampered by the lack of industry standardisation across virtually all aspects of the AVM process, including access to underlying data, models and accuracy results.
- Vendors' concerns about commercial sensitivity/intellectual property of their products prevail.
- Does the valuer possess additional information to that contained in the AVM?
- Are AVMs more accurate than physical valuations?

Final thoughts

- AVMs are extensively used around the world and have become part of the valuation environment. Given the widespread deployment of AVMs, their use is not an either-or question, but a question of *how* can an AVM enhance a valuer's estimate of value.
- The position has been well summarised by a significant provider of valuation technologies to the mortgage banking industry as follows:

"AVMs are going to get more and more mainstream, particularly as data and analytics get more sophisticated. AVMs won't take the place of an appraisal. There will always be a need for local knowledge and expertise, not to mention an on-site evaluation of the physical property."

- Hot off the press. TEGoVA has issued a new European valuation standard and guidance note on AVMs:

European Valuation Standard 6

AVMs cannot be used to produce a valuation report that complies with EVS independently of a valuation process founded, inter alia, on inspection of the property by the valuer and the application of valuation judgment by the valuer.

Where used, an AVM is never more than a tool contributing to the valuer's estimation of value, for which he remains responsible.

Would it be more fun to sit in the pub with this…?

Update W, the weights of the output layer

For a particular weight w_{JK} (from units Z_J to Y_k)

$$\frac{\partial E}{\partial w_{JK}} = \frac{\partial}{\partial w_{JK}}(0.5\sum_{k=1}^{m}(t_k - y_k)^2) = \frac{\partial}{\partial w_{JK}}(0.5(t_K - y_K)^2)$$

$$= (t_K - y_K)\frac{\partial}{\partial w_{JK}}(-y_K) = -(t_K - y_K)\frac{\partial}{\partial w_{JK}} f(y\_in_K)$$

$$= -(t_K - y_K) f'(y\_in_K)\frac{\partial}{\partial w_{JK}}(y\_in_K) \text{ (by chain rule)}$$

$$= -(t_K - y_K) f'(y\_in_K) z_J$$

The last equality comes from the fact that only one of the terms in $y\_in_K = \sum_{j=1}^{p} w_{jK} z_j$, namely $w_{JK} z_J$ involves w_{JK}

Let $\delta_K = (t_K - y_K) f'(y\_in_K)$. Then $\Delta w_{JK} = \alpha \cdot (-\frac{\partial E}{\partial w_{JK}}) = \alpha \cdot \delta_K \cdot z_J$

This is the update rule in Step 6 of the algorithm

…or this

Determining Class Assignment Rules (cont'd)

$$
\begin{aligned}
p(j^*|t) &= p(j^*, t_L|t) + p(j^*, t_R|t) \\
&= p(j^*|t_L)p(t_L|t) + p(j^*|t_R)p(t_R|t) \\
&= p_L p(j^*|t_L) + p_R p(j^*|t_R) \\
&\leq p_L \max_j p(j|t_L) + p_R \max_j p(j|t_R)
\end{aligned}
$$

$$
\begin{aligned}
r(|t) &= 1 - p(j^*|t) \\
&\geq 1 - \left[p_L \max_j p(j|t_L) + p_R \max_j p(j|t_R) \right] \\
&= p_L(1 - \max_j p(j|t_L)) + p_R(1 - \max_j p(j|t_R)) \\
&= p_L r(t_L) + p_R r(t_R)
\end{aligned}
$$

$$
\begin{aligned}
R(t) &= p(t)r(t) \\
&\geq p(t)p_L r(t_L) + p(t)p_R r(t_R) \\
&= p(t_L)r(t_L) + p(t_R)r(t_R) \\
&= R(t_L) + R(t_R)
\end{aligned}
$$

…or this valuer?

Ongoing work

Analysing residential prices & AVM valuations

Assertion

- Interested in testing the proposition that AVM valuations are unbiased estimates of actual transactions prices:

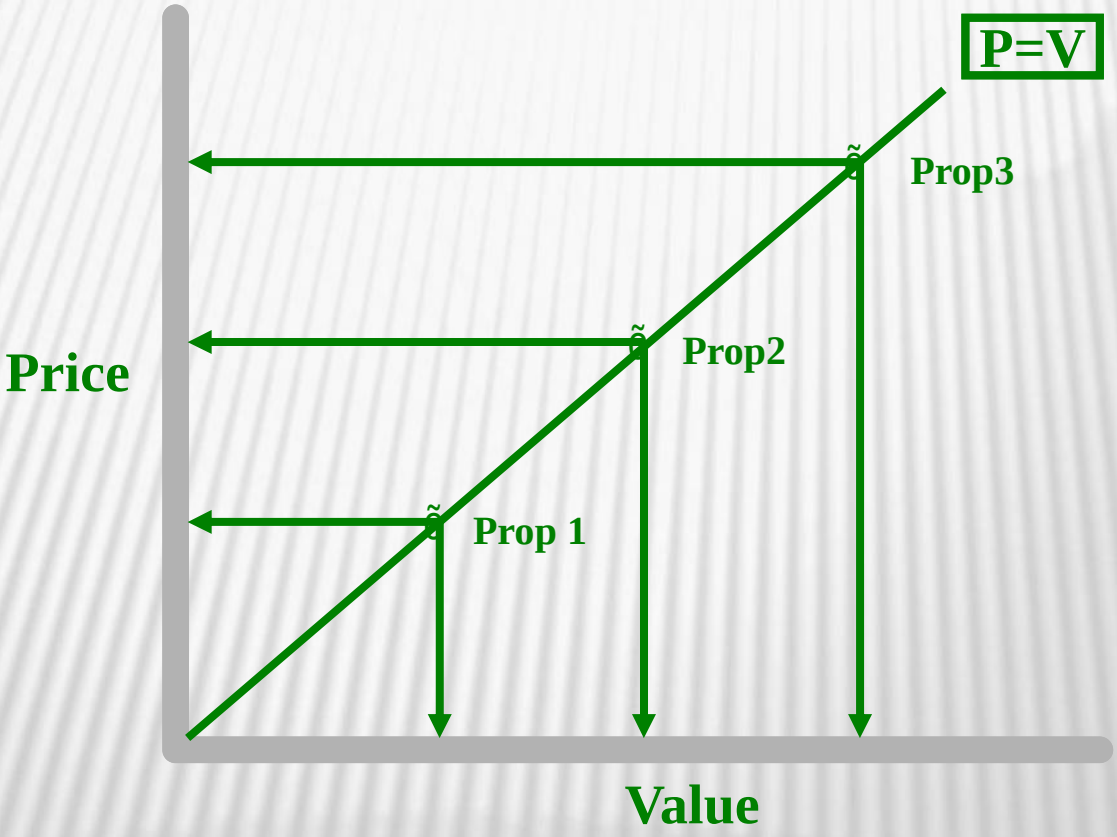
  Value=Expected Price=Actual Price

  $$V=E(P)=P$$

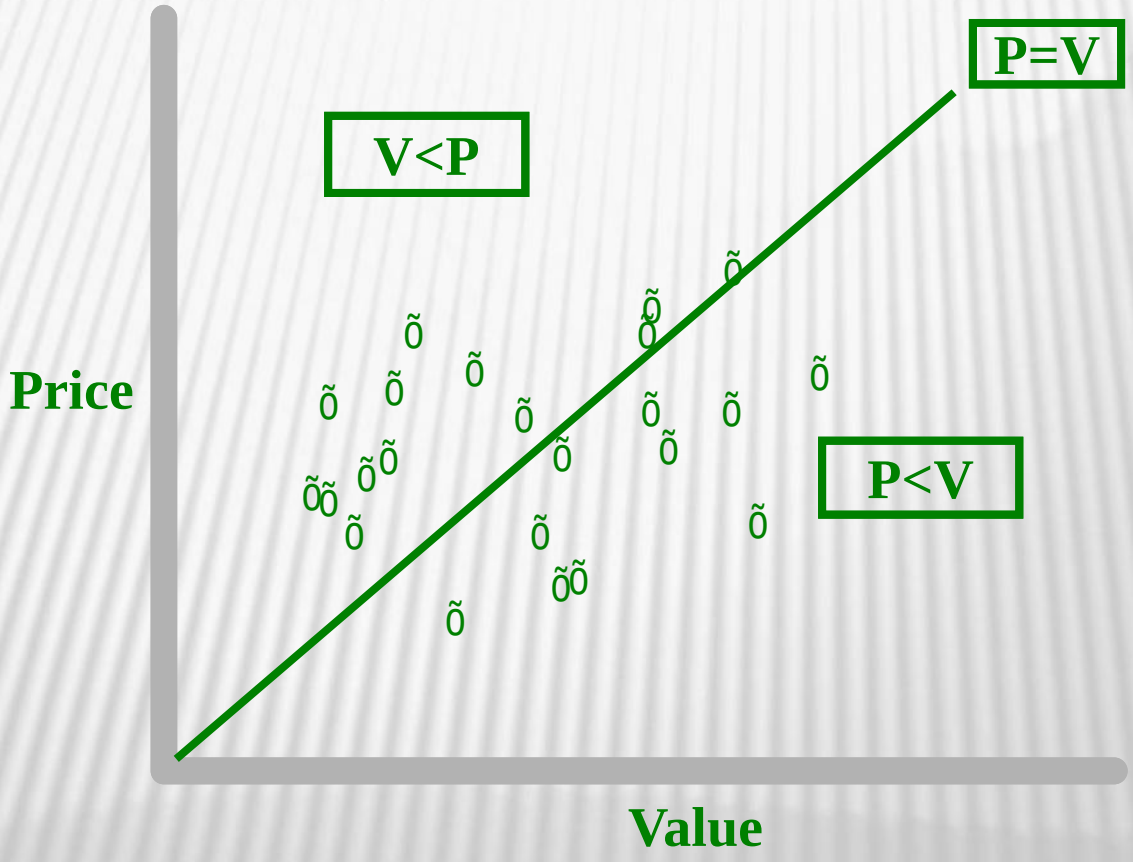
  simply:

  $$Ln\text{Price}_{it} = \ _0 + \ _1 Ln\text{Value}_{it} + £_t$$

Price-Value Space



Price-Value Space
Random Distribution



Conjecture

- AVMs are slow to respond to/cannot adequately capture market conditions in arriving at values. That is, market conditions are not fully incorporated in situations of *relatively* rapid market movements

  $\Rightarrow$ AVMs under/over react under certain market conditions

- Are there systematic tendencies for AVM valuations to lag market transactions prices i.e. biased AVM valuations in certain regimes/environments?

Augmented regression equation formulation

- Earlier bivariate regression formulation needs to be augmented by the inclusion of ‘market regime’ variables.
- What are market regimes?

Commercial real estate findings

- Appears to be some evidence for valuer appraisal bias, but not in every market environment i.e.

$$V \neq E(P) \neq P$$

- Wide range of uncertainty regarding appraisal accuracy (distribution of errors).

Questions

- How are the regimes identified?
- Are any biases/errors likely to be the same in all situations?
  - location
  - type of property
  - size of transaction
  - up market *versus* down market (asymmetric effect)
- Will the results be sample specific?

Further work: valuation error distributions

- Underlying distributions
  - + - not normal!
- Implications for tail accuracy estimates (underestimation if normal distributions are assumed!)
- Fat tails, so possibly Extreme Value Theory (EVT) application
  - + - Generalised Pareto Distribution (GPD)

Recent trends in the real estate market and its analysis
Narodowy Bank Polski & Warsaw School of Economics (SGH)

21st November, 2017

Assessing the accuracy of Automated Valuation Models (AVMs)

Professor George Matysiak
Krakow University of Economics