Schools and House Prices: Spatiotemporal Regression of Quantile Functions on Categorical Covariates

Luke A. Barratt
Humboldt-Universität zu Berlin
Germany

John A.D. Aston
University of Cambridge
United Kingdom

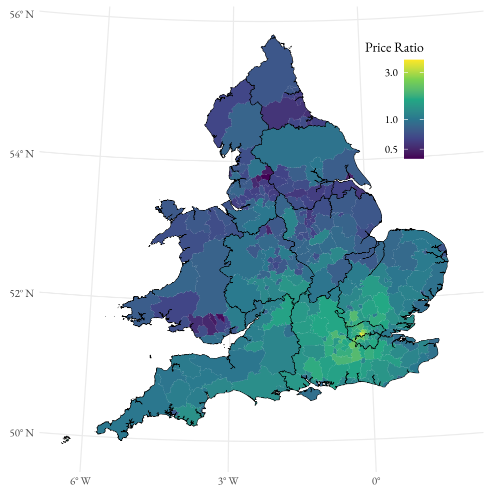
CompStat 2026
Athens, Greece
25 August 2026

House Prices in England

Housing in the UK is a defining economic and social preoccupation:

- Median house prices more than quintupled from 1995 to 2025.
- As a ratio to median earnings, housing has grown from 4.5 in 2002 to over 8 in England and Wales, and over 13 in London, by 2021.
- Residential property is the single largest asset class in the UK economy, estimated at over £9 trillion.

House Prices across Space



School Quality and Ofsted

In England, from 2005 to 2024, the school inspection body Ofsted gave schools an ‘overall effectiveness’ rating: Outstanding, Good, Satisfactory/Requires Improvement or Inadequate.

Quality of schools is not uniform across the country:

- In Westminster (London), an average of 68% of state secondary schools were deemed Outstanding in this period.
- 23 local authorities saw none of their schools achieve an Outstanding rating—none of which were in London.

How associated are school quality and the cost of housing in the local area?

To what extent is this long-term sorting versus capitalisation of Ofsted ratings in local housing stock?

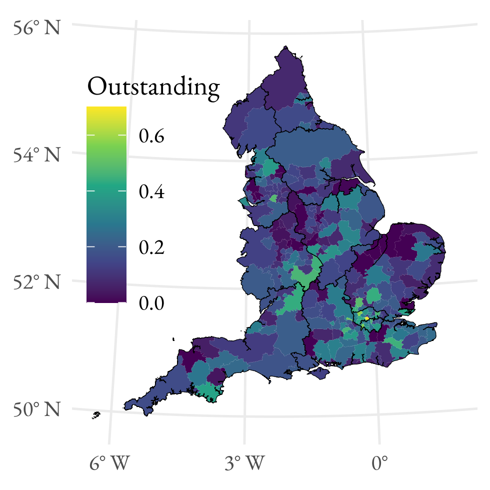
Ofsted Ratings as a Markov Chain

We will consider not Ofsted reports themselves, but the current most recent Ofsted rating of a school. This is a continuous time Markov chain, with the following estimated transition probabilities:

	Out.	Good	Unsat.	Inad.
Out.	32.6%	49.3%	13.4%	4.7%
Good	13.8%	44.7%	31.8%	9.7%
Unsat.	2.5%	43.6%	40.2%	13.7%
Inad.	1.3%	31.0%	56.0%	11.6%

It is worth noting that Outstanding schools get re-inspected less frequently than Inadequate schools.

Ofsted Ratings Across Space



The Distributional Perspective

It is well established that house prices reflect school quality. (SE Black, 1999; S Gibbons and S Machin, 2008)

What is not established is whether this effect is uniform across the distribution of house prices in the local area.

A methodology regression house price distributions in quantile space on the Ofsted rating categorical variable across spacetime is developed, and the results presented.

Setup

We divide England into a Voronoi tesselation centred at 3158 state-funded secondary schools, referenced by s_k, and time from 2010 until 2024 into 59 quarters, referenced by t_j.

For school cell k and quarter j, we then observe n_{kj} registered house sales at price P_{kji}, from assumed local distribution $\mathfrak{P}_{kj}$.

We also observe a compositional covariate $\kappa_{kj} \in [0, 1]^4$, reflecting the current overall effectiveness rating of the school in that quarter. In quarters with a rating change, the rating is divided commensurate with the time spent with each rating.

Regression Model

We model the house price distributions relative to Ofsted ratings by the following concurrent functional generalised additive model:

$$\log \mathfrak{P}_{kj}^{-1}(q) = \alpha(s_k; q) + \beta(t_j; q) + \kappa_{kj}^{\top}\gamma_0(q) + \epsilon_{kj}(q).$$

Note that the log of a quantile function is the quantile function of the logs, and so this is modelled in square-integrable quantile space, $\mathcal{Q} \subset L^2[0, 1]$.

Note that $\mathcal{Q}$ is not linear, but is a convex cone; it turns out that the fitted marginals remain in quantile space nonetheless.

The function β will be unsmoothed, i.e. time will act as a categorical covariate.

The function α will be considered at three levels of smoothness, with the resulting inference on γ_0 being of interest.

I-Spline Expansion of Quantile Functions

We begin by estimating $\log \mathfrak{P}_{kj}^{-1}$ from the data, before proceeding with the regression. We seek a functional data (smooth) representation that respects the constraints of $\mathcal{Q}$.

We expand the log-quantile functions as a median plus a centred I-spline basis:

$$\log \mathfrak{P}_{kj}^{-1}(q) = \eta_{kj} + \sum_{\ell=1}^{L} \pi_{kj}^{(\ell)} \phi_{\ell}^{*}(q) + \xi_{kj}(q).$$

The ϕ_{ℓ}^{*} are integrated M-splines based on L equispaced knots, shifted vertically such that $\phi_{\ell}^{*}(0.5) = 0$.

To ensure the fitted function is monotonically increasing, we restrict to $\pi_{kj}^{(\ell)} \in \mathbb{R}^{+}$.

Estimation of I-Spline Loadings

Given order statistics $\log P_{kj(i)}$ 'observed' at quantiles $q_{kji} := (i - 0.5)/n_{kj}$ (Hazen), we seek to minimise the residuals $\|\xi_{kj}\|^2$:

$$\text{minimise} \quad \sum_{i=1}^{n_{kj}} q_{kji}(1-q_{kji}) \left(\tilde{\eta}_{kj} + \sum_{\ell=1}^{L} \tilde{\pi}_{kj}^{(\ell)} \phi_\ell^*(q_{kji}) - \log P_{kj(i)} \right)^2 + \lambda \left\| \sum_{\ell=1}^{L} \tilde{\pi}_{kj}^{(\ell)} \phi_\ell'' \right\|^2$$
$$\text{over} \quad \tilde{\eta}_{kj} \in \mathbb{R}, \tilde{\pi}_{kj} \in \left(\mathbb{R}^+\right)^L, \quad (1)$$

The regularisation is required for well-posedness, chosen to be just large enough to ensure numerical stability.

This is achieved via an appropriately modified Lawson–Hanson algorithm.

Fitting the Regression Model

We now replace the functional additive model with a system of additive models on the loadings:

$$\pi_{kj}^{(\ell)} = \alpha^{(\ell)}(s_k) + \beta^{(\ell)}(t_j) + \kappa_{kj}^{\top}\gamma_0^{(\ell)} + \epsilon_{kj}^{(\ell)}, \tag{2}$$

with an analogous expression for η_{kj}.

This system of models is fitted independently by an appropriate algorithm to the degree of smoothing, with weights proportional to n_{kj}.

Uncertainty Quantification

An analysis of residuals is performed separately for spatial and temporal autocorrelation.

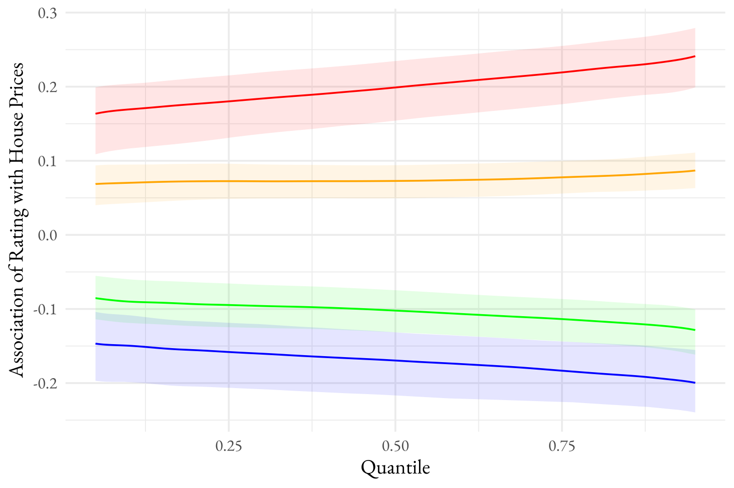
It is notable that temporal autocorrelation persists under all levels of spatial smoothing; thus, standard confidence bands will understate uncertainty.

Spatial autocorrelation persists only for the greater spatial smoothing; in the crucial school-as-factor case, no spatial autocorrelation persists.

Thus, a residual block bootstrap where the block is taken to be a school's entire temporal trajectory in all functional components is performed to approximate sampling distributions of the estimators.

Cross-Sectional Analysis

$$\alpha(s_k; q) = \mu(q)$$

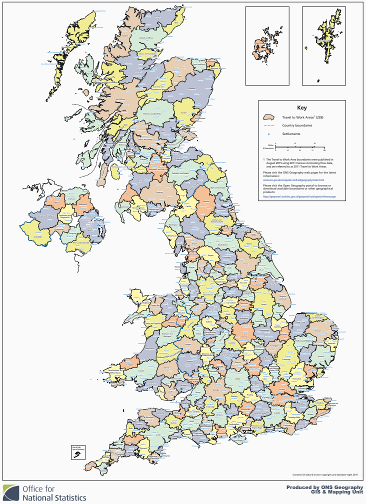


Travel-to-Work Areas (2011)

Travel-to-Work Areas were produced under the following guidelines:

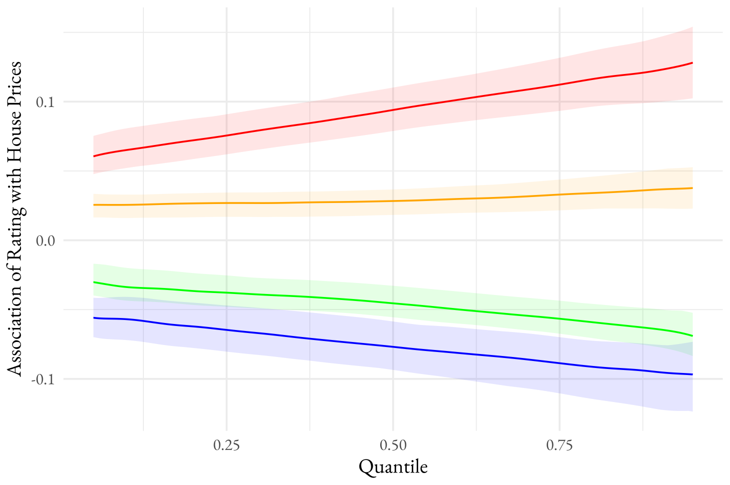
- 75% of the area's resident workforce works in the area; and
- 75% of the people who work in the area also reside in the area.

United Kingdom: 2011 Travel to Work Areas



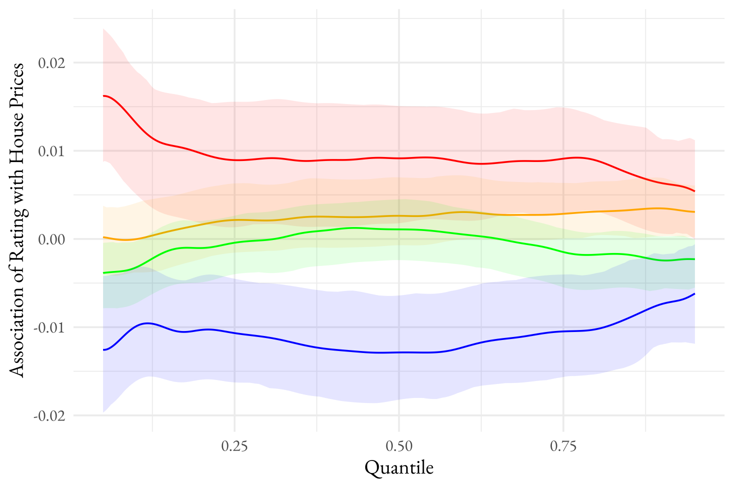
Regression on Travel-to-Work Areas

$$\alpha(s_k; q) = \alpha_{\mathrm{ttwa}(s_k)}(q)$$



Per-School Analysis

$$\alpha(s_k; q) = \alpha_k(q)$$



Thank you for listening.