

Supplemental Information (SI)

The Supplemental Information provides the end-to-end workflow, inclusion criteria, parameter estimation details and sensitivity checks.

A Data sources, scope, and definitions

Source: McGraw *American Street Railway Investments* Directories, 1894–1932 ([McGraw Publishing Co., 1932](#)). We analyse 1894–1926 due to coverage and model window.

Definition: Operating track only, excluding known under-construction segments when identifiable. Mode type not retained. Units are kilometres.

Spatial aggregation: Company-level entries were cleaned and aggregated to metro areas (as listed city names) and to states. Location names were harmonised to avoid duplication.

B Extraction and cleaning workflow

OCR and parsing: Page images were OCR'd ([ABBYY Production LLC, 2013](#)). Regex-based string matching captured company names, locations, and track lengths; ambiguous strings were flagged for manual review.

Manual checks and validation: We hand-checked edge cases and validated selected series against prior hand extractions and related historical syntheses ([Xie and Levinson, 2010](#); [Warner, 1978](#)).

Exclusions: Where the source listed planned or under-construction track that could not be separated, we excluded those entries. Years with known source gaps were left blank rather than imputed.

C Modelling details

For each metropolitan area, and for the United States total, we fit the logistic model in Equation (1) to the annual route-kilometre series $S(t)$. We use ordinary least squares regression, calculated with the Python numpy package, to estimate the model values.

For each series we constrain the asymptote to be at least as large as the observed maximum. Let

$$S_{\max} = \max_{t \in \mathcal{T}} S(t). \quad (2)$$

We require $K \geq S_{\max}$ and impose a loose upper bound K limited by [Equation 3](#), for systems where the observed maximum occurred during the analysed time period, and by [Equation 4](#) if the observed maximum occurred in 1926. The national S-curve is permitted to use $S_{\max} < K < S_{\max} + 5000$

Starting values for (K, b, t_i) are based on S_{\max} and on the years in which the series first reaches roughly ten and ninety percent of S_{\max} , and we treat all years equally in the objective function (no weighting by time or level).

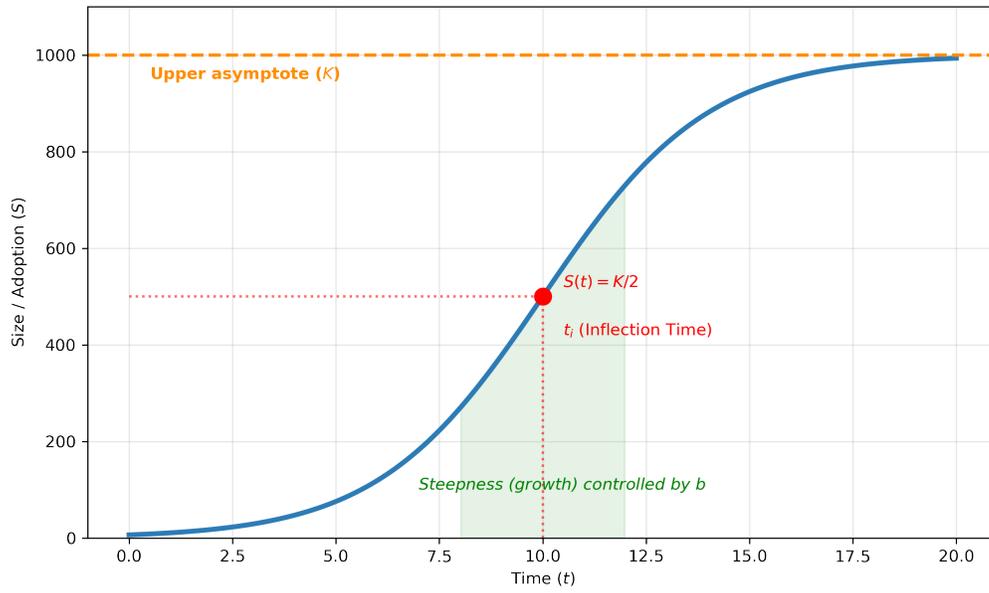


Fig. 5. Generic logistic growth model with asymptotic maximum = 1000.

$$S_{max} < K < S_{max} + 50 \quad (3)$$

$$S_{max} < K < 1.5S_{max} \quad (4)$$

with S_{max} being the maximum observed value.

The fitted parameters and goodness of fit for each series are reported in Tables S1–S4.

Additional detail on processing can be found in (Li, 2021).

D Summary figures

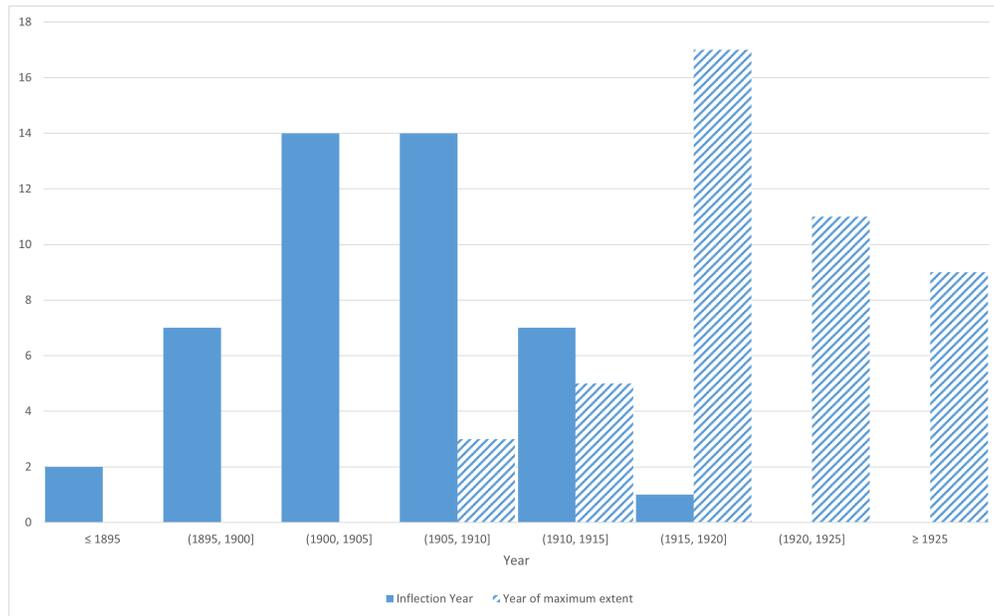


Fig. 6. States: inflection year and year of maximum extent.

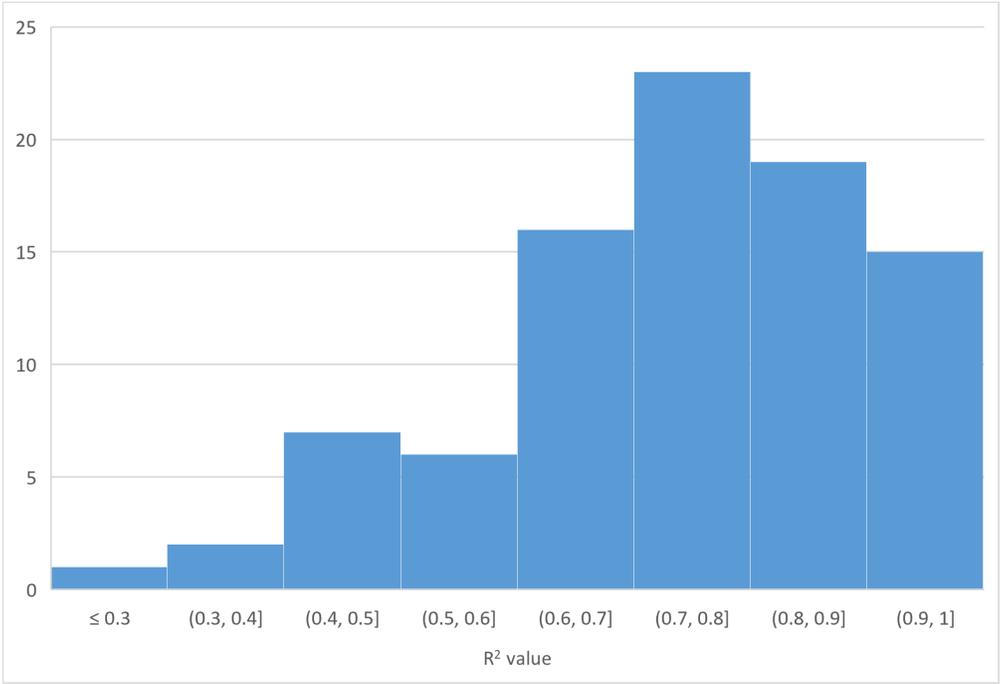


Fig. 7. R^2 for city systems where K reached the allowable limit.

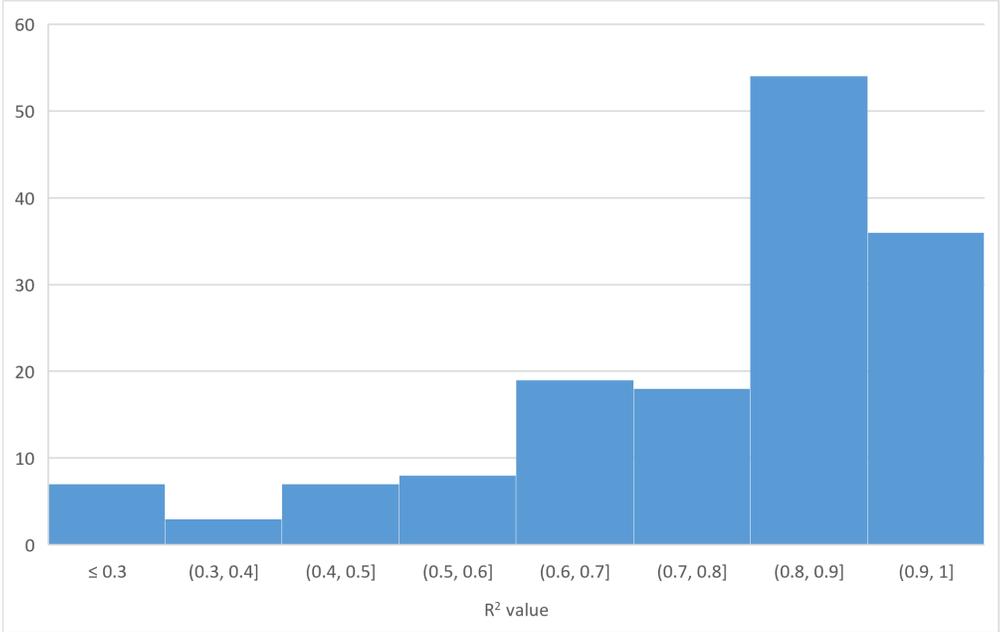


Fig. 8. R^2 for city systems under a higher K limit.

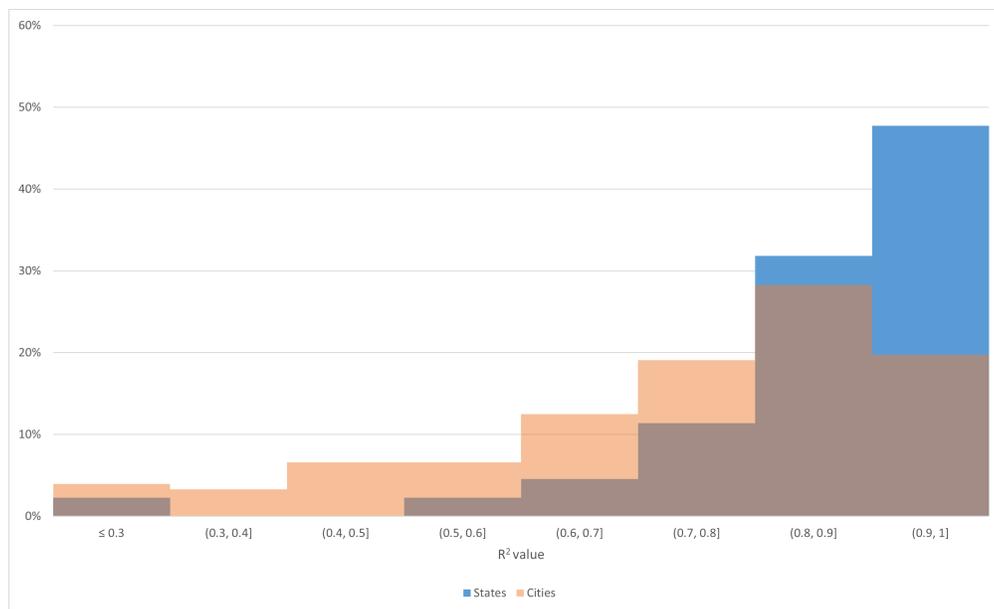


Fig. 9. R^2 distribution for S-curve fits.