Computation and Performance Studies
Volume 1 · Number 1 · 2026

Does Absolute-Loss Fitting Travel Across Forecast Horizons? Evidence from a Public Smart-Home Recording

Charles Russell[1] · Owen Allen[2] · Paul Lewis[3]

[1] Croatia · [2] Greece · [3] Lithuania

Original publication: October 28, 2026 | Article 632

Abstract

In a smart-home recording, an estimator that improves the next quarter-hour forecast need not show equally clear gains an hour ahead. We fitted ordinary least squares and median regression to an earlier SML2010 recording and evaluated both on a later recording from the same house. Daily mean absolute error favored median regression by 0.00035 degrees Celsius at fifteen minutes. At sixty minutes the point difference was larger, 0.00453 degrees, but its dependence-adjusted interval crossed zero. Both models used the same available predictors and evaluation rows within each horizon. The evidence supports a small short-horizon advantage under absolute error, while leaving the longer-horizon comparison uncertain over only fifteen evaluation days.

Chronological temperature-forecast reanalysis. The household recording is real; no new sensor experiment was conducted.

Does Absolute-Loss Fitting Travel Across Forecast Horizons? Evidence from a Public Smart-Home Recording

Chronological temperature-forecast reanalysis. The household recording is real; no new sensor experiment was conducted.

Abstract

In a smart-home recording, an estimator that improves the next quarter-hour forecast need not show equally clear gains an hour ahead. We fitted ordinary least squares and median regression to an earlier SML2010 recording and evaluated both on a later recording from the same house. Daily mean absolute error favored median regression by 0.00035 degrees Celsius at fifteen minutes. At sixty minutes the point difference was larger, 0.00453 degrees, but its dependence-adjusted interval crossed zero. Both models used the same available predictors and evaluation rows within each horizon. The evidence supports a small short-horizon advantage under absolute error, while leaving the longer-horizon comparison uncertain over only fifteen evaluation days.

1. A loss function is part of a forecasting decision

A temperature forecast can be used to plan a control action, detect an unusual change, or summarize a building's thermal state. These purposes need not assign the same cost to every error. Squared-loss fitting emphasizes large residuals, whereas absolute-loss fitting targets a conditional median. Neither choice guarantees lower error under a different evaluation loss, and neither guarantees that a benefit at one forecast horizon persists at another. This article examines that specific horizon dependence using a small public recording.

Our question is whether switching the fitting objective from squared error to absolute error improves later-period absolute prediction loss at both fifteen and sixty minutes. The comparison holds the recorded predictor set and evaluation rows fixed within each horizon. It therefore separates an estimator choice from changes in available information. The two horizons are estimated separately; we do not interpret their coefficient differences as a controlled intervention on a building's thermal dynamics.

The adaptive-loss perspective of Yan et al., 2021 is relevant to asking when a loss choice helps signals observed over time. The present study does not implement normalized LMP estimation or a graph update. It compares ordinary least squares with linear median regression, using a public house recording and a declared held-out period. This gives a transparent empirical setting in which a robust-loss intuition can meet a horizon-specific result without borrowing an unexecuted algorithm's claims.

2. The thermal setting and related evidence

Building models range from physical representations to statistical predictors, as discussed in Foucquier et al., 2013 and Amasyali and El-Gohary, 2018. Data-driven models can use easily observed variables without recovering every thermal mechanism. The appliance-energy dataset in Candanedo, 2017 and the accompanying analysis in Candanedo et al., 2017 illustrate a related public-data setting. We cite them to locate the application area, not to merge households or imply that their observations enlarge the SML2010 sample analyzed here.

Occupant behavior complicates interpretation of residential measurements. The questions raised by Hong et al., 2017 and the appliance and comfort relationships examined in Ruellan et al., 2016 suggest reasons why temperature changes can depend on unrecorded household actions. Appliance-use evidence in Firth

et al., 2008 and Jones and Lomas, 2016 also concerns contexts that a compact sensor regression cannot fully describe. Our analysis contains no inferred occupancy label and does not assign a temperature excursion to a particular appliance, resident, or behavioral decision.

Median regression follows the regression-quantile framework in Koenker and Bassett, 1978. Robustness discussions in Huber, 1964, Hampel, 1974, and Rousseeuw, 1984 help explain why reducing the influence of large fitting residuals may be attractive. Nevertheless, the residuals in this recording are not independently adjudicated sensor faults. They can reflect genuine changes in indoor conditions. An absolute-loss advantage, if observed, should therefore be stated as predictive loss performance rather than proof that the method removed bad measurements.

3. Two recordings from one house

The public resource is Romeu-Guallart and Zamora-Martinez, 2014, whose files contain fifteen-minute mean observations from a domotic house. The earlier file has 2,764 rows spanning March 13 to April 11, 2012; the later file has 1,373 rows spanning April 18 to May 2. The analysis preserves the gap between them. It does not interpolate an artificial continuous trajectory through the unobserved week or use later readings to reconstruct features in the earlier file.

The outcome is the future dining-room temperature minus its current value, in degrees Celsius. Predicting a change helps distinguish the forecast target from merely reproducing the current temperature level. For the fifteen-minute model, a row requires an exactly fifteen-minute predecessor and an exactly fifteen-minute future observation. The sixty-minute model requires the same predecessor and an exactly sixty-minute future observation. Rows at file boundaries or time discontinuities failing those checks are removed before fitting.

The predictor set is current dining-room temperature, its preceding fifteen-minute change, current outdoor temperature, current dining-room humidity, current sun measurement, and sine and cosine terms for time of day. All predictors are available by the forecast origin in the released recording. No future weather reading, subsequent indoor value, or later-file fit residual enters a forecast. The time-of-day terms supply a simple daily pattern without estimating an unrestricted effect for each evaluation date.

For fifteen minutes, the resulting training sample has 2,762 rows and the later evaluation sample 1,367. For sixty minutes, the corresponding counts are 2,759 and 1,358. The difference in evaluation eligibility is an explicit consequence of future matching. Within a horizon, both estimators use identical rows and predictors. We therefore avoid attributing a loss difference to the estimator when it could instead be caused by comparing different missing-data subsets.

4. Freeze the models, then summarize days

Ordinary least squares and conditional-median regression are fitted only on the earlier file. Median regression uses quantile 0.5, a fixed numerical tolerance, and a maximum of 10,000 iterations. The actual fits terminate in 189 iterations at fifteen minutes and 192 at sixty minutes, below that cap. These numerical checks confirm that the reported comparison does not silently substitute an unfinished iterative fit for the intended median estimate.

The later observations are predicted once with each frozen model. Absolute prediction errors are calculated for every eligible origin. Within each calendar date, we average those errors separately for the two estimators and subtract ordinary-least-squares loss from median-regression loss. The resulting daily difference is negative when median regression is better. There are fifteen retained evaluation dates at each horizon. Partial boundary days remain in the sample, and all retained days receive equal weight.

Equal-day weighting answers a different question from pooling every interval with equal weight. A partial day contributes fewer forecasts but still one day's mean. This is a declared choice intended to prevent dates with more eligible intervals from dominating the contrast. It does not make partial and complete days equivalent in precision. An alternative full-day restriction could be useful, but was not selected after inspecting which restriction gave the more favorable estimator result.

Forecast-comparison ideas in Diebold and Mariano, 1995 and West, 1996 motivate focusing on paired out-of-sample loss rather than training fit. Dependence is accommodated by an intercept-only daily-loss model with a fixed two-day HAC covariance based on Newey and West, 1987, small-sample correction, and a t reference with fourteen degrees of freedom. The short evaluation series limits how convincingly any such approximation can capture persistence. A narrow interval should consequently be read together with the very small absolute effect size.

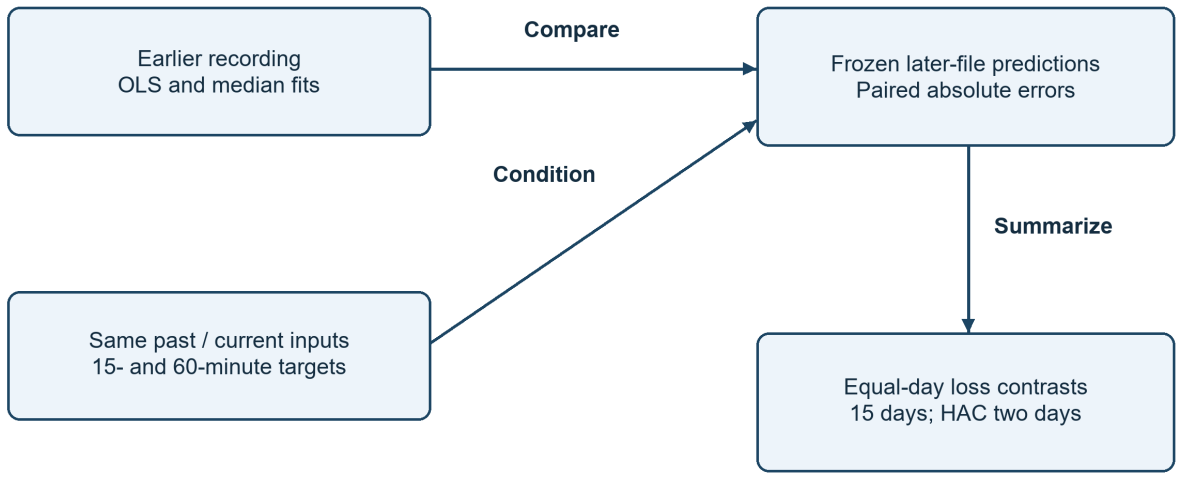


Figure 1. Earlier-file fitting and later-file daily loss evaluation, repeated at two declared horizons.

Table 1. Median-regression minus ordinary-least-squares daily absolute error, measured in degrees Celsius.

Model	Term	Estimate (SE)	95% CI	N
M1	Mean contrast	-0.000346 (0.000092)	[-0.000544, -0.000147]	15
M1: fit	R-squared = 0.0000	-	-	15
M2	Mean contrast	-0.004531 (0.002124)	[-0.009087, 0.000026]	15
M2: fit	R-squared = 0.0000	-	-	15

Model labels: M1 = 15-minute later-file LAD minus OLS daily MAE; M2 = 60-minute later-file LAD minus OLS daily MAE.

5. The fifteen-minute advantage is small

At fifteen minutes, equal-day mean absolute error is 0.018216 for ordinary least squares and 0.017871 for median regression. The mean difference is -0.000346, with standard error 0.000092 and a 95 percent interval from -0.000544 to -0.000147. On the chosen absolute-error criterion, the median fit performs better in this later recording. The difference is numerically tiny in Celsius and should not be promoted into an energy-saving or comfort benefit without a control decision and an appropriate cost model.

The earlier-file ordinary-least-squares R-squared for temperature changes is 0.9494, while the median-regression pseudo-R-squared is 0.8141. These measures use different objectives and are not a direct contest on a common scale. Neither establishes external forecasting performance. The intercept-only comparison in Table 1 has effectively zero R-squared because it estimates a mean loss difference, not a regression explaining why individual dates favor one model over another.

The practical importance of the difference also depends on the quality and resolution of the original measurements. The public recording does not establish an independently verified error-free temperature target. Measurement-error concerns in Carroll et al., 2006 remain relevant when reported differences are much smaller than a degree. Agreement with the recorded average is the observed endpoint. Whether the difference is consequential for a physical actuator or a resident's comfort is not tested by a confidence interval around that endpoint.

6. A larger point gain with less certainty at one hour

At sixty minutes, ordinary-least-squares daily error averages 0.092616 and median-regression error 0.088086. The difference is -0.004531, with standard error 0.002124 and interval -0.009087 to 0.000026. The point estimate favors the median fit, but the reported interval includes zero. Thus the data do not provide the same clear directional evidence as at fifteen minutes under the fixed covariance calculation, even though the absolute point difference is larger.

Longer-horizon changes may allow more unobserved actions and disturbances to accumulate. That is a plausible interpretation, not a mechanism identified by the present regression. We do not observe a randomized change in heating schedules or independently labeled disturbances. The source contains one house and a short later period. An explanation based on household behavior would require additional measurements or a design that separates those actions from weather and time-of-day conditions.

Earlier-file ordinary-least-squares R-squared at sixty minutes is 0.9214 and the median pseudo-R-squared is 0.7684. Again, the fit summaries describe their own objectives. The lower later-period loss of one method cannot be inferred from comparing those two training statistics. This distinction is especially important when a more favorable training criterion can encourage selection of a model whose particular loss advantage does not transport across periods or horizons.

The forecasting metrics discussed by Hyndman and Koehler, 2006 clarify why stating the scale matters. Here changes in Celsius permit a direct absolute-error comparison within the same series, without dividing by temperature levels or constructing a percentage error near zero change. Gneiting and Raftery, 2007 also emphasizes the relationship between a forecast functional and its evaluation objective. The median is naturally associated with absolute loss; the observed comparison still needs held-out evidence because linear specification and distribution change can matter more than that formal alignment.

7. What the temporal holdout protects, and what it cannot protect

The recording order prevents future observations from entering fitted coefficients. Validation guidance in Bergmeir et al., 2018 and Roberts et al., 2017 explains why temporal dependence requires care beyond an ordinary random row split. The separation does not create fifteen independent household replications: all evaluation dates belong to one environment. Nor does it make the earlier and later distributions interchangeable. The gap can coincide with changed occupancy, weather, or control behavior that the available variables do not fully capture.

Model-selection bias discussed by Varma and Simon, 2006 would become important if we tried many lag sets, horizons, and error summaries and reported only a favorable combination. This study retains the two stated horizons, common predictor set, and both loss contrasts. The uncertain one-hour result remains in the report. We do not choose a separate HAC lag after checking which value makes an interval exclude zero, and no claim is made that the two-day choice is uniquely correct.

Heteroskedasticity-only inference in White, 1980 and influence diagnostics in Cook, 1977 concern useful but different checks. The daily loss comparison uses serial covariance instead. Influential dates have not been removed because they make the one-hour result less certain. The missing-data framework in Little and Rubin, 2002 also cautions that observations absent from the public files need not resemble those retained after exact time matching. The analysis cannot recover the unrecorded period or certify a mechanism for every exclusion.

8. Implications for a small forecasting study

The main proposition receives horizon-dependent support. A fixed change of fitting loss gives a small observed benefit at fifteen minutes, while the one-hour interval remains compatible with no average difference. This is an empirical reason to report horizons separately. Averaging their errors together would mix different prediction tasks and obscure both the magnitude and uncertainty of each result. It would also tempt a general claim about robustness that the design does not establish.

The interpretation advice in Wasserstein and Lazar, 2016 and the research-validity concerns in Ioannidis, 2005 favor a cautious reading of such a short test period. Statistical separation from zero does not establish practical importance, while an interval crossing zero does not establish equivalence. A larger multi-house study could define a smallest consequential Celsius error difference in advance and examine whether benefits generalize across households, seasons, and operational decisions rather than merely adding more adjacent records from this same home.

Source provenance and transformations follow Gebru et al., 2021: original file counts, observation periods, temporal eligibility, predictor definitions, fitting tolerances, and daily weighting are recorded with the processed predictions. Wilkinson et al., 2016 motivates keeping those intermediate quantities available for independent recalculation. The resulting contribution is a bounded comparison of fitting losses under real temporal variation. It offers an auditable starting point for choosing a forecast objective, without claiming that a particular objective is universally preferable in residential control.

References

1. Roberts, David R., et al. “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure.” *Ecography*, vol. 40, no. 8, 2017, pp. 913-929. https://doi.org/10.1111/ecog.02881.

2. Foucquier, Aurélie; Robert, Sylvain; Suard, Frédéric; Stéphan, Louis; Jay, Arnaud. (2013). State of the art in building modelling and energy performances prediction: A review. Renewable and Sustainable Energy Reviews. https://doi.org/10.1016/j.rser.2013.03.004.

3. Yan, Yi; Adel, Radwa; Kuruoglu, Ercan E.. (2021). Adaptive Normalized LMP Estimation for Graph Signal Processing. 2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP), 1-6. https://doi.org/10.1109/mlsp52302.2021.9596181.

4. Ioannidis, John P. A.. (2005). Why Most Published Research Findings Are False. PLoS Medicine. https://doi.org/10.1371/journal.pmed.0020124.

5. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

6. Candanedo, Luis M.; Feldheim, Véronique; Deramaix, Dominique. (2017). Data driven prediction models of energy use of appliances in a low-energy house. Energy and Buildings. https://doi.org/10.1016/j.enbuild.2017.01.083.

7. Wasserstein, Ronald L.; Lazar, Nicole A.. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician. https://doi.org/10.1080/00031305.2016.1154108.

8. Hampel, Frank R.. (1974). The Influence Curve and its Role in Robust Estimation. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1974.10482962.

9. West, Kenneth D.. (1996). Asymptotic Inference about Predictive Ability. Econometrica. https://doi.org/10.2307/2171956.

10. Firth, S.; Lomas, K.; Wright, A.; Wall, R.. (2008). Identifying trends in the use of domestic appliances from household electricity consumption measurements. Energy and Buildings. https://doi.org/10.1016/j.enbuild.2007.07.005.

11. Little, Roderick J. A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002. https://doi.org/10.1002/9781119013563.

12. Rousseeuw, Peter J.. (1984). Least Median of Squares Regression. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1984.10477105.

13. Jones, Rory V.; Lomas, Kevin J.. (2016). Determinants of high electrical energy demand in UK homes: Appliance ownership and use. Energy and Buildings. https://doi.org/10.1016/j.enbuild.2016.02.020.

14. Newey, Whitney K., and Kenneth D. West. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, vol. 55, no. 3, 1987, pp. 703-708. https://doi.org/10.2307/1913610.

15. Amasyali, Kadir; El-Gohary, Nora M.. (2018). A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews. https://doi.org/10.1016/j.rser.2017.04.095.

16. Ruellan, Marie; Park, Herie; Bennacer, Rachid. (2016). Residential building energy demand and thermal comfort: Thermal dynamics of electrical appliances and their impact. Energy and Buildings. https://doi.org/10.1016/j.enbuild.2016.07.029.

17. Diebold, Francis X.; Mariano, Roberto S.. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics. https://doi.org/10.1080/07350015.1995.10524599.

18. Koenker, Roger; Bassett, Gilbert. (1978). Regression Quantiles. Econometrica. https://doi.org/10.2307/1913643.

19. Bergmeir, Christoph; Hyndman, Rob J.; Koo, Bonsoo. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis. https://doi.org/10.1016/j.csda.2017.11.003.

20. Varma, Sudhir and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.

21. Hong, Tianzhen; Yan, Da; D’Oca, Simona; Chen, Chien-fei. (2017). Ten questions concerning occupant behavior in buildings: The big picture. Building and Environment 114, 518-530. https://doi.org/10.1016/j.buildenv.2016.12.006.

22. Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; ’t Hoen, Peter A.C; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18.

23. Hyndman, Rob J. and Anne B. Koehler. “Another look at measures of forecast accuracy.” *International Journal of Forecasting*, vol. 22, no. 4, 2006, pp. 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

24. White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.

25. Romeu-Guallart, Pablo; Zamora-Martinez, Francisco. (2014). SML2010 [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5RS3S.

26. Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. “Measurement Error in Nonlinear Models.” 2006. https://doi.org/10.1201/9781420010138.

27. Huber, Peter J.. (1964). Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics. https://doi.org/10.1214/aoms/1177703732.

28. Gneiting, Tilmann; Raftery, Adrian E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. https://doi.org/10.1198/016214506000001437.

29. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.

30. Candanedo, Luis. (2017). Appliances Energy Prediction [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5VC8G.