Nesta...

WORKING PAPER SERIES

Nesta Working Paper No. 14/04

Estimating the Impact of Live Simulcast on Theatre Attendance: An Application to London's National Theatre

Hasan Bakhshi
Andrew Whitby

Estimating the Impact of Live Simulcast on Theatre Attendance: An Application to London's National Theatre

Hasan Bakhshi
Nesta

Andrew Whitby
Nesta

Nesta Working Paper 14/04
June 2014

www.nesta.org.uk/wp14-04

Abstract

Increasingly, performing arts venues are adopting live simulcast into cinemas as a means of increasing their overall audience reach. The effect on audience numbers at performing arts venues themselves is unclear, however: simulcasting may substitute for live attendance among existing audiences, but may also promote and engage new audiences. Using data for the UK's early National Theatre (NT) Live broadcasts, Bakhshi and Throsby (2014) conclude that live broadcasts generated greater, not fewer, audiences at the National Theatre. Using a new, extensive dataset of theatre ticket transactions for multiple theatre venues across England, and over a longer time period, we conclude that National Theatre Live is likely to have in addition boosted local theatre attendance in neighbourhoods most exposed to the programme.

JEL Classification: Z11; O33

Keywords: Digital technology; Theatre; Cannibalisation

The authors thank: The Audience Agency whose Audience Finder national cultural data set (funded by Arts Council England) forms the basis of our analysis; the National Theatre, for providing us data on live screening venues, and John Davies at Nesta for his contributions to the research in its early stages.

 Corresponding author: Hasan Bakhshi, Nesta, 1 Plough Place, London, EC4A 1DE, hasan.bakhshi@nesta.org.uk.

1 Introduction

Since 2006, when the Metropolitan Opera in New York began broadcasting by satellite HD performances of its operas into digital cinemas, there has been an explosion of interest by performing arts companies in live simulcasting (Towse, 2013). In 2008, the Berlin Philharmonic launched its Digital Concert Hall online streaming service and in 2009, the Royal National Theatre in London began broadcasting live theatre to cinemas. Other cultural forms have followed suit. In 2013, the British Museum began live broadcasts of blockbuster exhibitions to cinemas. In a large-scale survey of how arts and cultural organisations in England are using digital technologies, live simulcasting was found to be the fastest growing technology (Nesta, 2013).

No doubt this popularity is explained by the fact that live simulcasts allow venue-based organisations to expand their 'virtual capacity'. For example, the National Theatre's inaugural live broadcast of Racine's play, *Phèdre*, on 25th June 2009, allowed it in one night to double the overall audience for the play over its three-month run at the National (Bakhshi and Throsby, 2012). On 27th February 2014, 120,000 experienced the performance of the National's *War Horse* production in UK cinemas, compared with the 1,024 seat capacity of the New London Theatre where it was staged (breaking the National Theatre Live box office record).

What is less obvious is the impact of such live broadcasts on the box office at theatre venues themselves. On the one hand, they might be expected to increase theatre attendances insofar as live broadcasts serve to promote live shows at theatres. On the other hand, they might be expected to 'cannibalise' theatre attendance insofar as the live broadcasts are substitutes for theatre performances in the eyes of the theatre-going public. Bakhshi and Throsby (2014) study the impact of the National's *Phèdre* broadcast on its own box office bookings and conclude that the net effect was to add to, not cannibalise, attendance at the theatre.

As the phenomenon of live simulcasts such as National Theatre Live has grown in strength, there has been some speculation as to whether the broadcasts have impacted, either positively or negatively, on theatre attendance more generally (Towse, 2013). Audience survey data collected by one of the authors of the current paper for *Phèdre*, and for a second National Theatre Live production, Shakespeare's *All's Well That Ends Well*, suggested that a significant minority of audiences were inspired by the live broadcasts to want to go to more theatre. But, to date, there has been no evidence on whether this has in fact turned out to be the case, and conceivably even if it has been for the earlier broadcasts it may not be for later ones, given that live broadcasts have become a more established part of the theatre oeuvre (Towse, 2013).

These questions arise not only in respect of theatre: a recent study using survey data concludes that a (non-random) sample of audiences of live broadcasts of opera in London were mostly unaffected in their motivation to see future live performances, but became substantially more motivated to attend future live broadcasts (Wise, 2014, section 4iv). While these results are interesting, we are reluctant to place too much weight on stated preferences, and prefer to evaluate revealed preferences from actual behaviour.

In this paper, we use a new, 'big' data set of ticket transactions for 54 performing arts venues across England and spanning a period from early 2009, when National Theatre Live was launched, through to late 2013, to investigate this issue. To anticipate the conclusions, we find that National Theatre Live appears to have boosted local theatre attendance in neighbourhoods most exposed to the live broadcasting programme.

2 Modelling approach

A first-best modelling approach using observational data would be to track individuals, both those who attended NT Live screenings and those who did not, and compare theatre attendance for each group before and after (e.g. a diff-in-diff analysis) while attempting to control for any remaining endogeneity. Unfortunately two data limitations preclude such an analysis here: (a) we do not, for this paper, have access to uniquely identifiable individual theatre attendee data, but instead have only customer postcode (b) we do not have data on which individuals attended National Theatre Live screenings.

However, since UK postcodes allow very precise geolocation, we can build a model using aggregate data for small geographic areas, an approach that has previously been taken in predictive modelling of cultural attendance in London (Brook et al., 2010), and is also similar to the approach used in Bakhshi and Throsby (2014). We use the small areas defined by the Office for National Statistics for the 2011 census (ONS, 2014), namely:

OA Output areas, comprising at least 40 households (125 recommended)

LSOA Lower layer super output areas, comprising 400 to 1200 households

MSOA Medium layer super output areas, comprising 2000 to 6000 households

For comparison, the postcode districts used in Bakhshi and Throsby (2014) are around twice the size of an MSOA, on average.

In this aggregate setting, NT Live treatment is no longer a binary attended/did-not-attend variable, but must reflect instead distance from NT Live screenings (which can be interpreted as a proxy for probability of attending). This leads to the first of two hypotheses that we consider:

Hypothesis 2.1 *The effect of NT Live should decline with distance from participating cinemas.*

Intuitively we would expect this to be true whether the effect of NT Live is positive or negative: in either case its magnitude should diminish with distance. Secondly:

Hypothesis 2.2 *The effect of NT Live should be stronger on theatre consumption than non-theatre consumption.*

Note the relative statement here: if NT Live is a complement (or substitute) for theatre, and theatre has varying degrees of substitutability for other artforms, then we should expect NT Live to have some effect even on these non-theatre artforms: but anything less than perfect substitutability would suggest the magnitude should be less than on theatre itself.

3 Data

Our primary dataset consists of around 16 million transaction records for 54 performing arts organisations around England, from The Audience Agency's[1], *Audience Finder* national cultural data-set. These records represent 44 million tickets and span a period from early 2009 through late 2013, which coincides with the introduction and expansion of NT Live across the UK. They represent a convenience sample of such organisations in England, but one that has a reasonable spread of organisations by size and geography (see Appendix A for more detail).

As essentially raw transaction data from ticket-booking systems, substantial cleaning was undertaken to create a useable dataset of geo-tagged theatre attendance data. For example, records with invalid postcodes were either corrected or dropped and large group bookings were removed.

[1]a nonprofit audience development agency

3.1 Coding

The most substantial data exercise involved coding each transaction record to a particular artform—that is, a performance genre. This procedure varied across organisations, with three main approaches:

- For organisations that offered solely theatrical performances, all records were coded as 'theatre'. Conversely, for organisations that offered solely non-theatrical performances, all records were coded as 'non-theatre'.
- For organisations that offered a mix of performances and provided an organisation-specific coding of performances, this was mapped to 'theatre' or 'non-theatre'
- For organisations that offered a mix of performances and did not provide any coding, transactions were manually coded based on performance title, or matched to events listing data scraped from the organisation's website[2]

After cleaning and coding, around 12 million records remain, representing 28 million tickets, of which 32 per cent are for theatre performances, with the remainder for other non-theatre art forms.[3] Around 2 million records are dropped at the cleaning stage (lack of postcode, etc.) and around 2 million records are dropped at the coding stage.

The performing arts organisations remaining in the dataset at this point are shown in Figures 1 and 2. Also shown is a 30 km radius around each venue: this is used later in aggregation (described below), and reflects a balance between a reasonable catchment for theatre attendance and the computational limitations of including very large geographic areas in the analysis.

3.2 Aggregation

We then proceed to aggregate the data along three dimensions: by customer geography, by organisation and over time. For the unit of time, we chose annual aggregation. This eliminates the need to model seasonality and reduces the size of the final dataset, but sacrifices some time resolution that the large dataset would permit.[4] For the geographic unit, we currently use MSOA (recall, around 2,000-6,000 households) as this provides a

[2] In some cases an organisation's public-facing website provided event classifications even though their internal transaction systems had not recorded this information. Using this source of information reduced the burden of manual classification.

[3] The clean dataset represents around 8 million tickets sold, or around 1.6 million per year of data. As a reference point, one estimate puts total UK theatre tickets sold in 2012/13 at around 34 million (Theatrical Management Association (TMA), 2013).

[4] We are testing the robustness of our results to different time aggregations.

Figure 1: Map showing all performing arts organisations in the sample. Venues close together are shown as a circle with a number, indicating the number of venues. Darker shading indicates the 30km radius catchment area.

sufficiently high mean visit count for our modelling approach.[5] For each organisation, we include all MSOAs within the 30 km radius, for all years in which that organisation provided data. This results in an unbalanced panel. For any geography $\times$ organisation pairs within the radius, for which we observe no transactions in a given year, we record zero visits.[6]

When aggregated in this form, we have a total of $N = 3619$ geographic areas, $M = 54$ organisations and $T = 5$ time periods. The total number of aggregated observations is

[5] However, we are in the process of testing our models with smaller areas, e.g. OA and LSOA: data availabilities at these levels motivates some of our decisions regarding covariates.

[6] This reflects the fact that our database is an essentially complete record of attendance at the sample of theatres, so in our ecological framework we treat this not as a case of truncation or missing data, but as a genuine zero observation

Figure 2: Map showing all performing arts organisations in the sample, in London. Venues close together are shown as a circle with a number, indicating the number of venues.

113,025 (for theatre) and 128,505 (for non-theatre), substantially less than $N \times M \times T$. Not every possible triplet occurs in each of the theatre and non-theatre datasets due to (a) some organisations offering only theatre or non-theatre; (b) each organisation's catchment being different and (c) missing years for a small number of organisations.

3.3 Other data

The National Theatre provided us with data on the locations (postcodes) and screening dates for cinemas participating in the NT Live program. This represented around 12,000 unique events at 482 unique cinemas, with the numbers increasing over the period 2009 to 2013. This was used to generate treatment variables.[7]

Finally we used third-party data for annual small-area covariates in our models. Little official data is produced on the smallest statistical areas on an annual basis: population is an exception and we include it as an obvious control. Economic measures are general not available at such high geographic and time resolution: instead, we use the recently-opened Land Registry price-paid data to compute a median house price in each year and relevant geography. This serves as a proxy for wealth and some of the changing socio-economic character of an area, that might affect the demand for theatre.[8]

4 Panel analysis

We take as our dependent variable the number of theatre visits (ie. transactions). We can view the decision to purchase theatre tickets as having two parts: first, the decision to attend, and second the decision of with whom to attend. The mechanisms of effect we outlined in the introduction seem more likely to affect the former decision, hence the choice of visits, rather than tickets, as the dependent variable.

The dataset is a panel, though it is not a typical individual/time panel, having an extra dimension:

$$\text{audience geographic unit } i \times \text{organisation } j \times \text{year } t$$

We expect individual heterogeneity in any of these dimensions. For example,

- i: Socioeconomic differences affect taste for theatre, so to the extent that these differences are not scatter randomly by geography, this will influence the count of visits
- j: Organisations vary in size, so as a supply side effect, some organisations will simply have more visits in all areas and years
- t: Macroeconomic conditions varying from year to year will affect both the supply of and demand for theatre.

[7] We experimented with with various functional forms for a treatment-distance variable, including nearest venue inverse-distance (or more generally, constant elasticity), sum of inverse distance to all venues, and simple thresholds. Ultimately we preferred distance-band dummies, as these (a) imposed minimal assumptions and perhaps more importantly (b) allowed the most easily interpretable coefficient estimates.

[8] The Land Registry data is a relatively new but very valuable resource for researchers. It contains price paid, exactly location, and various attributes for every private property sale in England and Wales since 1995. It is available at `http://www.landregistry.gov.uk/market-trend-data/public-data`.

In addition, some sources of heterogeneity are likely to vary across pairs of dimensions. For example,

- $i \times j$: The geographic catchments for different organisations vary (mainly based on distance).
- $i \times t$: Geographically-differential change over time (e.g. gentrification or decline of areas) may affect theatre demand. (The NT Live treatment also varies on these dimensions, i.e., by year and by geographic area).
- $j \times t$: Organisations may vary their programmes from year to year, or may have part-year closures.

For each of these sources of heterogeneity we have a choice of either (a) ignoring, (b) controlling for, or (c) modelling the variation. Ignoring heterogeneity risks omitted variables bias, so we attempt to either control for (using fixed effects) or model all the sources of variation above. As usual, fixed effects modelling means that that we can ignore many sources of variation which influence the dependent variable but are fixed in one or two of the dimensions we control for. It also has the usual downside, which is that by focusing on only part of the variation in the data, efficiency of estimation will suffer: we compensate for this with a very large data set.

These considerations lead to the following flexible multiplicative specification for ticket sales, for each organisation j in each geographic unit i for each period t.

$$visits_{ijt} = \alpha_i \gamma_{jt} pop_{it}^{b_1} house_{it}^{b_2} d(i,j)^{b_3} \prod_{k=1}^{5} \delta_k^{treat_{k,it}} \epsilon_{ijt} \tag{1}$$

where α_i and γ_{jt} are fixed effects and $d(i,j)$ is the distances from area i to organisation j.

Here we have directly controlled for i and $j \times t$ variation using fixed effects (with the latter capturing unidimensional j and t variation also); we have modelled $i \times j$ variation using distance; and we have modelled $i \times t$ variation using the treatment variables and *pop* and *house* covariates. In this form, we can now view the model as a regular panel, in which the 'individual' dimension is geographic area i while the 'time' dimension is organisation $\times$ time ($j \times t$).

Rather than imposing a particular functional form, we include multiplicative 1km-banded dummies ($treat_{k,it}$) to allow a coarse non-parametric estimation of treatment as a function of distance. These are coded 1 if there was an NT Live participating cinema within that radius in the previous calendar year, and 0 otherwise.[9]

As the dependent variable, *visits*, is a nonnegative count, we would ideally estimate this using a fixed-effects count panel model (e.g. Cameron and Trivedi, 2013, s. 9.4). Unfortunately, we are not aware of existing panel count models that support two-way fixed effects.[10]

For count data with a sufficiently high mean, a normal approximation argument may justify the use of OLS without too much bias. Here, the mean number of visits varies depending on the level of geographic aggregation: at the highest level of aggregation (MSOA) the mean theatre visits is 27, and so we proceed cautiously to use OLS to estimate the relationship in log-linear terms. In doing so we assume that the error term in the multiplicative model is log-normally distributed. (This normality assumption seems to hold surprisingly well in practice: see Appendix B.)

Transformed in logs we arrive at a linear relationship.

$$\begin{aligned}\log(visits_{ijt}) =& a_i + g_{jt} + b_1 \log(pop_{it}) + b_2 \log(house_{it}) + b_3 \log(d(i,j)) \\ &+ \sum_{k=1}^{5} treat_{k,it} d_k + e_{ijt} \end{aligned} \tag{2}$$

where a_i and g_{jt} are fixed effects; $b_1 \dots b_3$ and $d_1 \dots d_5$ are coefficients to be estimated; and e_{ijt} is an error term, assumed normally distributed. In particular, the fixed effects are

$$\begin{aligned} a_i &= \log(\alpha_i), \text{ a geographic area } i \text{ fixed effect} \\ g_{jt} &= \log(\gamma_{jt}), \text{ an organisation } j \text{ / time } t \text{ fixed effect} \end{aligned}$$

and the treatment-distance coefficients are interpreted as

$$d_k = \log(\delta_k), \text{ (log of) a multiplicative treatment effect at } [k-1, k) \text{ km, for } k = 1 \dots 5$$

all against a base of $\geq$ 5km, which then becomes, by implication, the control region.

One further adjustment is needed to estimate the above equation: since many rows contain zero values for *visits*, we apply an ad-hoc correction, modelling $\log(visits_{ijt} + 1)$

[9]We recognise that assuming a full year lag in the treatment is rather crude, however assuming a contemporaneous effect with annual resolution would obviously be a poor choice. In a future model estimated at monthly frequency, we intend to more carefully test the *time-distribution* of the effect.

[10]Though this appears a relatively simple extension of the one-way case, analogous to same extension of the linear model, and we hope to have results for this model soon.

instead as our dependent variable.[11] The results are presented in Table 1.

Model 1A presents the regression we have described above, focusing on theatre visits. The treatment effect is positive, declining with distance, from 10 per cent within 1 km of an NT Live cinema to around 0% beyond 3 km. This means, for example, that areas within a 1km distance of a cinema participating in an National Theatre live broadcast in the previous year, will produce 10% more theatre visits, on average.[12] This decline is consistent with Hypothesis 2.1, which provides some reassurance that the modelled effect is real.

The other explanatory variables also have reasonable values, which can be interpreted as elasticities due to the log-linear form. House prices are significant in the model, but with a coefficient of 0.253, much less than one: so a doubling of house prices will result in a 19% increase in visits. The coefficient on distance is very close to -1, suggesting a plausible inverse-distance relationship. That the coefficient on population is insignificant is perhaps not surprising: only *spatially-differential time-variation* in population is not captured by fixed effects, and this relatively small over a five year period.

In Model 1B, we test for a possibly differential effect within and outside London, interacting a London dummy with the treatment variables (but not with other explanatory variables). This is motivated both by the qualitatively different London market (e.g. many more theatres competing) and a particular policy interest in capital vs. regional effects.[13] Indeed, the results for London appear to be driving the national results, while there is no consistent pattern to the non-London treatment effects.

As a test of Hypothesis 2.2, we rerun both models using the non-theatre visits data (Model 2A and 2B). This generates an unexpected result, in that the treatment effects appear to be similar to that on theatre (slightly smaller at the closest distances, but slightly larger at mid-distances). Other coefficients are directionally similar to the theatres model: the coefficient on house prices is smaller, and the coefficient on population is larger (and now significant).

Finally, in two further model variants (1C and 1D, full results not reported here), we use a single 3km distance dummy to capture the average effect over the 0-3km radius at which Models 1A and 1B suggest an effect. This resultd in a 5.3% average effect, breaking down as a 6.4% effect in London and an insignificant -0.1% effect outside London.

[11] A variety of monotonic transformations are commonly recommended for dealing with this issue, of which we adopt the simplest. It has the attractive feature of mapping zero to zero. We have estimated the model excluding zero cells and the results are qualitatively similar.

[12] More concretely, we might say that individuals in such a treated zone are 10% more likely to visit a theatre, in aggregate; however, the 'ecological' approach used here means caution must be used in interpreting these results in terms of individual behaviour in this way.

[13] e.g. see `http://www.theguardian.com/stage/2009/jan/14/national-theatre-live-broadcast`

	Model 1A	Model 1B		Model 2A	Model 2B	
Dependent variable:	theatre visits	theatre visits		nontheatre visits	nontheatre visits	
$\log(house)$	**0.253**	**0.251**		**0.088**	**0.085**	
	(0.023)	(0.027)		(0.018)	(0.018)	
$\log(pop)$	0.059	0.080		**0.358**	**0.386**	
	(0.112)	(0.027)		(0.091)	(0.092)	
$\log(d(i,j))$	**-1.068**	**-1.068**		**-1.293**	**-1.293**	
	(0.015)	(0.027)		(0.015)	(0.015)	
Lagged treatment effects		London	Non-London		London	Non-London
$[0,1)$ km	**0.095**	**0.103**	0.054	**0.070**	**0.085**	-0.006
	(0.014)	(0.027)	(0.027)	(0.012)	(0.013)	(0.036)
$[1,2)$ km	**0.063**	**0.073**	0.001	**0.059**	**0.071**	0.005
	(0.010)	(0.027)	(0.027)	(0.010)	(0.010)	(0.029)
$[2,3)$ km	**0.027**	**0.037**	-0.032	**0.063**	**0.070**	0.047
	(0.009)	(0.027)	(0.027)	(0.009)	(0.009)	(0.025)
$[3,4)$ km	0.003	0.005	-0.001	**0.038**	**0.045**	0.018
	(0.009)	(0.027)	(0.027)	(0.008)	(0.008)	(0.024)
$[4,5)$ km	-0.023	**-0.031**	0.026	0.012	0.014	0.009
	(0.009)	(0.027)	(0.027)	(0.009)	(0.009)	(0.021)
Percent effect on visits $=\exp(\text{lagged treatment effect})-1$						
$[0,1)$ km	**10%**	**11%**	6%	**7%**	**9%**	-1%
$[1,2)$ km	**7%**	**8%**	0%	**6%**	**7%**	1%
$[2,3)$ km	**3%**	**4%**	-3%	**7%**	**7%**	5%
$[3,4)$ km	0%	1%	0%	**4%**	**5%**	2%
$[4,5)$ km	-2%	-3%	**3%**	1%	1%	1%
R^2 **(within)**	0.34	0.34		0.43	0.43	

Table 1: Results from OLS estimation of models using equation (2). Standard errors (clustered by MSOA) in parentheses. Bold coefficients are significant at the 1% level.

5 Discussion

The results above represent a confirmation of Hypothesis 2.1 but a challenge to Hypothesis 2.2. We offer two alternative explanations for this.

The first interpretation is that NT Live has had an effect on the order we report, but that this effect has been on both theatre and non-theatre art forms equally. Mechanisms could be suggested to support this: for example, perhaps attending NT Live motivates further non-specific arts attendance, rather than theatre specifically, perhaps because once patrons are introduced to a venue they go on to attend multiple different artforms there.

An alternative explanation is that our analysis is suffering from some omitted variable which is correlated with the NT Live treatment variable. Using non-experimental data, we can of course not rule that out. However, given the range of effects that our panel model controls for, such an omitted variable would have to be (a) correlated with the NT Live roll-out in *both* space and time; and (b) not sufficiently captured by our house price and population variables. It is not straightforward to identify what these variables might be: One would be endogenous participation in NT Live by cinemas in anticipation of future growth in a high-cultural-participation demographic. We are further investigating the process by which cinemas participate in National Theatre Live in order to evaluate this possibility. Another possibility is that cinemas which have participated in NT Live have also tended to participate in other live simulcast programmes which boost non-theatre artforms.

Another interesting aspect of these results is the apparent difference between London and non-London organisations. Given the very different cultural landscapes represented, this is a plausible result. However it may also be an artefact of the model itself: (a) because MSOAs are likely be larger in areas of low population density, which affects our measure of distance[14] or (b) because the real effect of NT Live occurs over a larger scale outside London (which would be plausible if, for example, people are more likely to drive). The former issue is best dealt with by testing the model at various scales of spatial aggregation, while the second issue could be addressed by testing a variety of more flexible treatment-distance specifications (which we are currently doing).

We are also currently undertaking further analysis using non-linear count models, and we hope that these techniques will allow us to analyse more spatially fine-grained data, at the LSOA and OA (as opposed to MSOA) level. In addition, we hope to access

[14]This is a rather technical point. Because we represent an MSOA as its population-weighted centroid, more of the population is likely to live further from the centroid of spatially-large MSOAs than spatially-small MSOAs.

individual data (matched and unique, but anonymised), which would allow us to test much more specific hypothesis (for instance, the effect of NT Live may differ between regular theatre-goers and those new to the art form).

6 Conclusion

In this paper, we used a new, 'big' data set of ticket transactions for 54 performing arts venues across England and spanning a period from early 2009, when National Theatre Live was launched, through to late 2013, to test whether National Theatre Live broadcasts have impacted positively or negatively on theatre attendance.

Our results to date currently echo the conclusions of Bakhshi and Throsby's (2014) study of the impact on National Theatre audiences.. We find no evidence of cannibalisation on theatre attendance at a broad spread of English venues since the National Theatre Live programme was instituted in 2009. Indeed, as in that paper, we find some evidence of a complementarity between NT Live and in-person theatre attendance. Caution is, however, required in interpreting these results.

In particular, the unexpectedly strong effect on non-theatre attendance is worthy of further investigation: it may simply reflect the relatively close substitutability of the theatre and non-theatre productions in our sample. However, it may also suggest that another, unmodelled cause is driving the treatment effect. Although our modelling strategy controls for all of the most obvious sources of such bias, we cannot rule out that some other exogenous change, correlated in both time and space with the NT Live roll-out, but unrelated to it, is driving our results.

Further econometric analysis, which we are currently undertaking, and improved data (as the Audience Finder database grows) may shed light on this anomaly. However, in the meantime, we believe that our results add to the weight of evidence that live broadcast, at least in the case of NT Live, has created, rather than cannibalised audiences.

A A note on the sample of performing arts organisations

The data set was drawn from the Audience Finder national database. This database is still in development, but nonetheless includes millions of households engaging with theatres across England. The sample was taken from 54 performing arts venues over the period 2009–2013, a total original sample representing 44 million tickets sold. Of these venues, 21 (38 per cent) were in London, enabling us to compare effects on London as the National Theatre's home environment, and the rest of the country. Overall, it includes 38 venues that produce or receive quality drama, which is a good fit for the NT live proposition. There is a variation in location (London, city, market town), size (mid to large scale) and product type (producing / receiving). The sample does not include the West End, ATG or HQ group regional theatres. The non-London sample is focused on some specific geographical areas to enable exploration within a local market-place, each of which has a different level of NT Live exposure. For reference: the organisation UK Theatre creates national benchmarks with 102 majority theatre venues (15 per cent in London) and Arts Council England funds 61 (22 per cent in London).

A.1 Audience Finder Programme

The national database of theatre-attenders is part of the Audience Finder[15] programme run on behalf of the sector by The Audience Agency, and funded by Arts Council England. 500 organisations are currently sharing and comparing audience data and collaborating with the insight it generates. Audience Finder is accessible to all cultural organisations in England. There are over 30 geographic clusters sharing data in the programme, and a further 15 sector and art-form groupings, including: visual arts, museums, arts centres, classical and orchestras, rural touring and outdoor arts.

[15] `http://www.audiencefinder.org`

B Regression diagnostics

In using OLS we made a strong assumption that count variable *visits* could be treated as normally distributed, conditional on the covariates. Given the fixed-effect structure, this is difficult to test *ex ante*, however we can examine the regression residuals *ex post* to validate this assumption.

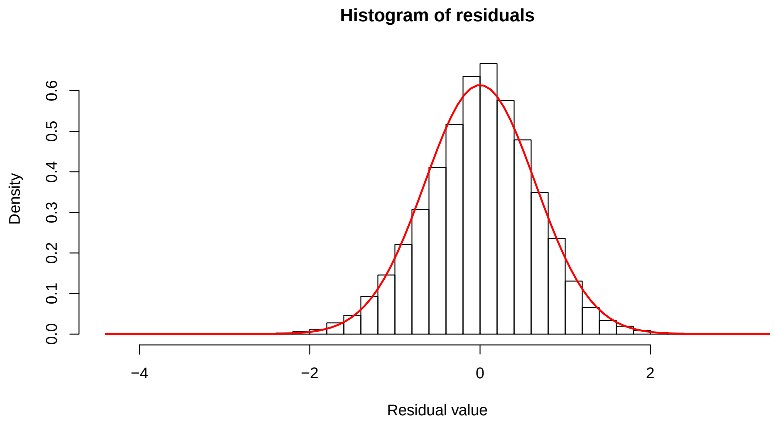

Figure 3: Histogram of residuals from Model 1B with overlaid normal distribution.

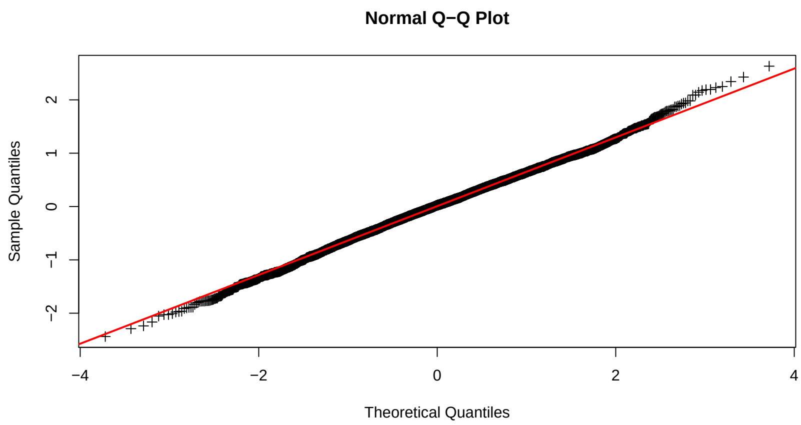

Figure 4: QQ plot of residuals from Model 1B with overlaid normal (using R's qqnorm/qqline).

Residuals for the other models are similar.

References

Bakhshi, Hasan and David Throsby (2012), 'New technologies in cultural institutions: theory, evidence and policy implications', *International Journal of Cultural Policy* **18**(2), 205–222.

Bakhshi, Hasan and David Throsby (2014), 'Digital complements or substitutes? a quasi-field experiment from the royal national theatre', *Journal of Cultural Economics* **38**(1), 1–8.

Brook, Orian, Paul Boyle and Robin Flowerdew (2010), Geographic analysis of cultural consumption, *in* 'Spatial and Social Disparities', Springer Netherlands, pp. 67–82.

Cameron, A. Colin and Pravid K. Trivedi (2013), *Regression Analysis of Count Data*, CUP.

Nesta (2013), Digital culture: How arts and cultural organisations in england use technology, Technical report, London.

ONS (2014), 'A beginner's guide to uk geography: Census geography', http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/index.html. Accessed: 2014-05-27.

Theatrical Management Association (TMA) (2013), 'Prompt'. Issue 69 (trade publication).

Towse, R (2013), Performing arts, *in* R.Towse and C.Handke, eds, 'Handbook on the Digital Creative Economy', Edward Elgar.

Wise, Karen (2014), 'English touring opera—'opera in cinemas' report'. Creativeworks London Working Paper, Number 3.