



# UK Job Vacancy Counts

- A Comparison of ONS & Innovantage Figures -

**Nigel Marriott**  
Chartered Statistician  
February 2013

# Innovantage

# Innovantage

- Innovantage ([www.innovantage.co.uk](http://www.innovantage.co.uk)) collate on-line job adverts in the UK using their web extraction technology.
  - This is a continuous operation covering nearly 200 commercial job boards.
- They now have a database starting from 2009 consisting of roughly 1.5 million job adverts every month.
- The raw data collated is unstructured, unstandardised, often incomplete and consists of many incompatible classifications.
- My involvement with Innovantage has been twofold.
  - Turn the dirty raw data into clean statistically sound data.
  - Turn the clean robust data into actionable labour market intelligence.

# Innovantage – Labour Market Intelligence

## 1. When are job vacancies being advertised?

- Analysis of daily trends & seasonality is possible.

## 2. Where are the job vacancies located?

- Analysis from region down to postcode is possible.

## 3. What kind of jobs are being advertised?

- Statistical classification model based on job title is used to classify jobs by discipline (occupation).
- Customised classifications are available.

## 4. How are job vacancies being advertised?

- Analysis of which job boards are being used.
- Also whether vacancies are posted directly by the employer or a recruitment agency.

## 5. Who is advertising job vacancies?

- Advertiser names are matched to Dun & Bradstreet DUNS numbers allowing all kinds of business information to be attached to the analysis.

## 6. Why are job vacancies being advertised?

- Requires extraction of key words from job descriptions which is currently being developed.
- Can also correlate vacancies data with other economic variables and use vacancies trends as a leading predictor of economic trends.

# Innovantage – The Statistical Challenge

- Clients will only use the data if they believe it is a true reflection of the real labour market.
- The biggest challenge is to DE-DUPLICATE multiple adverts of the same job vacancy.
- Every job advert can be duplicated in the following ways:
  - Reposted on the same job board every week or month.
  - Reposted on other job boards owned by the same company who owns the original job board.
  - Advertised by multiple recruitment agencies in addition to a direct advert by the employer.
- Each successive repost of an original job posting needs to be identified so that we can count the number of JOB VACANCIES being advertised rather than the number of JOB ADVERTS (or POSTINGS).

# De-Duplication – Original Algorithm

- Innovantage & I started with a simple deduplication algorithm.
- If two or more job postings had the same **JOB TITLE** and **JOB DESCRIPTION**, then these postings would be flagged as coming from the same job vacancy and given the same VACANCY\_ID code.
  - Minor variations such as spacing or paragraphs mark differences were ignored.
- We knew that this algorithm was biased and likely to lead to an overestimate of the true number of job vacancies.
- A common situation that would fool this algorithm:
  - A Statisticians job in Bristol might have the job title “Statistician” on Totaljobs and location field “Bristol” but on Jobsite, the same job title might be “Statistician – Bristol”.
  - Adding a location (or salary) to the job title enhances google searches but confused the algorithm.

**ONS**

# Enter the ONS Job Vacancies Survey

- To measure the extent of any bias, we compared our results with the ONS Job Vacancy Survey.
- Every month, the ONS Labour Market Bulletin publishes in Table 21 an estimate of the number of job vacancies in the UK.
- The figure that is reported is a seasonally adjusted 3 month moving average.
  - Non-seasonally adjusted figures are available but you have to ask for these.
  - Response times to such requests have been good in my experience.
- Figures are broken down by size of employer and SIC (standard industry classifications) codes.
  - Figures are not broken by location or by any other category.
  - This is where Innovantage can provide additional depth of detail.

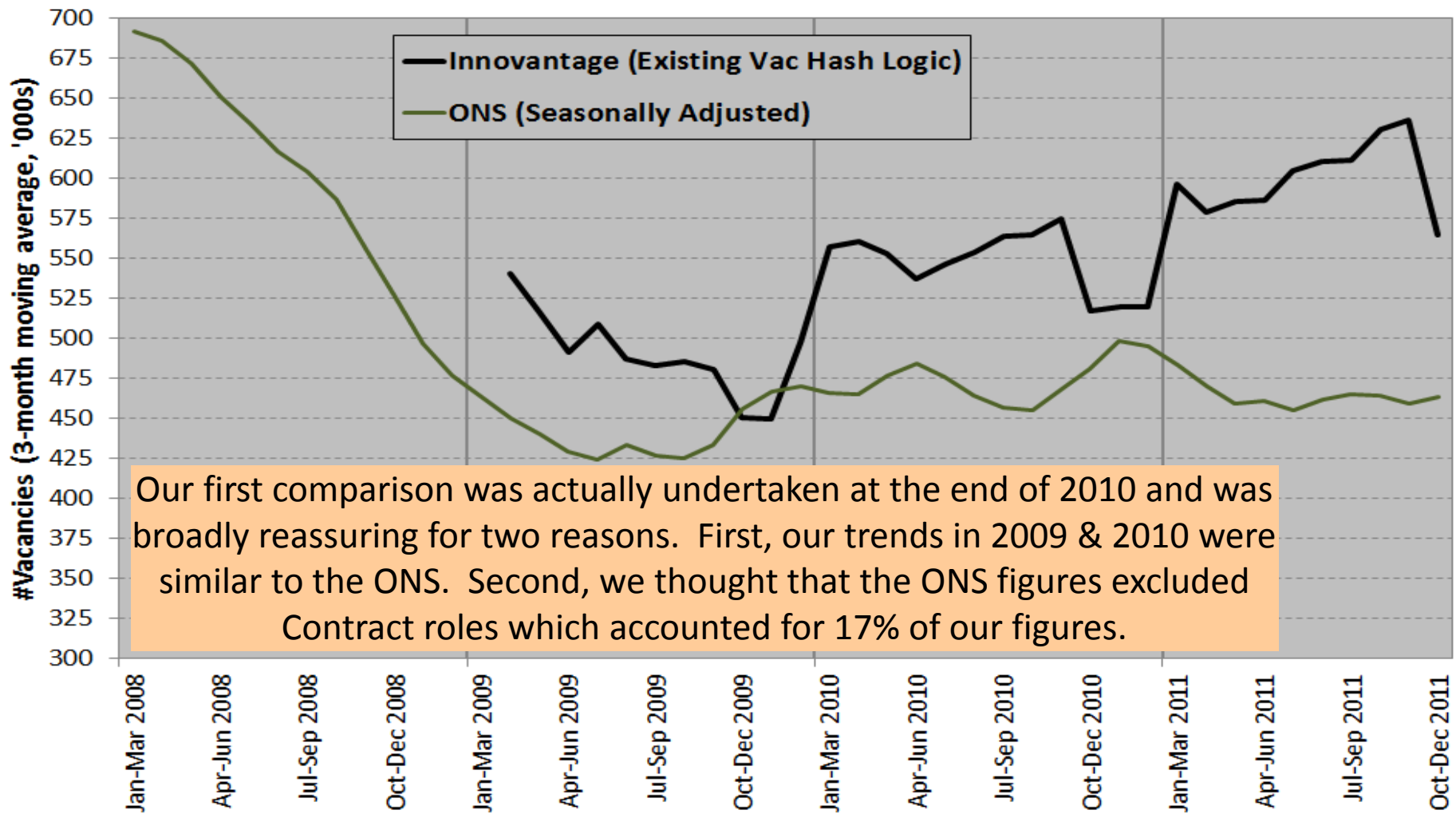


# ONS Job Vacancies Survey - Methodology

- Data is collected from a panel survey of 6000 enterprises.
  - 1500 large companies are permanent members of the panel and are surveyed every month.
  - Smaller companies are selected at random and remain in the panel for 5 to 9 quarters.
  - Each smaller company is surveyed quarterly with 4500 smaller companies surveyed every month.
  - Survey tends to focus on head offices (where applicable) which is why it is difficult to obtain a regional breakdown.
- Each company is asked a single question:
  - *“How many job vacancies did your business or organisation have on [ DATE] for which you were actively seeking recruits from outside your business or organisation?”*
- The standard error is estimated to be +/-1.5%.

# Innovantage Vs ONS

# Innov & ONS – 1<sup>st</sup> Comparison



In 2011 though our trends started to diverge from the ONS and so we undertook a detailed investigation to identify the extent to which the two series were comparable.

# Are ONS & Innov Figures Comparable?

|                    | Office of National Statistics  | Innovantage  |
|--------------------|--|--|
| Data Collection    | Monthly survey of 6,000 businesses for publicly advertised vacancies   | Hourly crawl of job postings on 177 commercial job boards. Corporate websites are also crawled but results are not included in vacancy count. Unique jobs identified using a vacancy hash logic based on the raw job title and job description only. |
| Job Types Included | Permanent, temporary and most contractor jobs included. Voluntary and pure consulting roles excluded. Internal job vacancies excluded. | All job types are collected. By definition internal job vacancies will not be picked up.   |
| Region             | Survey is just for Britain but results are adjusted to create a UK total.  | Figures are for Britain only. The number of job postings collected from Northern Ireland is very low in any case.  |
| Hours Worked       | A vacancy is counted as a vacancy regardless of whether the role is full or part time  | A vacancy is counted as a vacancy regardless of whether the role is full or part time.   |

The answer is **YES**. The key comparability is that both series focus on “publicly advertised” vacancies and exclude purely internal vacancies. We also discovered that the recruitment industry itself did not understand the distinction between Contract & Temporary jobs and Contract roles needed to be included.

The major difference is that Innovantage only collect vacancies advertised on job boards whereas ONS include those vacancies not so advertised.

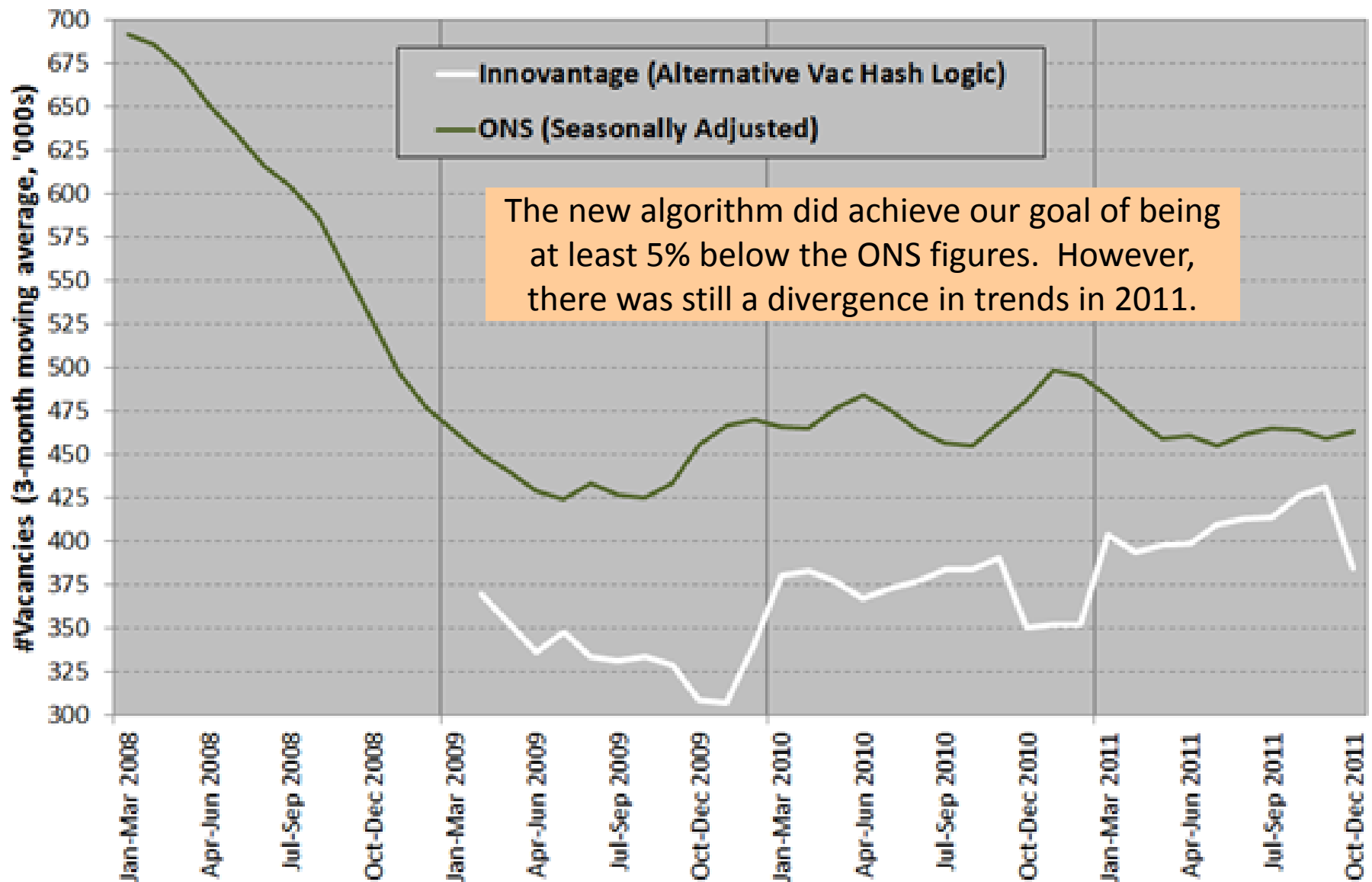
# So Why Were Innov Figures Too High?

- Our existing de-duplication algorithm was known to be crude and probably biased for the reasons given earlier in this presentation.
- Recruitment agencies account for 85% of all adverts & have been known to post non-job adverts for the purposes of collecting CVs.
  - There is a commercial incentive for this behaviour. Agencies often pay a lower fee per advert to the job board in return for posting an agreed volume of jobs.
  - I have developed an algorithm to spot some of these non-jobs but as yet this has not been implemented.
- We came to the conclusion that to be credible, our figures had to be at least 5% lower than the ONS figures for two reasons.
  - Innov do not include Northern Ireland (3% of the population), ONS do.
  - Purely offline adverts e.g. newspaper adverts are not included. However, nearly every newspaper is owned or affiliated to a larger publisher who repeat job adverts on a job board. In turn, job boards themselves are often owned by a publisher e.g. Reed. Therefore we believe this market is small.

# New Deduplication Algorithm

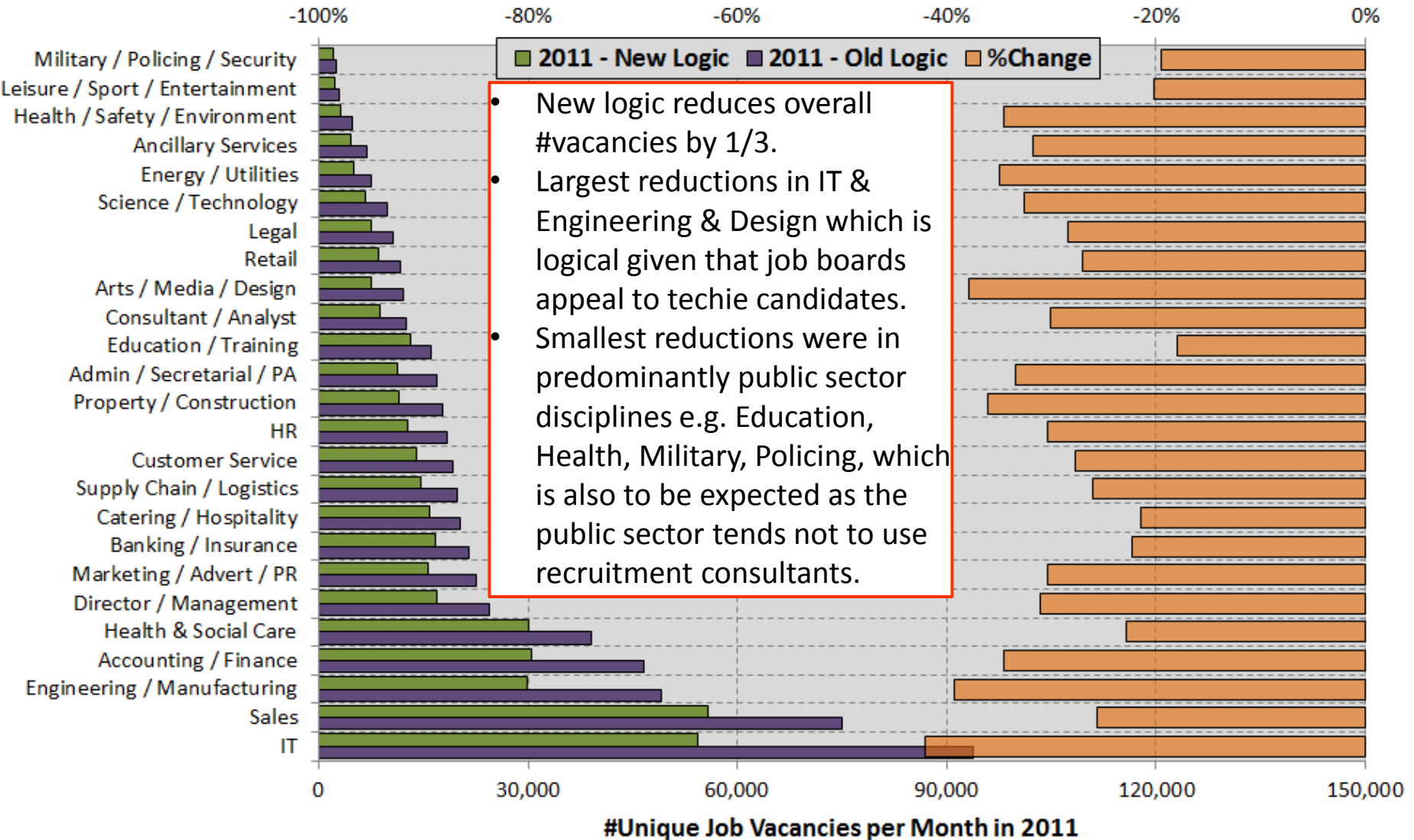
- By 2011, the algorithms I had developed for Innovantage to standardise & classify their dirty raw data were proving their worth.
  - For example, adverts by the Royal Bank of Scotland appeared under many company name variations e.g. “RBS”, “RBS Plc”, “The Royal Bank of Scotland”.
  - Regular audits showed that the misclassification error rate for some standardised fields such as location & salary were less than 1%.
- Therefore, I was able to develop an enhanced de-deduplication algorithm which made use of these additional clean & standardised fields to identify whether two or more postings were in fact for the same job.
- My plan was to achieve comparability with the ONS figures first and then test & validate the algorithm by selecting a random sample of vacancy ID groups and following up on the contact details in those adverts to manually verify if the adverts were for the same job.

# Innov & ONS – 2<sup>nd</sup> Comparison



# Innovantage – Comparison of Old & New

%Change in #Unique Job Vacancies due to change in logic



- New logic reduces overall #vacancies by 1/3.
- Largest reductions in IT & Engineering & Design which is logical given that job boards appeal to techie candidates.
- Smallest reductions were in predominantly public sector disciplines e.g. Education, Health, Military, Policing, which is also to be expected as the public sector tends not to use recruitment consultants.



# Why are trends diverging in 2011?

- 2011 saw a double dip recession, ONS was essentially flat but Innovantage saw growth in vacancies. Two possible reasons are:
  - A collapse of the offline recruitment market with everyone finally moving into online recruitment. Plausible but would expect this to be a gradual effect.
  - An increase in non-job adverts by agencies. More thought is needed as to what would be the incentive for this but we did see only 3% growth in 2011 for vacancies by direct advertisers (i.e. employers of the advertised roles) compared to 9% growth in vacancies from agencies & indirect advertisers.
- More importantly, this divergence forced us to look at the quality of our raw data in more detail.
  - Combined with another strand of work, we found a major job board (with 10%-15% market share) had subtly changed its text coding in mid 2011.
  - The effect of this was that two job titles from different job boards that appeared to be the same were not recognised to be the same words.
  - Thus the 2 postings would be treated as 2 vacancies instead of 1 vacancy.
  - This would artificially increase the 2011 figures of Innovantage.

# Conclusions

# Establishment Stats Enhance Street Cred

- Innovantage needed to establish its credibility as a provider of accurate big data which can be trusted by its clients when making commercial decisions.
  - The establishment statistics provided by the ONS was the best way by far of achieving this.
  - ONS are perceived as independent and reliable by the players in the labour market.
  - It is also fair to say that ONS are not perceived as a provider of timely & commercially relevant statistics.
  - Innovantage can bridge this gap (with myself providing the statistical expertise) and provide additional depth to what is provided by ONS.
- Usage of ONS Job Vacancy Stats enabled Innovantage to:
  - Identify & (more importantly) Test algorithms that can improve the way multiple job postings are de-duplicated.
  - Postulate hypotheses as to why the ONS & Innovantage figures differed which in turn led to market insights and improved data quality.