Economic History Working Papers No: 379

Predictive Modeling the Past

Meredith Paker, Grinnell College,
Judy Z. Stephenson, UCL
and
Patrick Wallis, LSE

June 2025

Economic History Department, London School of Economics and Political Science,
Houghton Street, London, WC2A 2AE, London, UK. T: +44 (0) 20 7955 7084.

Predictive Modeling the Past*

Meredith M. Paker† Judy Z. Stephenson‡ Patrick Wallis§

13th June 2025

Abstract

Understanding long-run economic growth requires reliable historical data, yet the vast majority of long-run economic time series are drawn from incomplete records with significant temporal and geographic gaps. Conventional solutions to these gaps rely on linear regressions that risk bias or overfitting when data are scarce. We introduce "past predictive modeling," a framework that leverages machine learning and out-of-sample predictive modeling techniques to reconstruct representative historical time series from scarce data. Validating our approach using nominal wage data from England, 1300-1900, we show that this new method leads to more accurate and generalizable estimates, with bootstrapped standard errors 72% lower than benchmark linear regressions. Beyond just bettering accuracy, these improved wage estimates for England yield new insights into the impact of the Black Death on inequality, the economic geography of pre-industrial growth, and productivity over the long-run.

JEL classification: J31, C53, N33, N13, N63
Keywords: Machine Learning, Predictive Modeling, Wages, Black Death, Industrial Revolution

*Corresponding author: Meredith M. Paker, Grinnell College. Email: pakermer@grinnell.edu. The authors would like to thank Gregory Clark, Yuqi Chen, Fernando Lozano, and the participants at the IHR Economic History of the Early Modern World seminar, the HKU Quantitative History webinar, the 2025 AALAC Economic History workshop, the University of Oxford's seminar, the Lund University economic history seminar, the Economic History Society 2025 conference, and the Yale Economic History workshop for their comments, questions, and ideas that have improved the paper. We are grateful to Gregory Clark for generously sharing data. To use the past predictive modelling method, please consult and download the files at https://github.com/merpaker/PastPredictiveModeling/

†Grinnell College
‡University College London
§London School of Economics

1 Introduction

When an economist begins a project on long-run economic growth or development, their first action is usually to download the benchmark data series that are held in one of the worlds' major repositories. These series of wages, population, income, and other variables appear robust and reliable. In reality, even the most authoritative of these are built on fragile foundations. Usually, they face two serious problems: the scarcity of the underlying data and the validity of the methodology used to derive a series from it.

While today's economic data are based on large numbers of observations, systematic sampling, and statistical adjustment, usually in the hands of reliable government agencies, data for earlier historical periods, which comprise the vast majority of economic time, are pieced together from fragmentary and incomplete records. They are a patchwork with significant temporal and geographic gaps.

Because historical records are riddled with holes, the method used to derive a long-run representative data series is critically important. The standard approach to addressing gaps in the quantitative historical record relies heavily on linear regression. Regressions are used to generate series of average fitted values for each period from scarce and inconsistent data. This produces a seemingly continuous time series, where annual estimates rely on parametric assumptions about the evolution of economic relationships. But there is a problem with this approach: as data become more scarce in each period, these standard curve-fitting exercises can lead to bias or overfitting to the few data points available. The resulting average fitted value for a period is less representative of the true population. To secondary users of the data, these limitations—and, sometimes, even the full extent of the underlying patchiness of the original data—are not always fully apparent. Yet they can introduce bias into results and lead to mistaken conclusions.

In this paper, we introduce "past predictive modeling," a framework that leverages machine learning and out-of-sample predictive modeling techniques to more accurately reconstruct long-run time series from fragmentary historical data. Our approach differs fundamentally from existing linear regression methods: instead of fitting a single curve to sometimes hundreds of years of data from the past, risking bias and overfitting in periods where data are more scarce, we use more recent data to make out-of-sample predictions of values in earlier periods. The standard (forward) predictive modeling literature addresses bias and overfitting by focusing on how well models can predict unseen out-of-sample data. We take this solution and apply it to the analogous problem of scarce data in the past. Rather than predicting future outcomes, we use more recent data to make predictions of past values, optimizing out-of-sample predictive accuracy to generate more representative estimates than standard regression methods.

In addition to developing an out-of-sample predictive modeling framework that addresses temporal dependencies in the data, we also advance the method of prediction. Our flexible framework enables any prediction algorithm to be used, so we move beyond linear regressions to test and implement multiple machine learning algorithms that are known make superior out-of-sample predictions.

FIGURE 1: COMPARATIVE PERFORMANCE: PAST PREDICTIVE MODELING VS. TRADITIONAL REGRESSION APPROACH

Note: Average performance metrics 1220–1900 for unskilled laborers. Linear regression on the left panel follows that of Clark (2005). Linear regression on the right panel run within the out-of-sample prediction framework. For full details, see Section 5.

Our main contribution is to elucidate this new past predictive modeling framework and to show empirically that it yields long-run historical economic time series that are more precise and generalizable than benchmark linear regressions in the face of patchwork historical data. Using English nominal wages as an example, we find that our method reduces bootstrapped standard errors by 60% on average relative to linear regression models and improves out-of-sample accuracy by over 28%. Figure 1 summarizes these results. We also illustrate that these methodological improvements meaningfully shift long-run nominal wage patterns, giving new insights on long-run economic development.

Our argument in this paper is developed in four steps. First, we assess the problem of data scarcity affecting long-run economic statistics and consider existing techniques with which it has been addressed. We show that the dominant methodology for generating annual series in the recent literature is fitting linear regressions, even in periods of limited data. Regression-based series have been collected in major data repositories, and scholars are increasingly relying on these series for analyses of very-long-run developments across many subfields. Then, we describe three major limitations of the standard regression approach and how they can be addressed in our new framework: the performance of the method under data scarcity, the dimensionality problem, and the inflexibility of the method.

In the second step, we set out our alternative methodological framework, past predictive modeling. We describe a simple predictive modeling framework that allows for backward temporal prediction of incomplete data. Our framework explicitly addresses temporal dependencies in the data and optimizes on out-of-sample predictive accuracy using an expanding window "walk backward" technique. The key innovation of this framework is using more recent data to make generalizable short-run predictions of historical values of an economic outcome. The framework is flexible to choice of machine learning algorithm and adaptable to many economic variables.

The third step is to empirically validate our new framework by directly comparing it to existing techniques. To do so, we select as an application long-run trends in nominal wages over the course of seven centuries in England (Clark 2005). This is an appropriate data setting in which to test our framework because nominal wages are a common measure of very-long-run economic development, and England is a classic case study.[1] We describe the nominal wage data available for England for the past seven centuries in Section 4. Then, in Section 5, we pit our past predictive modeling framework against linear models for estimating patterns of nominal wages. We show that the past predictive modeling framework leads to large reductions in bootstrapped standard errors and improved accuracy in predicting out-of-sample data. As an additional robustness check, we demonstrate improvements of a similar magnitude for another setting: servant's wages in Japan (Kumon 2022).

In the fourth step, we show that this framework not only leads to more accurate predictions, but that these new predictions have actual implications for our understanding of historical economies. Section 6 demonstrates that our new estimates for England yield novel insights on foundational questions about the evolution of wages, inequality, and productivity. We find paths of wages following the Black Death in the mid-1300s to be highly differentiated by skill, revising our understanding of the impact of this demographic shock on income inequality and the skill premium. We also show how our improved methodology can generate entirely new series previously unavailable to economists, generating for the first time long-run regional wage series that shed light on timings of key phases of development and

[1] These wages might more appropriately be thought of as labor costs. See Appendix A for a discussion of the distinction between wages and labor costs.

which can be aggregated into a population-weighted national average. Finally, we demonstrate that our revisions to this series impact secondary analyses where wages are a key variable by showing how they affect estimates of productivity growth from Bouscasse et al. (2025).

Related literature

Our paper contributes to several strands of literature. We contribute to the emergent subfield of machine learning for predictive modeling in economics, and to the use of machine learning for economics in historical settings. We also contribute the evaluation of long-run economic growth and the study of historical time series, and to literature on historical sources and data management in econometric analysis.

Recent work in economics has shown that, in cases where the goal is to make accurate predictions rather than to establish causal relationships, certain machine learning algorithms regularly outperform traditional OLS and time series regression techniques. For example, Mullainathan and Spiess (2017) demonstrate this for housing prices, Richardson et al. (2021) and Yoon (2021) for forecasting GDP, and Bluwstein et al. (2023) for predicting recessions, and Bajari et al. (2015) for demand estimation. These are cases, like ours, where obtaining $\hat{y}$ (in the language of Mullainathan and Spiess (2017)), is a core economic objective. Related works on "prediction policy" problems show improved accuracy with machine learning algorithms over regression techniques when the goal is prediction (Kleinberg et al. 2015; Chalfin et al. 2016). Goulet Coulombe et al. (2022) unpack the features driving the gains of machine learning over standard methods, finding that the flexibility and non-linearity of these methods the key to their strong performance. In general, the literature has found that machine learning provides advantages in predictive accuracy for complicated systems characterized by complex and non-linear relationships between variables under changing conditions.

Our paper is the first to apply these benefits of machine learning for predictive modeling to the problem of scarce historical data in long-run time series. We are not, however, the first to use machine learning in historical settings. Machine learning has revolutionized both the gathering of data sets from complex historical sources and their tractability. As Dell (2025) highlights, machine learning has aided in classification, document digitization, record linkage, and the analysis of text and image corpora for historical documents. Many papers rely on machine learning for linking individuals across censuses such as Price et al. (2021), Abramitzky et al. (2021), Helgertz et al. (2022) and Buckles et al. (2024). Natural language processing is a common technique for working with historical text data (see Ferguson-Cradler 2023).

Dell et al. (2023) develop a deep learning pipeline for extracting full-text articles from newspaper images, generating high quality data and better understanding of "historical world knowledge." We contribute to this literature by establishing the use of predictive modeling machine learning in historical settings.

Many papers have summarized and surveyed the available machine learning methods for prediction. Masini et al. (2023) provide a general framework and summary of these methods for time-series forecasting. Gu et al. (2020) discuss the methods in detail in the setting of empirical asset pricing. Dell (2025) provides an accessible summary of deep learning approaches. While the objective of our paper is not to summarize the various methods available, we base the structure of our neural networks on those developed in Gu et al. (2020).

We also contribute to literature on the fundamental task of constructing long-run economic indicators. Reliable long-run data are critical to all attempts to understand economic growth and development (Allen 2001, 2009; Bouscasse et al. 2025; Broadberry et al. 2015; Clark 2005, 2007a; Crafts 2021; Feinstein 1998; Humphries and Weisdorf 2019). Since the seminal work of Phelps Brown and Hopkins (1955), nominal wages have been among the most important series in use for this purpose. Advances in the scope of wage series and their comparative measurement since the turn of the century by Allen (2001, 2015) and Clark (2005, 2007b) have led to major shifts in our understanding of long run development, including debates on the role of factor prices in driving technological change in the first industrial revolution (Allen 2015), and the persistence of Malthusian forces (Clark 2007a). Our new estimates of long-run nominal wages contribute directly to this literature.

Our method also contributes more broadly to work on constructing long-run indicators and generating predicted values using regressions. We survey many such papers related to wages in Table 1 below (Clark 2005, 2007b; Allen et al. 2011; De Zwart and Van Zanden 2015; Pfister 2017; Ridolfi 2019; Humphries and Weisdorf 2019; Horrell and Humphries 2019; De Zwart and Lucassen 2020; Federico et al. 2021; Rota and Weisdorf 2021; Kumon 2022; Chambru and Maneuvrier-Hervieu 2023; Buscemi 2025; Carvalhal et al. 2024; Liu 2024). Similar methods have been used to predict other economic phenomena (for example: Zhai 2024; Du Plessis et al. 2015; Mitchener and Weidenmier 2015; Etro 2018; Koepke and Baten 2005; Samy 2015; Karagedikli and Tunçer 2021; Raff et al. 2013; Edquist 2010; Coşgel and Ergene 2012).

A final contribution of our paper is the contextualization of the ready availability of historical data from resources such as the Bank of England and FRED. These repositories are increasingly valuable to researchers in many subfields. We highlight the scarcity of data underlying some of the series in these repositories. We also show empirically that shifts in predicted values arising from past predictive modeling meaningfully shape secondary analyses where wages are a core component.

2 The problem of scarce data

In an age of big data, the scarcity of original observations from reliable sources remains a critical challenge for long-run economic measurement. In this section, we document the presence of data scarcity even in estimates that appear founded in large samples. Then, we describe approaches to dealing with scarce data, using the case of wages. We explore the limits of the current best practice methodology, linear regression, and describe how, in the presence of scarce data, estimates from regression models can be biased and unrepresentative.

2.1 Data scarcity

Data scarcity is a serious problem in long-run economic datasets. We illustrate this with the case of nominal wage series as these are among the best-founded datasets. The wage estimates for skilled and unskilled workers in England between 1209 and 1914 made by Clark (2005) are, to our knowledge, based on the largest collection of historical wage observations from a country with an unusually rich and well-preserved set of records.[2] After cleaning the data and removing duplicates, Clark's data set contains 39,223 observations of day wages from multiple sources, including 9,591 observations of laborers or unskilled workers, leaving, on average, 136 observations of unskilled wages per decade. However, data are unevenly distributed, and even in this setting data scarcity generally increases with distance from the present. As Figure 2 shows, while some decades have abundant laborer data (such as the 1890s, with over 900 total observations) others have almost no data points (such as the 1240s, with only three total observations). On average, there are only thirteen observations of laborers' wages in each year. The problem of scarce data is especially acute after the Black Death and in the mid fifteenth century, late seventeenth century, and mid eighteenth century—all key periods of interest.

Data scarcity is a general issue with long-run economic indicators, as we can see in Table 1, which summarizes the sample, observations, and parameterization of a range of papers using regression models to

[2]By one measure, the data in Carvalhal et al. (2024) is larger, but it is based on expanding a much smaller number of observations that specify the wages for large groups of individual workers who receive the same pay into individual observations for each worker within that group.

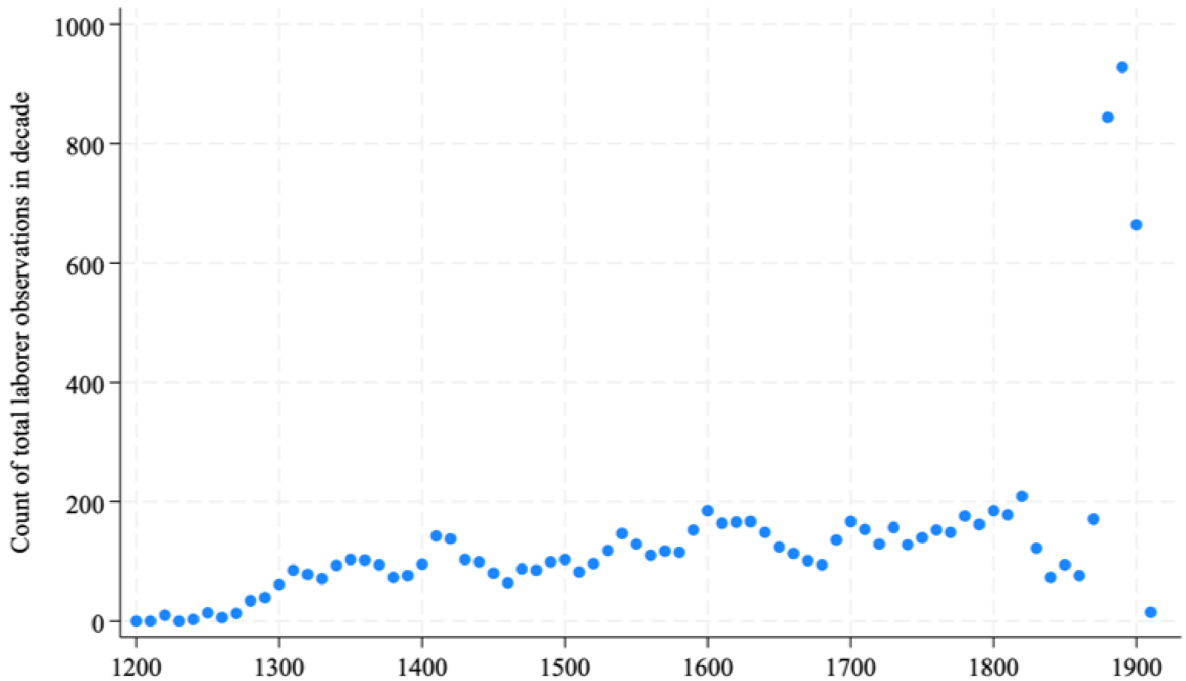

FIGURE 2: DECADAL OBSERVATION COUNTS IN THE CLARK (2005) LABORER SAMPLE

estimate historic wages from around the globe. Despite the large size of the datasets built by researchers, few contain more than fifty observations per year studied, even after heroic work with idiosyncratic primary sources. The inclusion of controls for factors that might plausibly affect the wage, such as gender, occupation, payment type, source, lead to a high number of parameters in each estimation. As a result, even studies with a relatively large number of observations per year frequently have periods and categories in which the estimation is reliant on very small samples. These estimates are, nonetheless, presented and used as nationally representative wage data.

We are unlikely to solve the problem of data scarcity, which is driven by broader issues of data survival and collection. The challenge to the researcher is how to respond to this limitation in order to achieve the best possible estimates of economic indicators.

2.2 Existing methodological approaches

The papers listed in Table 1 contain the best estimates of long run day wages from different regions to date. All apply the current best practice methodology of regression analysis. In this section, we briefly discuss the development of this methodological approach to constructing long-run wage series and its limitations.

Table 1: Review of papers with wage time series based on regression models

Paper	N	Years	R2	Param.	Obs. per year	Obs. per decade	Obs. per param. per year
Clark (2005)	46,000	795	-	47	57.86	578.62	1.23
Clark (2007b)	19,417	660	-	39	29.42	294.20	0.75
Allen et al. (2011)	327	80	0.408	12	4.09	40.88	0.34
De Zwart and Van Zanden (2015)	587	190	0.48	3	3.09	30.89	1.03
Pfister (2017)	2,187	350	0.77	18	6.25	62.49	0.35
Ridolfi (2019)	26,332	610	0.991	328	43.17	431.67	0.13
Humphries and Weisdorf (2019)	6,800	590	-	12	11.53	115.25	0.96
Horrell and Humphries (2019)	3,873	580	0.567	12	6.68	66.78	0.56
Gary and Olsson (2020)	28,500	350	-	8	81.43	814.29	10.18
De Zwart and Lucassen (2020)	7,586	280	-	17	27.09	270.93	1.59
Federico et al. (2021)	14,513	52	-	16	279.10	2,791	17.44
Rota and Weisdorf (2021)	439	350	0.86	5	1.25	12.54	0.25
Kumon (2022)	1,736	280	-	41	6.20	62.00	0.15
Chambru and Maneuvrier-Hervieu (2023)	19,786	249	-	118	79.46	794.62	0.67
Carvalhal et al. (2024)	69,325	129	0.52	31	537.40	5374.03	17.34
Liu (2024)	6,006	310	0.936	46	19.37	193.74	0.42
Buscemi (2025)	23,490	310	-	82	75.77	757.74	0.92

Note: The measure presented of observations per parameter/year treats sums the number of parameters in separate control variables and expresses them over the years in observation. In use in the model, the potential set of variations increases geometrically, so readers are encouraged to treat this measure as a heuristic device only.

Constructing estimates of wages for occupations, regions, or nations presents an empirical and conceptual challenge that economic historians have addressed in a number of ways over more than a century of research. Those working on more modern periods are, of course, able to rely on series gathered from official sources, deferring judgement about the representativeness of the wages they use to original investigators (De Zwart 2011; Frankema and Waijenburg 2012; Cvrcek 2013; Allen and Khaustova 2019; Mijatović and Milanović 2021, for example). Work on earlier periods generally takes one of two main approaches: identifying a local average wage directly or estimating a representative wage using a linear regression.

The first approach, evident from the earliest work in the late nineteenth-century, focused on identifying an average (modal, mean, or representative) wage paid to men in given industries on a daily or weekly basis (Bowley 1900; Gilboy 1934). For work on single institutions or cities this approach is still widely used (Allen 2001; Paker et al. 2025; López Losa and Zarauz 2021).[3] In all these studies, the reliance on a single site and context simplifies the problem of achieving a plausibly representative wage estimate,

[3] Alternatives to simpler averages are sometimes used: Rota and Weisdorf example, fit a local polynomial regression in their work on St Peter's in Rome and on rural workers in Tuscany (Rota and Weisdorf 2020, 2021). They also estimate rural wages using Clark's regression approach as a cross-check on their results (Rota and Weisdorf 2021, pp. 461-2).

at the price of not being generalizable to the regional or national level.

The second approach, estimating nominal wage values using regressions, was introduced by Robert Margo in his analysis of American wages and generalized to a wider audience by Greg Clark in his seminal paper on English day wages (Margo 2000; Clark 2005). It provided an econometric solution to the challenge of how to estimate a representative wage for a larger polity—usually a region or nation—from a diverse range of data sources. Regression analysis overcame the fragmented nature of the surviving sources, providing a formal solution to the problem of representativeness that otherwise constrained estimates from single sites or institutions.

It is Clark's models for building craftsmen and laborers' day wages (or labor costs) that have been most widely cited and emulated.[4] These include a tightly defined set of variables that allow for some variation between occupations, across space, and over time, which were used to predict log-transformed annual national day wages. As an example, the model for craftsmen includes categorical indicators for 29 crafts; an indicator if the wage was for a combination of workers, such as a master and servant; 50-year period dummies, interacted with region dummies for five regions; location fixed effects; and the year of observation (Clark 2005, p. 1322). Clark (2007b) conducts a similar analysis for agricultural day wages, averaging observations within each year, place, season and task, as well as including controls for threshing rates and winnowing, and for seasonality (Clark 2007b, p. 101). These models set the standard form for using regressions to generate a national annual metric from individual-level data.

Some sense of the influence of these methods and the importance of wage series relying on these methods can be obtained from citation counts. At the time of writing, Clark (2005) has over 700 citations, many from secondary users of his laborer and craftsman wage series. His series form the basis of the average weekly earnings reported in the Bank of England's "Three Millennium of Macroeconomic Data" whose associated bulletin has over 200 citations (Thomas et al. 2010).[5] The average weekly earnings series back to 1260 is readily available on FRED.[6]

While the discussion so far has focused on the use of these methods for wages, regressions have been used to generate other predicted series in economic history such as agricultural labor productivity in China (Zhai 2024), human heights from archeological long bones (Koepke and Baten 2005), prices and

[4]Clark (2005) acknowledged that many of the wages that he gathered were probably not directly paid, deflating nominal wages by 0.905 to account for this. Stephenson (2018) demonstrates that in the seventeenth and eighteenth centuries, the quoted day rates in institutional accounts may have overstated wages paid to actual workers 20-30%. See Appendix A for more details.

[5]Note that this dataset is often used without a reference in the bibliography, so this is an underestimate.

[6]Bank of England, Consumer Price Index in the United Kingdom [CPIUKA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIUKA

marginal productivity of enslaved workers in the Cape Colony (Du Plessis et al. 2015), currency risk premia during the classical gold standard (Mitchener and Weidenmier 2015), sale prices of paintings during the Italian Renaissance (Etro 2018), historical house price indices globally (Samy 2015; Karagedikli and Tunçer 2021; Raff et al. 2013), electric motor prices in Sweden (Edquist 2010), and wealth inequality in the Ottoman Empire (Coşgel and Ergene 2012). As in the case of wages, in these other examples, regressions are used to generate a more representative prediction of an outcome by controlling for various factors.

2.3 Challenges of regression curve-fitting to scarce data

Margo and Clark's method has become standard, as the long list of papers in Table 1 demonstrates, though the details of its application have varied. In this section, we highlight three connected issues with this regression methodology that affect the quality of the annual estimates that are generated: data scarcity, the extent to which additional controls have been incorporated, and the inflexibility of the approach. For each of these limitations, we briefly discuss how it will be addressed by our past predictive modeling framework.

First, and most broadly, data scarcity directly affects the reliability of estimates from regressions. In periods with less data available, fitted values become more like interpolations than estimates of nationally representative values. This risks bias or overfitting depending on how closely the model tracks the few data points available in a year or a decade (Hastie et al. 2009, pp. 21-27, 220-224) The resulting estimates of wages or other economic phenomena in the periods in which data are most scarce are less representative and reliable than estimates from the same model for years in which data are less scarce.

To address this overarching challenge, we shift from a curve-fitting estimation framework to a prediction framework that takes into account the time series nature of the data (Shmueli 2010). Privileging out-of-sample prediction accuracy allows us to generate more representative estimates of unseen data, especially in the periods in which data are most scarce.

Second, a tendency to expand the number of control variables has increased the risk of parameter instability. While the earliest wage regressions used relatively spartan models, subsequent work has often increased the number and variety of additional control variables. Variables have included the effects of, among other things, urban vs. rural settings, source biases, types of payment, duration of contract, and gender.

There are good reasons to expand the regression model to incorporate these factors, as we have might well expect them to impact wages. However, the inclusion of a larger set of controls in linear models presents a 'curse of dimensionality' when data are scarce that increases the risk of overfitting, causing unstable parameter estimates (Hastie et al. 2009, pp. 24–27). Evidence that this may be an issue can be seen in the measures of model fit reported in Table 1. Several have an R^2 in excess of 0.9 (Ridolfi 2019; Liu 2024).

To address this limitation, our past predictive modeling framework employs algorithms with regularization and dimensionality reduction. This allows for the inclusion of a rich set of controls while avoiding the challenges of high-dimensional estimation on limited data (Varian 2014). Our technique of using more recent data to predict historical data also ensures that the largest possible sample is used in each year's prediction.

One reason for the increase in control variables in recent papers has been to allow models to predict wages from datasets that include males and females, or skilled and unskilled workers, or multiple occupations at the same time, increasing the sample size as a result. This approach contrasts with Margo and Clark's separation of samples, models, and predictions for different occupational groups. The recent trend of estimating a single model and predicting separate occupational wage series from it follows a common logic in regression analysis. It does, however, build in assumptions about the consistency of the relationship between different control variables and occupational categories.

This leads us to a third issue: unless extensive interactions are incorporated, which is rarely the case, such an approach implicitly assumes that a control variable such as 'urban' has, for example, the same effect on skilled and unskilled workers alike, and that this effect remains relatively constant over time and space. This 'hedonic' aspect of curve-fitting approaches to scarce data, a necessary characteristic of inflexible linear models, has been strongly criticized by historians (Hatcher 2018; Nicholas and Oxley 1993).

Margo's initial discussion of wage regressions underlined this issue (Margo 2000). He highlighted the distinction between the bundle of job and worker characteristics, estimated through the coefficient β, and the time-period dummies in his model. He notes, "[i]n a less restrictive specification, β would be allowed to vary across time periods—ideally, for each time period. However, allowing β to vary over time greatly increases the number of coefficients to estimate, producing the trade-off noted above between sampling error and historical detail" (Margo 2000, p. 38). Margo deliberately chose to estimate his model separately for occupational groups and regions so that beta could vary between them, allowing for flexibility in the contribution of various factors to wages over time and between groups.

Past predictive modeling addresses these issues by parsimoniously generating estimates for different groups from the same model using machine learning algorithms which are more flexible than linear models. The increased flexibility of these non-parametric models allows the impact of variables on the prediction to vary over time and space.

The large body of existing research on long-run nominal wages has achieved a great deal with the current method. However, it faces difficult problems that are shared by other equivalent efforts to estimate long-run economic statistics: the unavoidable scarcity of data and the inherent rigidity of the estimation of relationships between variables achieved by least squares regression.

3 Methodology

To address these limitations of previous approaches to working with scarce historical data, we develop and implement a new framework, past predictive modeling, that generates robust out-of-sample predictions of historical economic series using machine learning strategies. In this section, we describe the core elements of this framework. The innovation is first to approach the problem of scarce data using a prediction framework; then, rather than predicting future values, to adapt these methods to predict historical values of economic series; and finally, to use machine learning algorithms to generate the best predictions. We first present the general framework, followed by specific implementations and estimation procedures.

3.1 General framework with sample splitting and "walk backward"

In the most general sense, we model economic outcomes $y_{i,t}$ using a simple predictive framework with error,

$$y_{i,t} = E_t(y_{i,t}) + \epsilon_{i,t},$$

where the conditional expectation of $y_{i,t}$ at time t is,

$$E_t(y_{i,t}) = g^*(x_{i,t}).$$

Our goal is to isolate a representation of $E_t(y_{i,t})$ as a function of predictor variables, given in the P-

dimensional vector $x_{i,t}$, that minimizes the out-of-sample prediction error for realized $y_{i,t}$. The function $g^*(\cdot)$ is a flexible function of $x_{i,t}$ that maps predictor variables to the outcome. We describe various algorithms for $g^*(\cdot)$ below.

Our past predictive modeling framework differs from standard predictive modeling by making predictions backwards in time rather than forwards. We adapt the expanding window forecasting approach, commonly used for forward-looking predictions, to address the historical data scarcity problem by reconstructing series backward in time. In this setting, we use the more recent data points as the "known" information to train a model that predicts historical economic values for periods with incomplete records. Just as in standard predictive modeling, we predict where there is the most uncertainty, but in historical settings, the greatest uncertainty is further in the past.

To maintain temporal consistency, we follow the most common approach in the literature and divide the sample into three disjoint periods: a training sample, a validation sample, and a testing sample. The training sample is the data used to estimate the model. Using the trained model, we then generate predictions for the validation sample and use these predictions to tune the hyperparameters of the model. We evaluate the prediction error in the validation sample and then iteratively search for hyperparameters that minimize the error in the validation sample, re-fitting the model to the training data under each set of hyperparameters. The validation sample is out-of-sample and temporally disjoint from the training data, so evaluating the prediction error of the validation sample allows us to optimize the model for out-of-sample performance on an unseen time period. Finally, we report measures of model fit for the testing sample, which was not used for model estimation nor for hyperparameter tuning. We then use the fully trained and validated model to make a forecast for the next year.

Where our method differs from the common expanding window "walk forward" approach—in which the training window grows year by year to make successive forward predictions—is that we instead "walk backward." In our framework, the training window expands in reverse, using all of the most recent data and incorporating older historical data at each step to predict values for earlier and earlier years.[7] In each iteration, we predict only a single year, generating a short-horizon forecast that captures historical trajectories without imposing patterns from one century onto another.

Figure 3 illustrates this framework for years $t = 0$ to $t = 9$. The first iteration gives the process that generates a prediction for $t = 3$. The training data includes years $t = 7$ through $t = 9$. The model is tuned

[7] We show in Appendix C that our results for England are also robust to using a variety of rolling window approaches.

using $t = 5$, the model fit is computed for $t = 4$, and the prediction is ultimately made for $t = 3$. $t = 6$ is left out to ensure there is no leakage between the training data and the testing data.

In the second iteration, we get the prediction for $t = 2$. Data from future years $t = 6$ to $t = 9$ are used to estimate the model that ultimately predicts the past value in $t = 2$. The training window expands back as the prediction is made further back in time, increasing the amount of data used to train the model. This walk-backward, expanding window process continues until each year in the sample has a prediction.[8]

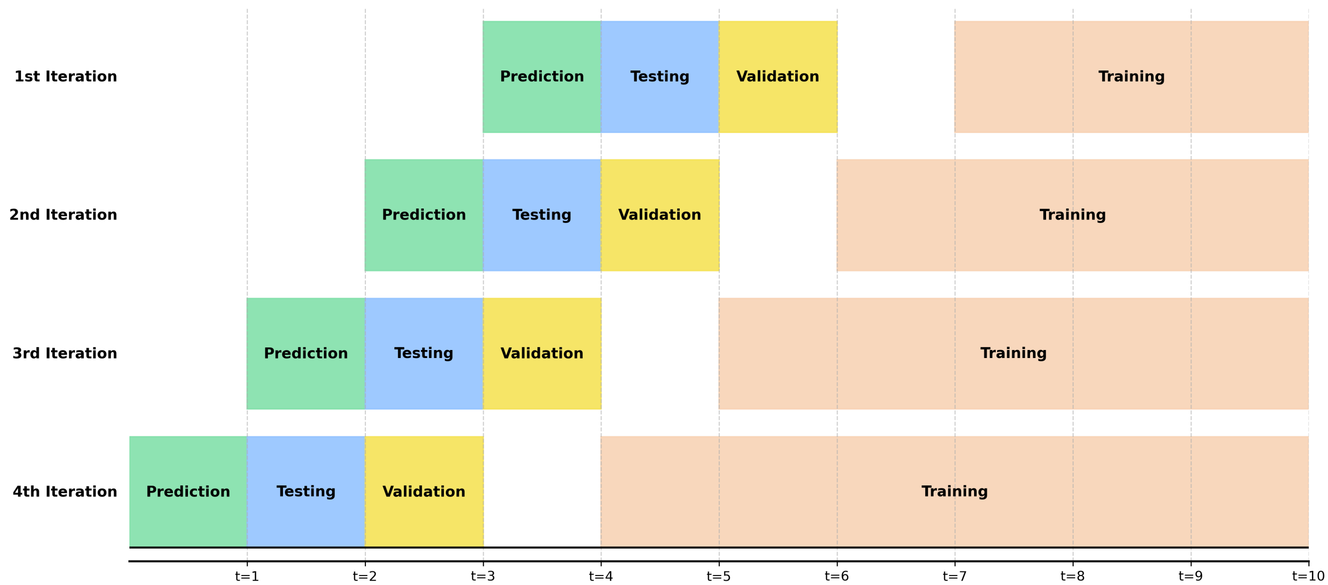

FIGURE 3: ILLUSTRATION OF WALK-BACKWARD EXPANDING WINDOW

This framework is more flexible than existing methods in two key aspects. First, it accommodates highly flexible, non-linear prediction algorithms through $g^*(\cdot)$, which can better capture complex and changing relationships between predictor variables and outcomes. Second, the walk-backward expanding window approach allows the model to change dynamically over time. By re-estimating $g^*(\cdot)$ each year, we relax the assumption built into the regression models discussed above that predictor variables have largely time-invariant impacts on the outcome. This allows our approach to more effectively capture structural changes and evolving relationships over extended historical periods. To balance this flexibility, we impose in our framework the restriction that the hyperparameters governing regularization of $g^*(\cdot)$ are stable over time. This constraint ensures that even as the model adapts dynamically, the fundamental model architecture is consistent, which ensures that our flexible estimates are still coherent across periods.

To determine these hyperparameters, we follow common practice and iteratively test combinations of

[8]Note that only original, raw data observations are in the training data for each annual prediction. Model-generated predictions are never used recursively to inform other predictions.

hyperparameters on the validation sample using a grid search to determine the combination that minimizes the out-of-sample prediction error. To find a combination of hyperparameters that works well across the full sample period, we set a fixed interval length and conduct the grid search once every interval. From the resulting best combinations of hyperparameters from each search, we select the modal combination.

The estimation method is not a black box. Although the re-estimation of $g^*(\cdot)$ means that the parameters connecting variables and outcomes change dynamically, they can be observed and interpreted. We are able to see how much different variables in the dataset are contributing to the estimation in each prediction window from the score assigned to each feature, which we discuss in more detail in Appendix D.

3.2 Model selection and evaluation

Our choice of machine learning algorithms for $g^*(\cdot)$ is based on the existing literature, which generally finds that gradient boosted decision trees and deep neural networks outperform linear regressions for prediction and forecasting problems. We compare all non-linear algorithms to a benchmark linear model with no regularization. We focus on three algorithms for our gradient boosted decision trees: XGBoost (Chen and Guestrin 2016), LightGBM (Ke et al. 2017), and CatBoost (Prokhorenkova et al. 2018). All three algorithms are known for performing well on prediction tasks in scenarios with complex and non-linear relationships by creating ensembles of decision trees and iteratively boosting to improve model performance. This is balanced by regularization to reduce overfitting. We also build four feed-forward neural networks with hidden layers ranging from two (shallow) to five (deep), all of which also incorporate regularization through ensembling, early stopping, and learning rate shrinkage.

Any of these algorithms for $g^*(\cdot)$ can be seamlessly integrated into our past predictive modeling framework. Central to the framework is the expanding window, walk-backward approach where more recent data are used to predict past data. The key is to correctly subset the data into disjoint training, validation, testing, and forecasting blocks and then expand the size of the training data in each step. Once these steps are in place, any algorithm that can predict continuous-valued outcomes from labeled data is appropriate to incorporate as $g^*(\cdot)$. Because the framework itself is algorithmically agnostic, it will be adaptable to future innovations in predictive algorithms.

In Appendix B, we test the various machine learning strategies described above to determine what is initially most successful with very-long-run economic data. To determine which algorithm makes the strongest out-of-sample predictions, we calculate the out-of-sample prediction error using the root mean squared error. This is defined as the square root of the average of the squared prediction error $(y_{i,t} - \hat{y}_{i,t})^2$ where $\hat{y}_{i,t}$ denotes our predicted value from the estimation of $g^*(\cdot)$ on unseen data. We find that LightGBM outperforms the benchmark linear model, the other gradient boosted decision trees, and the neural networks, so we use that algorithm throughout the paper.

3.3 Use cases and extensions

The past predictive modeling framework described above is designed to work best with scarce historical data for which the underlying observations are available at a regular frequency. This is because the method relies on consecutive observations for the split of the training, validation, testing, and prediction samples. The framework could be modified for data available at, e.g., decadal levels by simply indexing t in terms of decades.[9] If consistency between the timing of the training data and prediction data is not important in an application, t could also simply index the next non-missing year of data.

While the framework is designed for scarce data, it is important to keep in mind that if the training data are too limited, machine learning algorithms will not be able to produce reliable results. Given the backwards prediction structure, the sample size of the training data is most limited in the years closest to the present. One therefore must include years after the first prediction year to form the training data for the first prediction.[10] While there is no universal threshold for what constitutes an adequate sample size in machine learning applications, for tree based models, one guideline is to ensure that the training data includes at least the number of leaves times the minimum data in each leaf.

The backwards prediction structure also has a further implication that should be recognized: it cannot anticipate exogenous shocks that are not yet in its training data. As a result, the time at which a sudden shock becomes visible in the predictions is shifted by the four years it takes for it to enter the training sample. Analytically, this is easy to accommodate, but it needs to be kept in mind when interpreting short-run annual predictions.[11]

[9] In this case, to ensure that the training data are close to the prediction data, one might consider eliminating the space between the training and validation data, or eliminating the validation data entirely. This would, however, prevent any tuning of the hyperparameters or reporting of feature importance.

[10] For example, in our application to English wages below, we use data through 1914 and make our first prediction in 1900.

[11] We note that this does not seem to present a large issue in the application to English wages we explore. In Appendix F, we show that the correlation between changes in annual nominal wages and annual grain prices is stronger with past predictive modeling than with regression estimation.

4 Application

To empirically validate our framework, we apply past predictive modeling to the well-studied case of nominal wages in England over the course of seven centuries.

We leverage Clark's 47,000 observations of nominal wages paid to workers in the English building industry from 1200 to 1914 (Clark 2005)[12]. This dataset is the foundation of much influential research on the economic history of England over six centuries. As Humphries and Weisdorf (2019) and others have since pointed out, the degree to which these wages represent an average worker's income or real wages is limited because they omit women and workers on annual contracts, and because we have no accepted estimates of how many days were worked per year.[13] Nevertheless, the dataset remains one of the largest and best researched for any country in Europe.

Clark (2005) uses these observations to estimate a representative wage for laborers and craftsmen in each year in a linear regression model. We apply past predictive modeling to these data, using the variables in the original 2005 model as well as some information that was collected but not included, presumably because of the dimensionality issue inherent in the regression approach.[14]

We modify the raw data in three ways to improve the validity of the analysis. First, rather than constructing two separate samples for unskilled laborer and skilled craftsman wages, we incorporate all observations into a unified dataset with an indicator variable to distinguish among laborers, craftsmen, and assistants to craftsmen. This approach allows our estimates of, for instance, laborers' wages to be informed by broader wage patterns across occupations, taking full advantage of the enhanced flexibility of machine learning algorithms to allow for these spillovers.

Second, we refine the classification of occupations into craft types based on the original occupations to ensure that all wages pertain exclusively to building workers and that the correct occupations are grouped together.[15] This results in nineteen types of building craft categories.

[12]We are exceptionally grateful to Gregory Clark for sharing the raw data underlying the regression models with us. To the best of our knowledge, these data have not been used in projects other than in Clark's own work.

[13]Our approach is agnositc to how to adjust for these concerns.

[14]The additional variables are: whether the worker was an assistant, a laborer, or a craftsman; county identifiers; whether the wage was paid by the day, hour, or week; whether the worker was male or female; the season of the year; and whether the wage was complemented by food or other provisions.

[15]Specifically, we drop coopers, furbishers, millwrights, boat builders, shipwrights, blacksmiths, locksmiths, smiths, wheelwrights, and whitesmiths as these occupations are not traditionally considered to be in the building industry. We also combine various names for thatchers, for plasterers, and for roofers into three craft categories. The remaining nineteen categories of crafts include carpenters, joiners, or wrights; bricklayers; laborers (e.g. carry earth, make wells); carvers; daubers (e.g. clayers, wattle); glaziers; tilers; lath layers (e.g, rending lath, lathier, lathing); masons or stonecutters; painters; pavers; plasterers or

Finally, we address duplicates in the data, where multiple workers receiving identical wages in the same place for the same occupation in the same year were entered on separate rows. We consolidate these rows in order to not overstate our sample size, combining repeat observations into single row of the data and generating a new variable that captures the count of observations of each wage-year-place combination. After these modifications, we are left with 39,223 total observations.

We also employ regional population data for our analysis. When working with nominal wages, researchers have typically reported separate wage series for laborers (unskilled wages) and craftsmen (skilled wages). These estimates are meant to be taken as 'national' estimates representative of average wage levels across England. We improve on this approach by presenting population-weighted averages of regional wage estimates in order to generate a more representative 'national' estimate. To achieve this, we build upon established population data for the counties of England (Wrigley 2007, 2009; Broadberry et al. 2015, table 8B).[16] This enables us to fit values to population levels as they changed over time, improving the accuracy of a 'national' wage estimate.

5 Validation results

This section assesses the extent to which our past predictive modeling framework improves on conventional regression-based approaches using our case study of English nominal wages. We evaluate performance along two dimensions. First, we compute bootstrap standard errors to compare the precision of estimates generated by the past predictive modeling framework and traditional linear regression. The results indicate a substantial reduction in uncertainty, with past predictive modeling yielding bootstrap standard errors that are on average 60.2% lower than those from linear regressions. Second, we evaluate predictive performance out-of-sample, comparing the accuracy of machine learning algorithms to that of linear models within the past predictive modeling framework. We conservatively find that the past predictive modeling framework implemented using LightGBM improves out-of-sample accuracy by more than 28%.

pargetters; plumbers; roofers (e.g. slaters, heliers); sawyers; reeders (e.g. thatchers or arundinators); waller; whitelimers (e.g. whitewashers); and other craftsmen.

[16]Our regional population estimates are constructed as follows. We first use linear interpolation to estimate the share of the national population in each county between the benchmarks estimates given in Broadberry et al. (2015) and Wrigley (2007, 2009). We then estimate county populations by assigning them their share of the national population estimates (Broadberry et al. 2015). We then aggregate the counties into the regions used in Clark (2005).

5.1 Reduction of uncertainty

Does the past predictive modeling framework yield more precise estimates than traditional in-sample regression-based approaches? As discussed in Section 2.3, when data are scarce, regression models are prone to overfitting. This leads to unstable predictions that poorly reflect underlying population parameters and are less generalizable to unseen data. In theory, using appropriate predictive modeling techniques that optimize for out-of-sample performance and account for the temporal nature of the data should address this problem and improve precision. Is this borne out in practice?

Using our case study of nominal wages in England, we test whether the predicted values from our past predictive modeling approach are more stable and precise than those from traditional regression methods. To do this, we generate bootstrap standard errors of the average estimate for each year using both methods. First, we replicate the sample and models for laborers and craftsmen from Clark (2005) and confirm that we generate the same annual average wage estimates. Then, we estimate the bootstrap standard error by creating 100 bootstrapped samples, resampling with replacement, and generating a new prediction for each year from each bootstrapped sample.[17] A lower standard deviation of these predictions indicates a narrower confidence interval and thus a more precise and stable set of predictions.

Second, we calculate the bootstrap standard error for the predictions from the past predictive modeling framework, implemented with LightGBM.[18] We follow the same procedure but within the past predictive modeling framework, where predictions are generated separately for each year, making the process more computationally intensive. To calculate the bootstrap standard errors, as before, we generate the standard deviation of the prediction across 100 bootstrap samples for each year.

Figure 4 presents the results of the bootstrap analysis for both estimation approaches, separately for laborers (unskilled) and craftsmen (skilled), with standard errors scaled by the mean and plotted as decadal averages. The blue line depicts the bootstrap standard errors from the replicated estimates of Clark (2005), while the orange line corresponds to the standard errors generated by the past predictive modeling framework implemented using LightGBM. A higher standard error indicates more uncertainty in the prediction. For both laborers and craftsmen, the past predictive modeling framework produces much more precise estimates in almost every decade. On average, the standard errors are reduced with the new framework by 60.2% for laborers and by 69.0% for craftsmen. These substantial reductions in

[17] If the sample size is N, each bootstrap sample is size N. Then, each observation has a $\frac{1}{N}$ chance of being drawn so a $1 - \frac{1}{N}$ chance of never being drawn. Over the course of N draws, the probability of an observation not being drawn is $(1 - \frac{1}{N})^N$

[18] This decision is discussed below.

the standard error suggest that the regression-based models are overfitting to the limited historical data, resulting in greater variability in predictions under resampling. The past predictive modeling framework addresses this overfitting to generate more robust and stable estimates of annual wages.

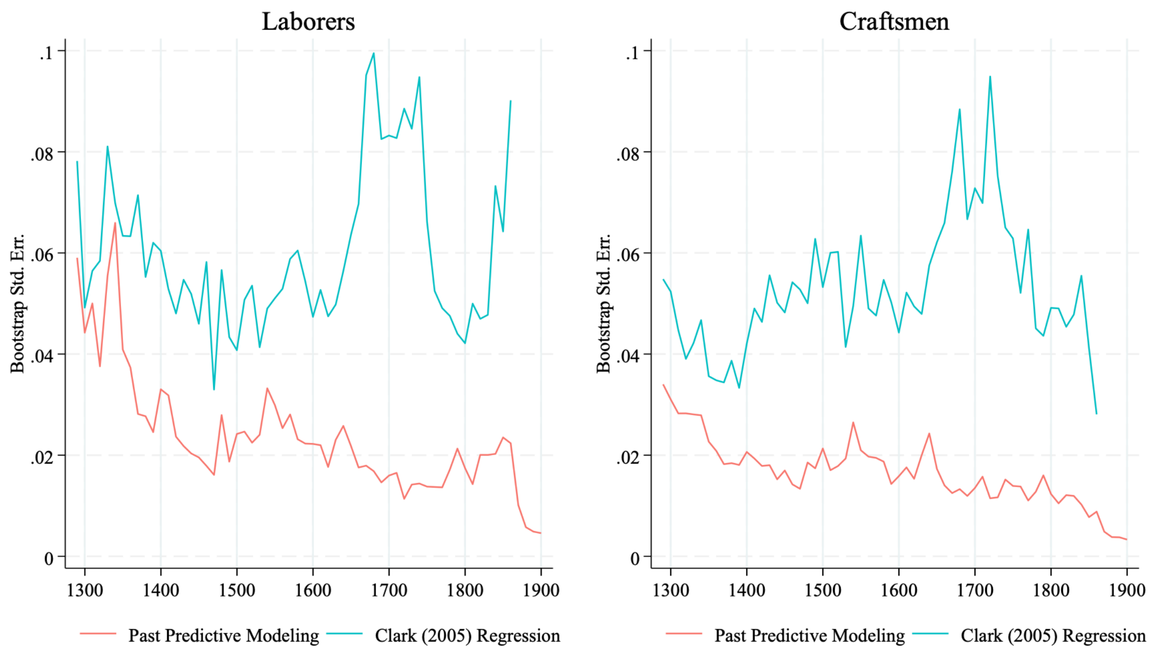

FIGURE 4: BOOTSTRAP STANDARD ERRORS, REGRESSION VS. PAST PREDICTIVE MODELING

The instability of wage estimates from regression models, particularly in the presence of scarce data, makes any given prediction less representative of the true population parameter and therefore less generalizable to unseen data. A key innovation of the past predictive modeling approach is its prioritization of out-of-sample model fit, explicitly taking the temporal nature of the data into account by using an expanding window approach. This fundamental difference in perspective increases the ability of the model to generate stable, generalizable estimates that better reflect the true underlying wage distribution.

We conduct three additional robustness checks. First, we check that our framework presents these advantages outside of the English nominal wages setting. In Appendix E, we replicate this bootstrap analysis using the data from Kumon (2022) for servant wages in Japan 1610–1890. Our implementation of the past predictive modeling framework with LightGBM reduces bootstrap standard errors by over 40%.[19] This large reduction in uncertainty in a distinct setting with limited data indicates that the usefulness of this method is not just confined to the case of English nominal wages.

[19] We implement LightGBM with just its default hyperparameters. With hyperparameter tuning, it is likely this reduction in uncertainty could be even greater.

Second, in Appendix F we relate the laborers' wages from the Clark (2005) regression and from our new past predictive modeling framework to grain prices for the years 1300-1600. Wage prices and grain prices should be related through both income and in wage effects for a period in which agriculture is the dominant sector (Broadberry et al. 2015) and consumption baskets are dominated by grain-based foods (Clark 2005, p. 1328). We show that changes in wage estimates from the past predictive modeling framework are slightly more correlated with changes in wheat prices than the wage estimates from Clark (2005). Also in Appendix F, our third check shows that when the sample size in a given prediction year is smaller, our estimates diverge more from those in Clark (2005), suggesting that, as expected, our improvements are coming from years in which data are more scarce.

5.2 Improvements in accuracy

The previous section shows that the past predictive modeling framework generates more precise estimates than traditional regression-based methods by explicitly accounting for temporal dependencies in the data using an expanding window "walk backward" approach. An additional benefit of this shift in methodology is that it enables the use of machine learning algorithms that optimize predictive power. This can lead to more accurate out-of-sample predictions than linear regressions.

As we described in Section 3, any algorithm for $g^*(\cdot)$ can be integrated into the past predictive modeling framework, so we evaluate three gradient boosted decision trees and four neural networks in Appendix B. Our metric for evaluating which of these algorithms makes the strongest out-of-sample predictions is the root mean squared error for the testing sample. A lower root mean squared error indicates that our algorithm better predicts unseen data. The out-of-sample prediction error is particularly salient in the context of scarce data, where, by definition, the true value of some economic outcome is unseen. Optimizing out-of-sample predictive power reduces the risk of bias or overfitting that can arise from scarce or noisy data, generating the most representative and accurate prediction. We therefore select the algorithm that, for our specific case, produces the lowest average root mean squared error across all years.

As Appendix B shows, when compared to a benchmark linear model implemented within the past predictive modeling "walk-backward" framework, LightGBM is 28% more accurate out-of-sample.[20] This demonstrates that the past predictive modeling approach not only improves on existing methods in terms

[20] This estimate of the improvement in accuracy to unseen out-of-sample data is necessarily an underestimate. In order to evaluate out-of-sample, out-of-time accuracy, the linear model has to be implemented in a walk-backward, time-aware framework. There are likely further gains to accuracy that arise just from using the framework which this estimate cannot capture.

of precision, but that these more generalizable estimates are also more accurate to unseen data.

One final, fundamental point that needs to be acknowledged is that, irrespective of method, the true population value we wish to predict is unknowable, made unrecoverable by the choices and accidents that shape what data from hundreds of years ago survive today. While it is impossible to recover data that were never recorded, in this setting, prioritizing algorithms that minimize out-of-sample prediction error offers a pragmatic alternative. This approach is superior to running a linear regression on the full sample, which assumes that the observed data are representative and that true structural parameters can be recovered. Instead, we seek the most representative approximation of unrecorded economic outcomes using the data we do observe by focusing on out-of-sample predictive power. Yet while we can confidently state that past predictive modeling improves out-of-sample accuracy relative to a linear model, it remains impossible to quantify the extent of that improvement relative to the true, but unrecoverable, population value.

6 The economic benefits of past predictive modeling

More accurate predictions of economic indicators have an obvious appeal, but does this added statistical precision have meaningful benefits for economic analysis? In this section, we examine the differences between the predictions that the past predictive modeling framework produces and those obtained from the existing best-in-class regression methodology in order to evaluate whether they are significant enough to justify adopting this new approach.

The broad trends of our predicted long-run wage series for 1200 to 1900 are reported in Figure 5, with 95% confidence intervals around 9-year moving averages. The benefits of the past predictive modeling framework can be demonstrated usefully through three examples that explore some applications of these new wage estimates.

First, we show that the enhanced precision of predictions in periods of scarce data provides new insights into the effect of the Black Death on income inequality. Second, we show how the flexibility of parameter estimates in this approach allows regional wage variations to be documented, generating a deeper understanding of the economic geography of industrializing England. Third, we exploit these new regional series to produce nationally representative weighted wage estimates that remove major biases introduced by the changing spatial composition of the data, a frequent issue with historical datasets.

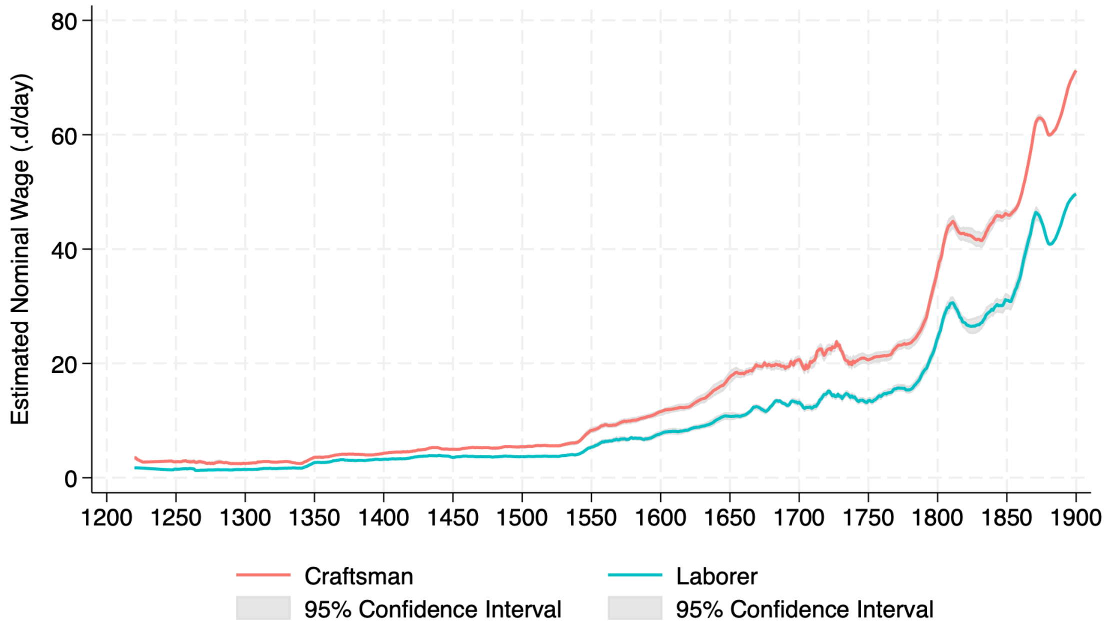

FIGURE 5: PAST PREDICTIVE MODELING WAGE ESTIMATES OF CRAFTSMEN AND LABORER'S WAGES, 1200-1900

Finally, we examine how the new wage series affect estimates of productivity growth over the long run, revising our understanding of developments before 1600.

6.1 The Black Death and inequality: predictions with limited observations

Meaningful differences between the wage estimates generated by the past predictive modeling approach and regression emerge in several periods. The past predictive modeling approach has particular strengths in generating predictions in settings where data are sparse. We illustrate this through the impact of a defining economic shock, the Black Death of 1348 that killed somewhere between a third and a half of the English population.

The Black Death was the largest single labor market shock in recorded history. Figures 7 reports our predicted wages and Clark's estimates in the decades around the Black Death (Clark 2005).[21] Focusing on wage differences across the period after shock, we are able to exploit the advantages of these predictions. Most notably, we find a more rapid, larger and more sustained increase in unskilled laborer's wages in the years immediately following the plague. Laborers' nominal wages increased by almost 70% between the early 1340s and the mid-1350s, compared to 45% in Clark's estimation. Our predictions for skilled

[21] Following Clark, to make these comparisons we scale our predicted wages by 0.905, which is his adjustment to reflect that most day wages were not directly paid to workers. See Appendix A for more on these issues.

craftsmen's wages are consistently lower than Clark's in this period, but the rate and size of their wage increase after the Black Death was more similar to his series, with our predicted wages increasing by 48% and Clark's by 39% over the same period.

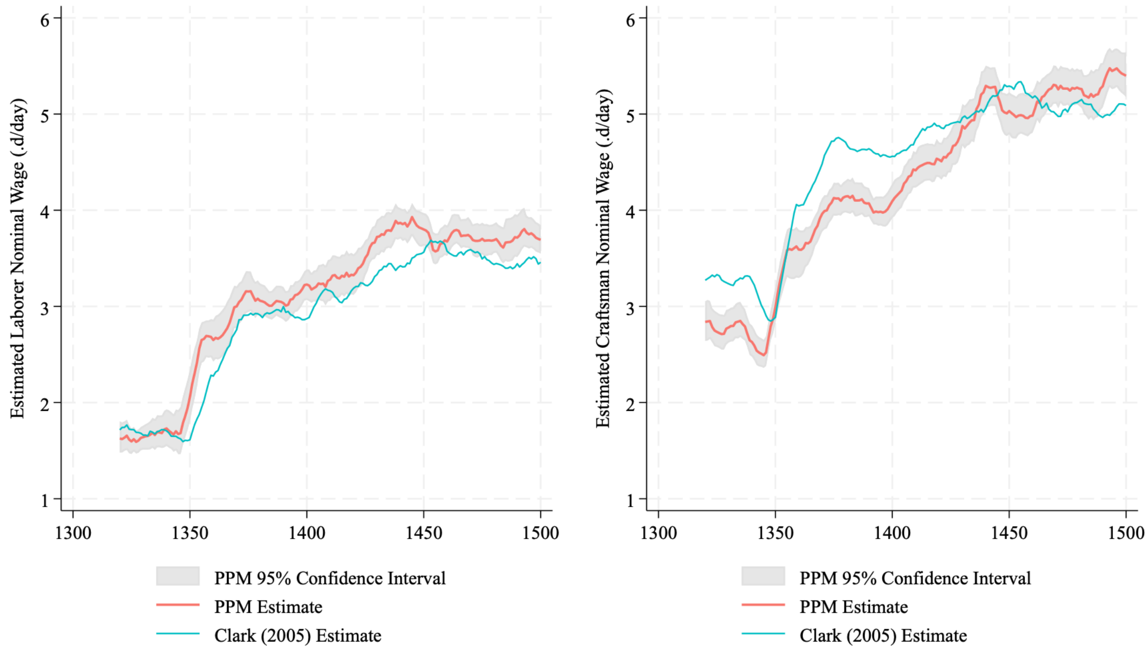

FIGURE 6: COMPARISON OF PAST PREDICTIVE MODELING WAGE ESTIMATE WITH EXISTING SERIES FROM CLARK (2005), 1320-1500

The differential impact of the Black Death that we see in these new wage series pushed the skill premium—the ratio between the average unskilled and average skilled wage—down to its lowest-ever level in these decades, as Figure 7 shows. This is substantially below other estimates (Van Zanden 2009; Jensen and Luo 2024), implying much lower returns to skill in the late fourteenth and early fifteenth century. Where existing series show stability from around 1400, we observe the skill premium steadily recovering until the early sixteenth century, when population growth began to accelerate, and, perhaps more importantly, real interest rates fell (Schmelzing 2020). For the next two hundred years, the skill premium moved around 60%, with peaks and troughs in periods of economic and political crisis such as the 1590s and the civil war.

The divergence we observe between skilled and unskilled wages after the Black Death indicate that the impact of the epidemic was economically differentiated. Unskilled laborers on casual day work contracts, with an outside option in agricultural work, benefited most from this dramatic Malthusian shock. The lower elasticity of skilled wages is plausibly explained by the state's intervention in the labor market,

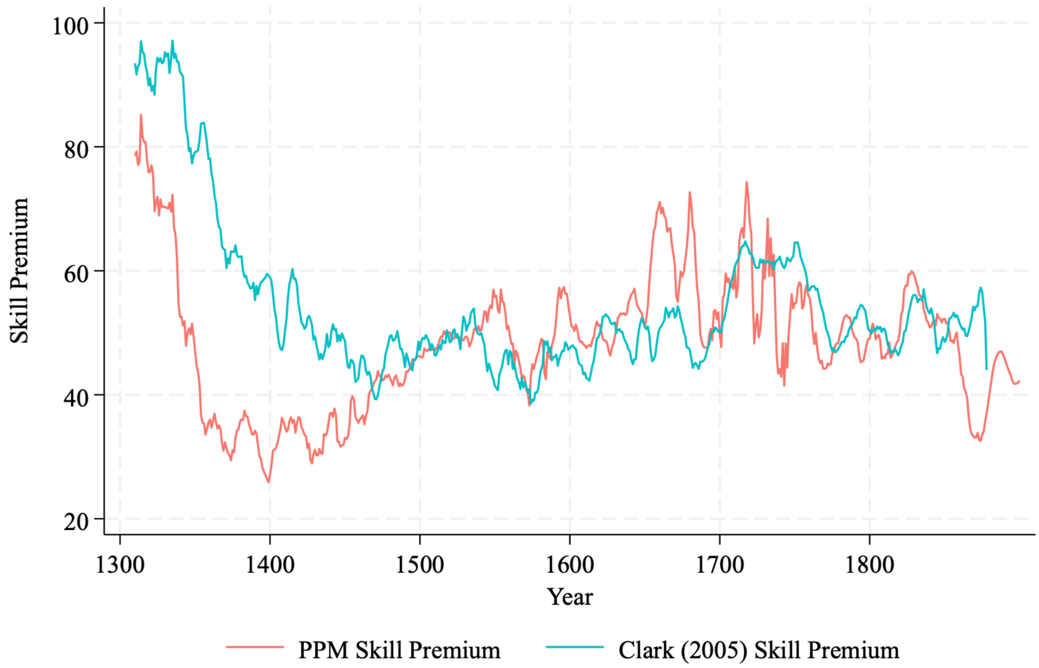

FIGURE 7: COMPARISON OF SKILL PREMIUM WITH EXISTING SERIES FROM CLARK (2005), 1320-1500

as it introduced novel legal constraints on wages through the 'Ordinance of Labourers' (1349) (Cohn 2007).[22] If the Black Death's impact on the capital-labor ratio generated a major reduction in economic inequality, reducing the power of the landed elite (Scheidel 2018; Alfani 2021), these results suggest that the epidemic also lowered inequality *within* the waged labor force, revealing a new mechanism by which epidemics impact inequality.[23]

6.2 Regional growth patterns before the Industrial Revolution

One of the major benefits of past predictive modeling is its flexibility. In most linear regression specifications, it is difficult to identify shifts in paramaters. Coefficients on variables are estimated for the full sample, or vary only intermittently between restricted sets of interactions. To give a concrete example from the setting we are exploring, regression models struggle to capture shifts in inter-regional wage ratios over time. When estimated in a single model, regional effects are usually fixed or, at best, estimated for a limited number of pre-determined periods. When linear models are estimated separately for different regions and then aggregated, inter-regional spillovers are missed as wage dynamics in one region

[22] Annual wages adjusted even more slowly to the Black Death (Claridge et al. 2024; Humphries and Weisdorf 2019, p. 2874)
[23] Evidence for an equalizing effect in agricultural wages suggests a similar trend in the larger rural workforce (Claridge et al. 2025, p. 38)

cannot influence another.

The greater flexibility of machine learning can address these issues, allowing the identification of relative changes in regional wage levels while still parsimoniously estimating a single model. This enhanced ability to examine spatial variation (or similar variation in other parameters) is one of the major advantages of past predictive modeling.

In our application, English nominal wages, we use this flexibility to predict regional wages based on time-varying region effects. This allows us to examine regional differences in wage costs, a feature of premodern labor markets that has been identified as major cause of industrialization via directed technical change (Allen 2009). Our new predicted regional wage series offer a novel insight into the economic geography of England in the centuries before the first Industrial Revolution.

Figure 8 presents the first consistent, long-run predictions of regional wages in England.[24] They provide new evidence on the timing of two key phases of economic development that have only been sketched from patchy urbanization data previously. The first is the sixteenth and seventeenth century divergence between London and the rest of the nation that was primarily driven by international trade (Wrigley 1967; Zahedieh 2010; Acemoglu et al. 2005). This started in the mid-sixteenth century and involved a doubling of nominal wages in London by 1650 compared to much slower wage growth elsewhere in England. By the 1670s, the 'London premium' relative to northern wages was around 100% (Woodward 1995). These high London wages were then sustained for almost a century The second is the later eighteenth-century process of convergence, as productivity growth surged in the industrializing north due to the increasing returns to skill and human capital in newly mechanised industries. The wage evidence suggests this did not begin until the 1770s, and the 'great reversal' in wages brought about by industrialization (Kelly et al. 2023, 70) did not fully emerge until after 1800.

These new regional wage series underline the distinctive dominance of London as the economic center of skills, training, and labor demand in England throughout the seventeenth and eighteenth centuries. England's 'high wage' economy was limited to the capital until the Industrial Revolution was well under way.

[24] See Appendix G for comparisons with alternative shorter-run measures.

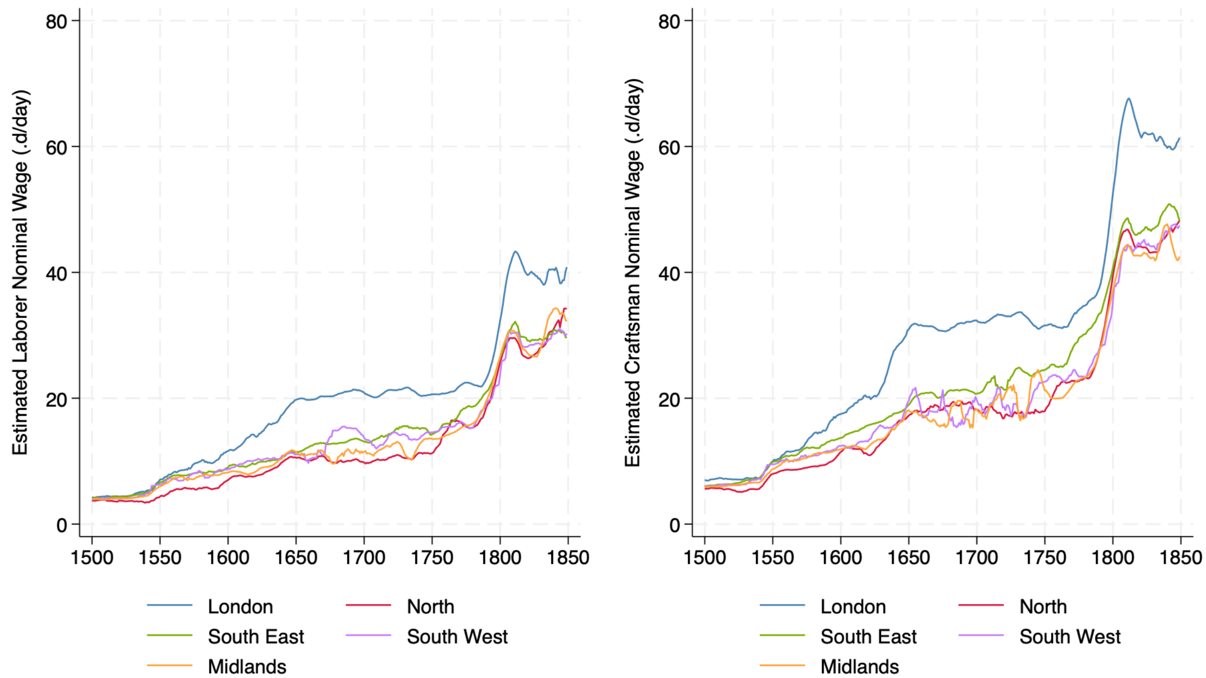

FIGURE 8: REGIONAL WAGE PATTERNS, 1500 TO 1840

6.3 Constructing representative national wage series

Regional wage series have a further important advantage. As we discuss in Section 4, in combination with population data we can use them to estimate a population-weighted national wage.

The large relative changes in regional wage levels presented in Section 6.2 have economically meaningful implications for national wage estimates. We illustrate this in Figure 9, which presents the new past predictive model wage series for building laborers and craftsmen. Each panel reports two series that presented as nine-year moving averages. The unweighted series is the average wage prediction in each year from our model. The weighted series is a population-weighted average of the regional predictions.[25] The weighted series is more representative of national wage trends and is the approach taken by modern national statistical agencies.

For most periods, the two series are similar, as we would expect given the empirical richness of Clark's original data. However, the weighted series and the unweighted series diverge substantially from around 1660 until the middle of the eighteenth century due to the divergence between regional wage levels and the large share of observations from high-wage London. Both the weighted series show significantly

[25] Appendix H discusses the weighting method fully.

lower national wages during this time than the unweighted series, with large differences of 15-20% for some decades. In the absence of weighting, the series show a marked rise after 1660 followed by a substantial fall in the 1740s. In the weighted series, both laborers and craftsmen's wages grow much more slowly after 1660 before entering a period of sustaining growth from around 1760. There is no evidence of any change in the trend in nominal labor income after the Glorious Revolution (1688), or of a fall in nominal wages during the mid-eighteenth century when demographic pressure was pressing on output growth, suggesting the presence of substantial downward wage stickiness.

In cases of scarce data, where some regions may be better represented in the extant data than others, weighting regional estimates by population generates a more nationally-representative estimate than an unweighted average. The flexibility of past predictive modeling permits the use of aggregation strategies that correct for the existence of potentially large biases.

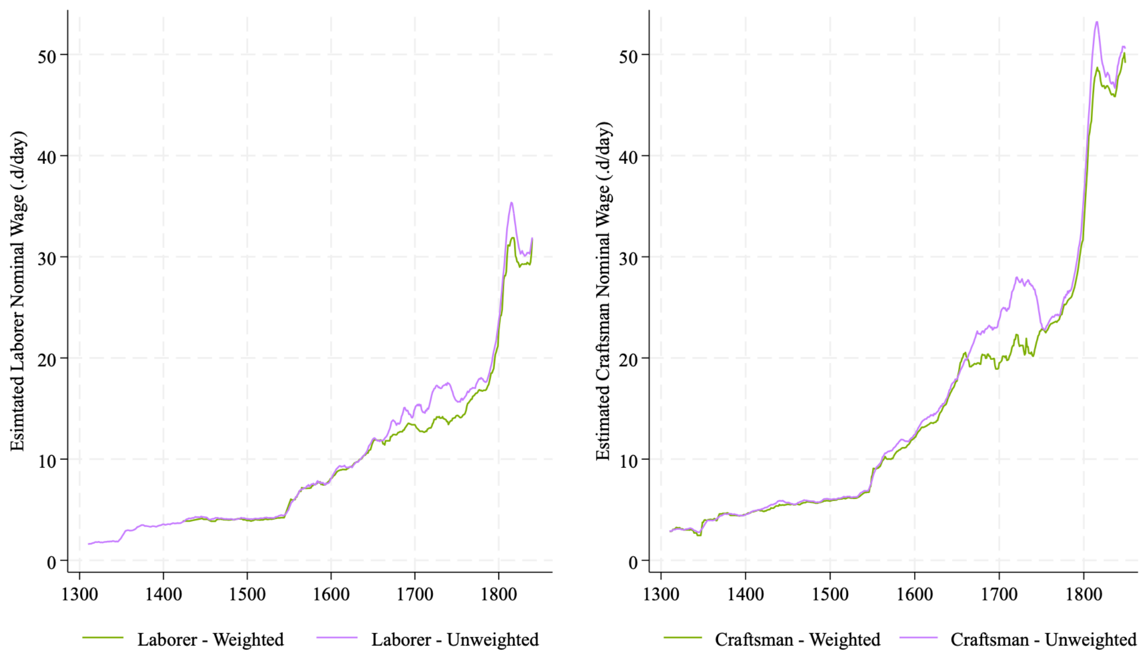

FIGURE 9: WEIGHTED AND UNWEIGHTED WAGES IN D. PER DAY

6.4 Productivity growth in England

As a final illustration of the implications of using the past predictive modeling framework, we revisit Bouscasse, Nakamura and Steinsson's recent analysis of productivity and growth in England over the long run (Bouscasse et al. 2025). Their paper estimates productivity growth in England between 1250 and 1870, providing important new evidence for when productivity growth began and the rate of growth from 1600-1800.

Wages are a critical ingredient in their analysis, which derives estimates of productivity from the labor demand curve. They choose this approach because real wages (and population) "are arguably among the best measured series of all economic time series over our long sample period" (Bouscasse et al. 2025, 838). They estimate productivity in two ways. First, they use the Black Death as an exogenous shock, and exploit the change in real wages that accompanied the fall in population. Second, they structurally estimate a Malthusian model using population and wages across six centuries. The data they use for their main estimations are Clark's series for unskilled building workers (Clark 2005). We replaced these with our wage predictions, deflated in the same way, as explained in Appendix I.

Reassuringly, we produce essentially unchanged results for the two major results in Bouscasse et al. (2025): the breakpoint from which persistent productivity growth is observed remains around 1600, and the rates of growth in the period before and after this are similar.

We do, however, find suggestive evidence that one of the surprising features of medieval economic history that Bouscasse et al. believed had disappeared with their analysis did in fact exist. This is the rise and fall of productivity between the Black Death and the middle of the sixteenth century. While Bouscasse et al. saw little sign of this in their results, the productivity rates obtained with the new wage estimates shows a significant increase by the 1450s, as Figure 10 shows. These estimates are still slightly below those produced by Clark (2016) using a dual approach technique to estimate TFP, and Allen's estimates for agricultural productivity (Allen 2005).[26]

How to explain this rise and fall in productivity has stood as a challenge since Clark identified it. That it appears again in these new estimates suggests that there were real changes in productivity during these centuries, perhaps reflecting the countervailing impact of improvements in market integration and political turmoil on the English economy, and this deserves further research.

More broadly, this application demonstrates that the improvements brought by past predictive modeling impact not just the estimated series, but also secondary analyses that use those estimations as a core component.

[26]Figure 10 plots the baseline model from Bouscasse et al. (2025). In Appendix I, we show that this result is robust to both the simple and structural models. We also show the full productivity breakpoint results.

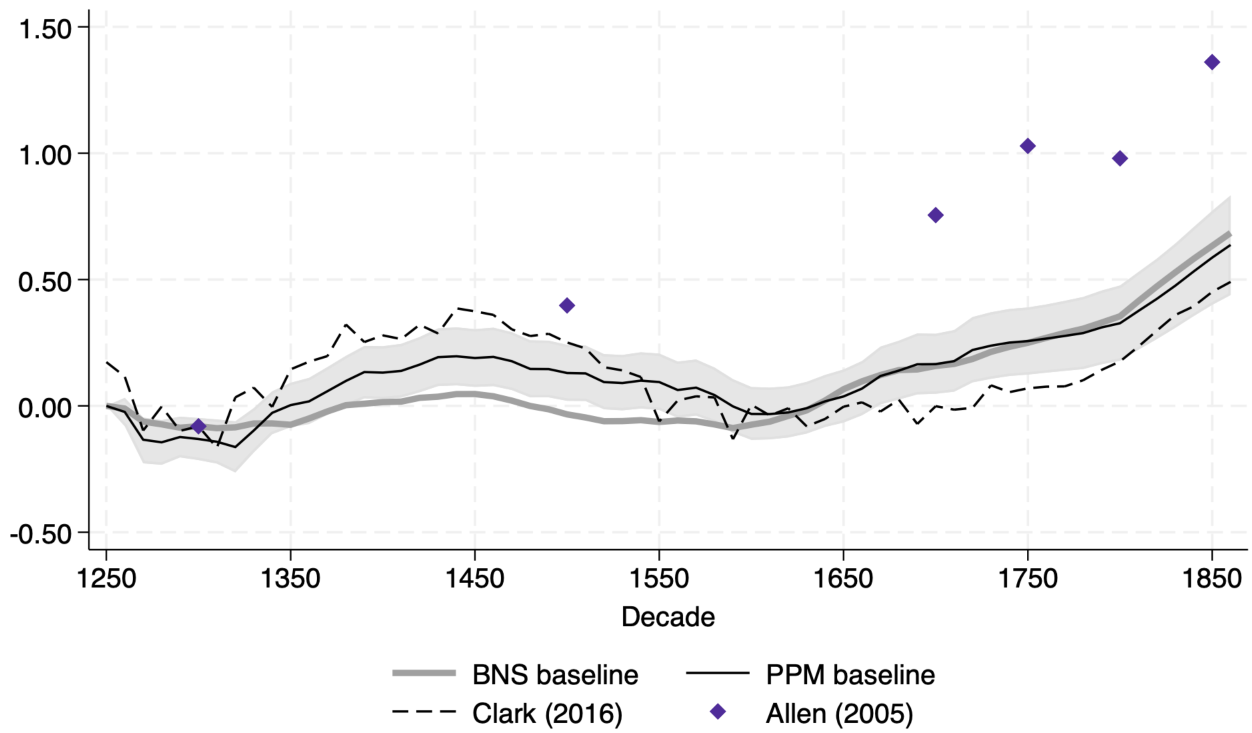

FIGURE 10: ESTIMATES OF PRODUCTIVITY IN ENGLAND
Note: Each series is the natural log of productivity, as reported in (Bouscasse et al. 2025, Fig. I).

7 Conclusion

Economists interested in long-run analyses must contend with the fragmentary and incomplete nature of historical data. This paper introduces past predictive modeling, a framework that leverages machine learning and out-of-sample predictive modeling techniques to address this challenge of reconstructing economic time series from scarce historical data. We improve on conventional methods that fit regression curves to limited data, which risk bias and overfitting, by using a prediction framework that optimizes on out-of-sample prediction accuracy.

Applying this framework to English nominal wages from 1300 to 1900, we demonstrate that this past predictive modeling approach yields substantial improvements in precision and accuracy of predicted values. Relative to benchmark linear regressions, past predictive modeling reduces bootstrap standard errors by 60.2% and out-of-sample root mean squared errors by 28.4%. These performance gains arise from leveraging well-established techniques for addressing bias and overfitting and using more flexible machine learning algorithms.

Beyond just improvements in accuracy, our new wage estimates shape our understanding of long-run economic development. We find that the Black Death had a skill-differentiated impact on wage rates, leading to a much lower skill premium than previously recognized. By estimating novel long-run regional wage series and weighting by population, we illuminate regional economic dynamics prior to the Industrial Revolution and also show the importance of population weighting for national wage estimates to be representative. Finally, our new estimates reconfirm the existence of a medieval productivity puzzle after the Black Death.

The broader implications of our method extend to development economics, macroeconomics, and economic history, where reliable historical data are critical to our understanding of economic transitions and long-term growth yet difficult to come by. While we have chosen to use nominal wages as our case study, the framework we outline here can be applied to a wide range of historical economic variables, opening up new avenues for future research and shifting fundamentally our approach to scarce data.

References

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum, and S. Pérez (2021). Automated Linking of Historical Data. *Journal of Economic Literature 59*(3), 865–918.

Acemoglu, D., S. Johnson, and J. Robinson (2005). The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth. *American Economic Review 95*(3), 546–579.

Alfani, G. (2021). Economic Inequality in Preindustrial Times: Europe and Beyond. *Journal of Economic Literature 59*(1), 3–44.

Allen, R. C. (2001). The Great Divergence in European Wages and Prices from the Middle Ages to the First World War. *Explorations in Economic History 38*(4), 411–447.

Allen, R. C. (2005). English and Welsh Agriculture, 1300-1850: Output, Inputs, and Income. *Working Paper*.

Allen, R. C. (2009). *The British industrial revolution in global perspective*. Cambridge: Cambridge University Press.

Allen, R. C. (2015). The high wage economy and the industrial revolution: a restatement. *The Economic History Review 68*(1), 1–22.

Allen, R. C., J. Bassino, D. Ma, C. Moll-Murata, and J. L. Van Zanden (2011). Wages, prices, and living standards in China, 1738–1925: in comparison with Europe, Japan, and India. *The Economic History Review 64*(s1), 8–38.

Allen, R. C. and E. Khaustova (2019). Russian real wages before and after 1917. *Explorations in Economic History 72*, 23–37.

Allen, R. C. and J. L. Weisdorf (2011). Was there an 'industrious revolution' before the industrial revolution? An empirical exercise for England, c. 1300-1830. *The Economic History Review 64*(3), 715–729.

Bajari, P., D. Nekipelov, S. P. Ryan, and M. Yang (2015). Machine Learning Methods for Demand Estimation. *American Economic Review 105*(5), 481–485.

Baulant, M. (1971). Le salaire des ouvriers du bâtiment à Paris de 1400 à 1726. *Annales. Histoire, Sciences Sociales 26*(2), 463–483.

Bluwstein, K., M. Buckmann, A. Joseph, S. Kapadia, and Şimşek (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics 145*, 103773.

Bouscasse, P., E. Nakamura, and J. Steinsson (2025). When Did Growth Begin? New Estimates of Productivity Growth in England from 1250 to 1870. *The Quarterly Journal of Economics 140*(2), 835–888.

Bowley, A. L. (1900). *Wages in the United Kingdom in the nineteenth century: notes for the use of students of social and economic questions*. Cambridge: Cambridge University Press.

Broadberry, S., B. M. Campbell, A. Klein, M. Overton, and B. Van Leeuwen (2015). *British Economic Growth, 1270-1870*. Cambridge, UK: Cambridge University Press.

Buckles, K., A. Haws, J. Price, and H. E. B. Wilbert (2024). Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project. *NBER Working Papers* (31671).

Buscemi, T. (2025). From the little divergence to the little divide: Real wages in the Kingdom of Sicily (1540–1850). *The Economic History Review 78*(2), 646–672.

Carvalhal, H., J. Lucassen, and P. De Zwart (2024). After da Gama: real wages in Western India, c. 1500–c. 1650. *European Review of Economic History 28*(3), 311–334.

Chalfin, A., O. Danieli, A. Hillis, Z. Jelveh, M. Luca, J. Ludwig, and S. Mullainathan (2016). Productivity and Selection of Human Capital with Machine Learning. *American Economic Review 106*(5), 124–127.

Chambru, C. and P. Maneuvrier-Hervieu (2023). The evolution of wages in early modern Normandy (1600–1850). *The Economic History Review 76*(3), 917–940.

Chen, T. and C. Guestrin (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Claridge, J., V. Delabastita, and S. Gibbs (2024). (In-kind) Wages and labour relations in the Middle Ages: It's not (all) about the money. *Explorations in Economic History 94*(101626).

Claridge, J., V. Delabastita, and S. Gibbs (2025). The commercialization of labour markets: Evidence from wage inequality in the Middle Ages. *LSE Economic History Working Papers* (375).

Clark, G. (2005). The Condition of the Working Class in England, 1209–2004. *Journal of Political Economy 113*(6), 1307–1340.

Clark, G. (2007a). *A Farewell to Alms*. Princeton, NJ: Princeton University Press.

Clark, G. (2007b). The long march of history: Farm wages, population, and economic growth, England 1209–1869. *The Economic History Review 60*(1), 97–135.

Clark, G. (2016). Microbes and Markets: Was the Black Death an Economic Revolution? *Journal of Demographic Economics 82*(2), 139–165.

Cohn, S. (2007). After the Black Death: labour legislation and attitudes towards labour in late-medieval western Europe. *The Economic History Review 60*(3), 457–485.

Coşgel, M. M. and B. A. Ergene (2012). Inequality of Wealth in the Ottoman Empire: War, Weather, and Long-Term Trends in Eighteenth-Century Kastamonu. *The Journal of Economic History 72*(2), 308–331.

Crafts, N. (2021). Understanding productivity growth in the industrial revolution†. *The Economic History Review 74*(2), 309–338.

Cvrcek, T. (2013). Wages, Prices, and Living Standards in the Habsburg Empire, 1827–1910. *The Journal of Economic History 73*(1), 1–37.

de Vries, J. (2008). *The Industrious Revolution: Consumer Behavior and the Household Economy, 1650 to Present*. Cambridge: Cambridge University Press.

De Zwart, P. (2011). South African Living Standards in Global Perspective, 1835–1910. *Economic History of Developing Regions 26*(1), 49–74.

De Zwart, P. and J. Lucassen (2020). Poverty or prosperity in northern India? New evidence on real wages, 1590s–1870s. *The Economic History Review 73*(3), 644–667.

De Zwart, P. and J. L. Van Zanden (2015). Labor, wages, and living standards in Java, 1680–1914. *European Review of Economic History 19*(3), 215–234.

Dell, M. (2025). Deep Learning for Economists. *Journal of Economic Literature 63*(1), 5–58.

Dell, M., J. Carlson, T. Bryan, E. Silcock, A. Arora, Z. Shen, L. D'Amico-Wong, Q. Le, P. Querubin, and L. Heldring (2023). American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers. *37th Conference on Neural Information Processing Systems*.

Du Plessis, S., A. Jansen, and D. Von Fintel (2015). Slave prices and productivity at the Cape of Good Hope from 1700 to 1725: Did everyone win from the trade? *Cliometrica 9*(3), 289–330.

Edquist, H. (2010). Does hedonic price indexing change our interpretation of economic history? Evidence from Swedish electrification. *The Economic History Review 63*(2), 500–523.

Etro, F. (2018). The Economics of Renaissance Art. *The Journal of Economic History 78*(2), 500–538.

Federico, G., A. Nuvolari, L. Ridolfi, and M. Vasta (2021). The race between the snail and the tortoise: skill premium and early industrialization in Italy (1861–1913). *Cliometrica 15*(1), 1–42.

Feinstein, C. H. (1998). Pessimism Perpetuated: Real Wages and the Standard of Living in Britain during and after the Industrial Revolution. *The Journal of Economic History 58*(3), 625–658.

Ferguson-Cradler, G. (2023). Narrative and computational text analysis in business and economic history. *Scandinavian Economic History Review 71*(2), 103–127.

Frankema, E. and M. V. Waijenburg (2012). Structural Impediments to African Growth? New Evidence from Real Wages in British Africa, 1880–1965. *The Journal of Economic History 72*(4), 895–926.

Gary, K. E. and M. Olsson (2020). Men at work. Wages and industriousness in southern Sweden 1500–1850. *Scandinavian Economic History Review 68*(2), 112–128.

Gilboy, E. W. (1934). *Wages in Eighteenth Century England*. Cambridge, MA: Harvard University Press.

Goulet Coulombe, P., M. Leroux, D. Stevanovic, and S. Surprenant (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics 37*(5), 920–964.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies 33*(5), 2223–2273.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Element of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer Series in Statistics.

Hatcher, J. (2018). Seven Centuries of Unreal Wages. In *Seven Centuries of Unreal Wages*, Palgrave Studies in Economic History, pp. 15–69. Cham: Palgrave Macmillan.

Helgertz, J., J. Price, J. Wellington, K. J. Thompson, S. Ruggles, and C. A. Fitch (2022). A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 55*(1), 12–29.

Horrell, S. and J. Humphries (2019). Children's work and wages in Britain, 1280–1860. *Explorations in Economic History 73*(101272).

Humphries, J. and J. Weisdorf (2019). Unreal Wages? Real Income and Economic Growth in England, 1260–1850. *The Economic Journal 129*(623), 2867–2887.

Hunt, E. H. (1986). Industrialization and Regional Inequality: Wages in Britain, 1760–1914. *The Journal of Economic History 46*(4), 935–966.

Karagedikli, G. and A. C. Tunçer (2021). House prices in the Ottoman Empire: evidence from eighteenth-century Edirne. *The Economic History Review 74*(1), 6–33.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *31st Conference on Neural Information Processing Systems (NIPS)*.

Kelly, M., J. Mokyr, and C. Ó Gráda (2023). The Mechanics of the Industrial Revolution. *Journal of Political Economy 131*(1), 59–94.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015). Prediction Policy Problems. *American Economic Review 105*(5), 491–495.

Koepke, N. and J. Baten (2005). The biological standard of living in Europe during the last two millennia. *European Review of Economic History 9*(1), 61–95.

Kumon, Y. (2022). The Labor-Intensive Path: Wages, Incomes, and the Work Year in Japan, 1610–1890. *The Journal of Economic History 82*(2), 368–402.

Liu, Z. (2024). Wages, labour markets, and living standards in China, 1530–1840. *Explorations in Economic History 92*, 101569.

López Losa, E. and S. P. Zarauz (2021). Spanish subsistence wages and the Little Divergence in Europe, 1500–1800. *European Review of Economic History 25*(1), 59–84.

Margo, R. A. (2000). *Wages and Labor Markets in the United States, 1820-1860*. University of Chicago Press.

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys 37*(1), 76–111.

Mijatović, B. and B. Milanović (2021). The real urban wage in an agricultural economy without landless farmers: Serbia, 1862–1910. *The Economic History Review 74*(2), 424–448.

Mitchener, K. J. and M. D. Weidenmier (2015). Was the Classical Gold Standard Credible on the Periphery? Evidence from Currency Risk. *The Journal of Economic History 75*(2), 479–511.

Mocarelli, L. (2004). Wages and the Labour Market in the Building Trade in 18th Century Milan. *Jahrbuch für Wirtschaftsgeschichte / Economic History Yearbook 45*(2), 61–84.

Mullainathan, S. and J. Spiess (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives 31*(2), 87–106.

Nicholas, S. and D. Oxley (1993). The Living Standards of Women during the Industrial Revolution, 1795-1820. *The Economic History Review 46*(4), 723–749.

Paker, M. M., J. Z. Stephenson, and P. Wallis (2023). Job Tenure and Unskilled Workers before the Industrial Revolution: St Paul's Cathedral 1672–1748. *The Journal of Economic History 83*(4), 1101–1137.

Paker, M. M., J. Z. Stephenson, and P. Wallis (2025). Nominal wage patterns, monopsony, and labour market power in early modern England. *The Economic History Review 78*(1), 179–206.

Pfister, U. (2017). The timing and pattern of real wage divergence in pre-industrial Europe: evidence from Germany, c. 1500–1850. *The Economic History Review 70*(3), 701–729.

Phelps Brown, E. H. and S. V. Hopkins (1955). Seven Centuries of Building Wages. *Economica 22*(87), 195–206.

Price, J., K. Buckles, J. Van Leeuwen, and I. Riley (2021). Combining family history and machine learning to link historical records: The Census Tree data set. *Explorations in Economic History 80*, 101391.

Prokhorenkova, L., G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin (2018). CatBoost: unbiased boosting with categorical features. *32nd Conference on Neural Information Processing Systems (NeurIPS)*.

Raff, D., S. Wachter, and S. Yan (2013). Real estate prices in Beijing, 1644 to 1840. *Explorations in Economic History 50*(3), 368–386.

Richardson, A., T. Van Florenstein Mulder, and T. Vehbi (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting 37*(2), 941–948.

Ridolfi, L. (2019). Six Centuries of Real Wages in France from Louis IX to Napoleon III: 1250–1860. *The Journal of Economic History 79*(3), 589–627.

Rota, M. and J. Weisdorf (2020). Italy and the Little Divergence in Wages and Prices: New Data, New Results. *The Journal of Economic History 80*(4), 931–960.

Rota, M. and J. Weisdorf (2021). Italy and the little divergence in wages and prices: evidence from stable employment in rural areas. *The Economic History Review 74*(2), 449–470.

Samy, L. (2015). Indices of House Prices and Rent Prices of Residential Property in London, 1895-1939. *University of Oxford Discussion Papers in Economic and Social History* (135).

Scheidel, W. (2018). *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*. Princeton: Princeton University Press.

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science 25*(3).

Stephenson, J. Z. (2018). 'Real' wages? Contractors, workers, and pay in London building trades, 1650-1800. *The Economic History Review 71*(1), 106–132.

Stephenson, J. Z. (2019). Mistaken wages: the cost of labour in the early modern English economy, a reply to Robert C. Allen. *The Economic History Review 72*(2), 755–769.

Stephenson, J. Z. (2020). *Contracts and Pay: Work in London Construction 1660–1785*. Palgrave Studies in Economic History. Cham: Springer International Publishing.

Thomas, R., S. Hills, and N. Dimsdale (2010). The UK recession in context — what do three centuries of data tell us? *Bank of England Quarterly Bulletin* (Q4).

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives 28*(2), 3–28.

Wallis, P., J. Colson, and D. Chilosi (2018). Structural Change and Economic Growth in the British Economy before the Industrial Revolution, 1500–1800. *The Journal of Economic History 78*(3), 862–903.

Woodward, D. M. (1995). *Men at work: labourers and building craftsmen in the towns of northern England, 1450-1750*. Cambridge: Cambridge University Press.

Wrigley, E. A. (1967). A Simple Model of London's Importance in Changing English Society and Economy 1650-1750. *Past and Present 37*(1), 44–70.

Wrigley, E. A. (2007). English county populations in the later eighteenth century. *The Economic History Review 60*(1), 35–69.

Wrigley, E. A. (2009). Rickman revisited: the population growth rates of English counties in the early modern period. *The Economic History Review 62*(3), 711–735.

Yoon, J. (2021). Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. *Computational Economics 57*(1), 247–265.

Zahedieh, N. (2010). *The Capital and the Colonies: London and the Atlantic Economy, 1660-1700*. Cambridge: Cambridge University Press.

Zhai, R. (2024). Toward The Great Divergence: Agricultural Growth In The Yangzi Delta, 1393–1953. *SSRN Electronic Journal* (No. 4838585).

Appendices

A Labor costs or wages

The series of day wage observations in Clark's dataset are the amounts reported in original historical records as paid by institutions to craftsmen or building firms for labour costs (Clark 2005, p. 1321). Most economic historians working with long-run series of construction wages have interpreted these as 'day wages' paid to workers; however, as Clark points out, before 1860 or so, the amounts actually received by the workers would have been less than these payments. This is because institutions typically contracted with master contractors or 'firms' for different types of work, and account books show that the masters or contractors deducted their overhead costs from the billed amounts before paying labour (Stephenson 2018).

In order to convert this series into a nominal day wage series that represents the wage received by workers, not the cost billed to the client, it is necessary to deduct the contractor's overhead or margin from the recorded values. The value of the difference between the 'day rate' in the institutional account and the wage paid to the worker will have varied depending on the value of the contract to the contractor, the costs of credit, other supply chain considerations, seasonality, stage dependency, and so on. Clark (2005) deducts 9.5 percent from the bill or recorded amount. Stephenson (2018) finds that for the period 1650–1800 in England the difference lay somewhere between 20 and 30 percent. Deductions for other places will have been subject to local institutional arrangements (López Losa and Zarauz 2021; Baulant 1971; Mocarelli 2004; Stephenson 2019, pp. 765-6).

To then move from day wages to estimate average annual male wage income, it is necessary to establish or estimate a number of days that workers will have worked per year. It is generally accepted that this will have varied (de Vries 2008; Broadberry et al. 2015), and it will have been a function of investment in and demand for building services, seasonality, project specificity, and outside options for labor (Stephenson 2020). Estimates of the annual number of days worked come from a variety of sources (see description in Broadberry et al. 2015, p. 264) including working records from building sites (Stephenson 2020; Paker et al. 2023; Woodward 1995) and the costs of consumption goods (Allen and Weisdorf 2011; Humphries and Weisdorf 2019). In this paper, the series that we present are for day rates and are only directly comparable to nominal day wages. We do not make assumptions or estimations about the days worked per year, so therefore we do not predict annual real wages or living standards.

The predicted 'wage' series for England 1300–1900 from this paper should therefore be understood as the cost of labor less 9.5 percent. We deliberately follow Clark on this to enable direct comparisons to his series, as we say in footnote 21, and for all comparisons we scale our predicted wages by the same 0.905 factor.

B Comparison of machine learning algorithms

In this appendix, we test the effectiveness of three gradient-boosted decision tree algorithms and four neural networks, pitting them against a benchmark linear model estimated within the framework.

We focus on three algorithms for our gradient boosted decision trees: XGBoost (Chen and Guestrin 2016), LightGBM (Ke et al. 2017), and CatBoost (Prokhorenkova et al. 2018). All three algorithms are known for performing well on prediction tasks in scenarios with complex and non-linear relationships by creating ensembles of decision trees and iteratively boosting to improve model performance. This is balanced by regularization to reduce overfitting. The methods differ in their optimization: XGBoost uses first and second-order gradient descent, LightGBM uses histograms to approximate first-order gradients more efficiently, and CatBoost computes gradients on dynamically updated subsets of the data.

For each of the gradient boosted decision trees, model and tree complexity hyperparameters are tuned using the validation data to minimize out-of-sample prediction error. To determine these hyperparameters, we follow common practice and iteratively test combinations of hyperparameters on the validation sample using a grid search to determine the combination that minimizes the out-of-sample prediction error. To find a combination of hyperparameters that works well across the full sample period, we set a fixed interval length and conduct the grid search once every interval. From the resulting best combinations of hyperparameters from each search, we select the modal combination.

When applied to our case study of English nominal wages, Table A1 gives the best hyperparameters for XGBoost, Table A2 for CatBoost, and Table A3 for LightGBM.

We also build four feed-forward neural networks with hidden layers ranging from two (shallow) to five (deep). It is not common practice to tune neural networks using a hyperparameter grid owing to their computational complexity, so we follow Gu et al. (2020) in designing the architecture of our neural networks with fixed parameters. The number of neurons is selected according to the geometric pyramid

TABLE A1: XGBOOST TUNING PARAMETERS

Parameter	Value
objective	reg:squarederror
eval_metric	rmse
max_depth	10
min_child_weight	1
eta	0.3
n_estimators	200
colsample_bytree	1
alpha	0
reg_lambda	0

TABLE A2: CATBOOST TUNING PARAMETERS

Parameter	Value
loss_function	RMSE
iterations	500
learning_rate	0.1
depth	6
l2_leaf_reg	3
min_data_in_leaf	5

rule and all layers are fully connected. The activation function is Leaky ReLU, and we use an ADAM optimizer with a learning rate of 0.001 for 200 epochs with a batch size of 128 observations. Following Gu et al. (2020), we incorporate three regularization techniques into the network architecture. We use the learning rate shrinking algorithm with factor 5 if validation metrics plateau and incorporate early stopping. For each estimate, we also construct five separate networks and take the average to ensure the results are not biased by one-off variation in the random seeds.

Our metric for evaluating which of these algorithms makes the strongest out-of-sample predictions is the root mean squared error. This is calculated for the testing sample, which is not used in the training or tuning of any of the machine learning algorithms. A lower root mean squared error indicates that our algorithm better predicts unseen data. The out-of-sample prediction error is particularly salient in the context of scarce data, where, by definition, the true value of some economic outcome is unseen. Optimizing out-of-sample predictive power reduces the risk of overfitting that can arise from scarce or noisy data, generating the most representative and accurate prediction. We therefore select the algorithm that, for our specific case, produces the lowest average root mean squared error across all years.

Figure A1 presents the average root mean squared error across all years from each of the eight models

TABLE A3: LIGHTGBM TUNING PARAMETERS

Parameter	Value
`objective`	regression
`metric`	rmse
`max_depth`	10
`num_leaves`	31
`learning_rate`	0.2
`n_estimators`	500
`min_data_in_leaf`	10
`lambda_l1`	1
`feature_fraction`	0.9

applied to the English nominal wage data: the benchmark linear model (LM), XGBoost (XGB), LightGBM (LGB), CatBoost (CATB), and feed-forward neural networks with 2-5 hidden layers (FFNN2-5).

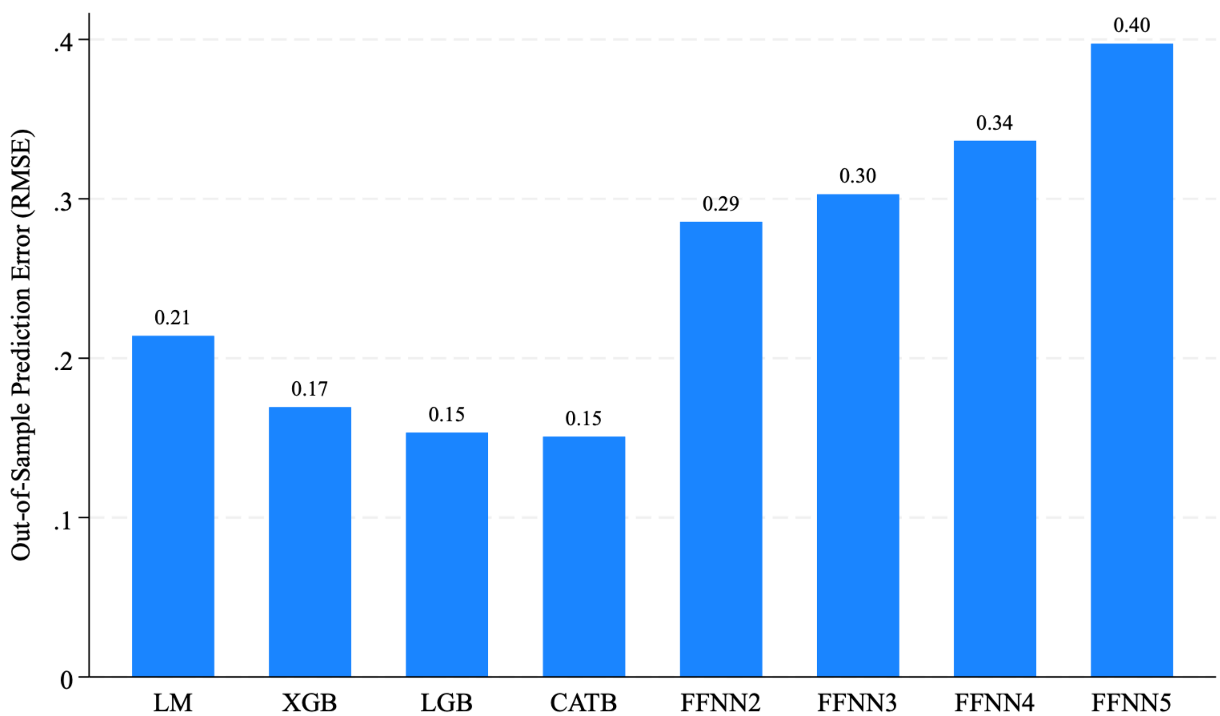

FIGURE A1: OUT-OF-SAMPLE ROOT MEAN SQUARED ERROR, LINEAR REGRESSION AND MACHINE LEARNING MODELS

Figure A1 shows that LightGBM and CatBoost generate the most accurate out-of-sample predictions, while the neural networks generate the least accurate predictions. This accords with the literature which generally finds that gradient boosted decision trees have the greatest prediction power in cases with many categorical features over long run time series. Compared to the benchmark linear model estimated within the framework, LightGBM reduces the prediction error by 28.36%.

Our results in Figure A1 show that CatBoost marginally outperforms LightGBM in terms of reducing the root mean squared error. Because this difference is negligible, we choose to use LightGBM in our subsequent tests of whether the more accurate predictions from the past predictive modeling framework are meaningfully different from existing estimates because it is faster and easier to work with.

C Robustness to rolling window approaches

As an additional robustness check, we run the model with 50-year, 100-year, 250-year, and 500-year rolling windows instead of an expanding window. A rolling window uses the same number of years in each model estimation and is another standard technique for time series prediction. In our case, this limits the data used in the estimation to a fixed number of years following the year for which the prediction is being generated.

Figure A2 shows that our results are essentially identical regardless of whether an expanding window or rolling window approach is used, for both laborers and craftsmen.

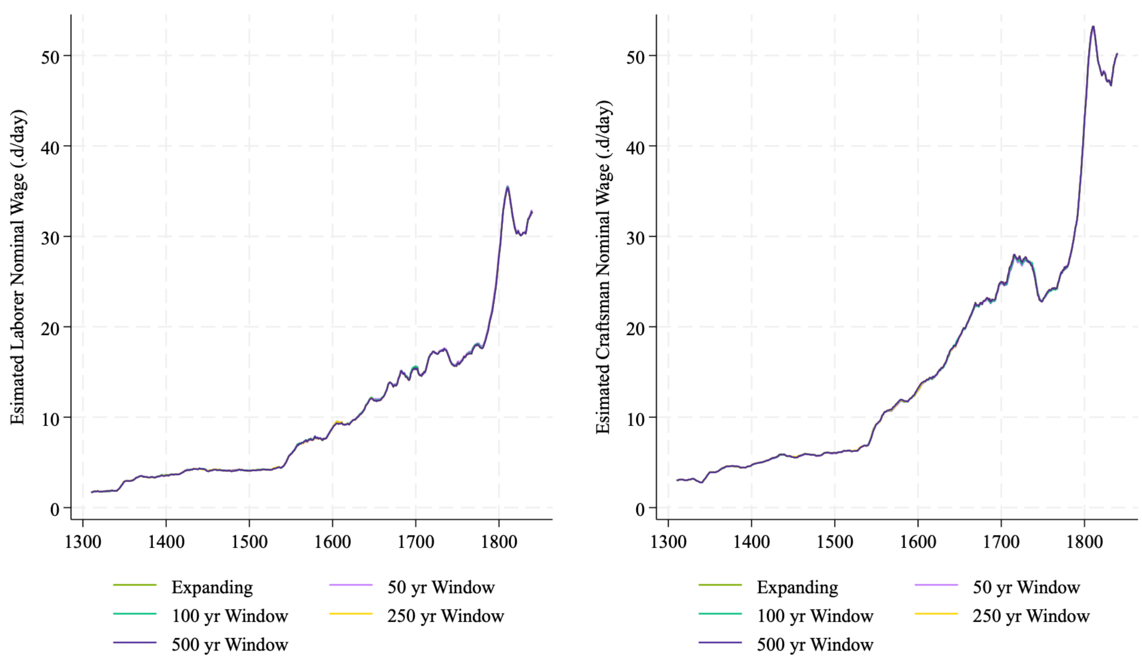

FIGURE A2: EXPANDING WINDOWS VS. ROLLING WINDOWS FOR LABORERS AND CRAFTSMEN

Table A4 shows that the average bootstrap standard errors are very similar between the different approaches. The expanding window provides the lowest average standard error for the laborers estimate, while the 100-year rolling window provides the lowest average standard error for the craftsmen estimate.

Note that all of the approaches give a lower standard error than the linear regression in Figure 1.

TABLE A4: AVERAGE BOOTSTRAP STANDARD ERRORS, EXPANDING VS. ROLLING WINDOW

	Laborers	Craftsmen
Expanding Window (baseline)	0.229	0.239
50-yr Rolling Window	0.241	0.241
100-yr Rolling Window	0.234	0.234
250-yr Rolling Window	0.230	0.237
500-yr Rolling Window	0.230	0.236

D Feature importance in the predictive model

One advantage of the past predictive modeling framework is that it allows us to evalute the contribution each variable ("feature") makes to the model's predictions. We calculate a permutation importance score for all the features in the model, giving us information about the contribution that each makes to the wage prediction for each year. The permutation importance score is a measure of how much the model's performance drops if the values of a single feature are randomly shuffled. It has an appealingly obvious intuition. If shuffling the values does not affect performance much, then that feature is not making a large contribution to the prediction, and vice versa.

In our application to English wages, we report permutation importance scores for each year, giving the contribution of each feature to the wage prediction for that year, and then express this as a percentage. As a result, we are able to decompose the importance of individual features, such as an area like the county of Kent, and of different categories of feature, such as counties in general. We can also observe how the importance of different features changes over the period being studied.

Table 1 reports the average percentage score of each category of feature across all 653 models estimated from 1220 to 1900 (for some years in the thirteenth century where wage observations are missing no model can be estimated), and the minimum and maximum share contributed by each category.

The type of work involved—whether the wage is paid to a craftsman or laborer—has the largest average permutation importance score. This is entirely plausible given the large difference in the sums paid to each, captured in the skill premium estimate discussed in Section 6.1. We also see that Region, and to a lesser degree County, also have relatively large scores, which is consistent with the substantial spatial variation in wages across England. Most other features have smaller permutation importance scores on

TABLE A5: SUMMARY OF FEATURE IMPORTANCE ACROSS ALL YEARS

Feature		**Category**	
	Mean (%)	Max (%)	Min. (%)
CountObs	0.16	6.36	0.00
County	8.83	46.81	0.00
Craft	6.53	50.72	0.00
Female	0.51	17.52	0.00
Food	3.05	56.72	0.00
FoodMissing	0.46	7.98	0.00
Joint	0.12	0.13	0.12
JointDummy	1.19	1.21	1.15
Region	17.47	57.79	0.00
Season	0.42	5.67	0.00
Source	4.91	36.13	0.00
Status	3.82	46.39	0.00
Type	52.23	92.26	0.00
WageType	0.29	7.37	0.00

average, but occasionally have larger effects in specific years.

Together, the three most important parameters account for, on average, 78% of the permutation importance score across these predictions. The model derives significant information from the multiple other pieces of information it can incorporate into its estimation process.

The contribution of different features evolves over time. In Figure A3, we illustrate the shifting share of the prediction score supplied by different categories of feature. The contribution of the Wage Type rises and then declines over time, while Region becomes visibly more important in the second half of our period, as divergence within England becomes greater.

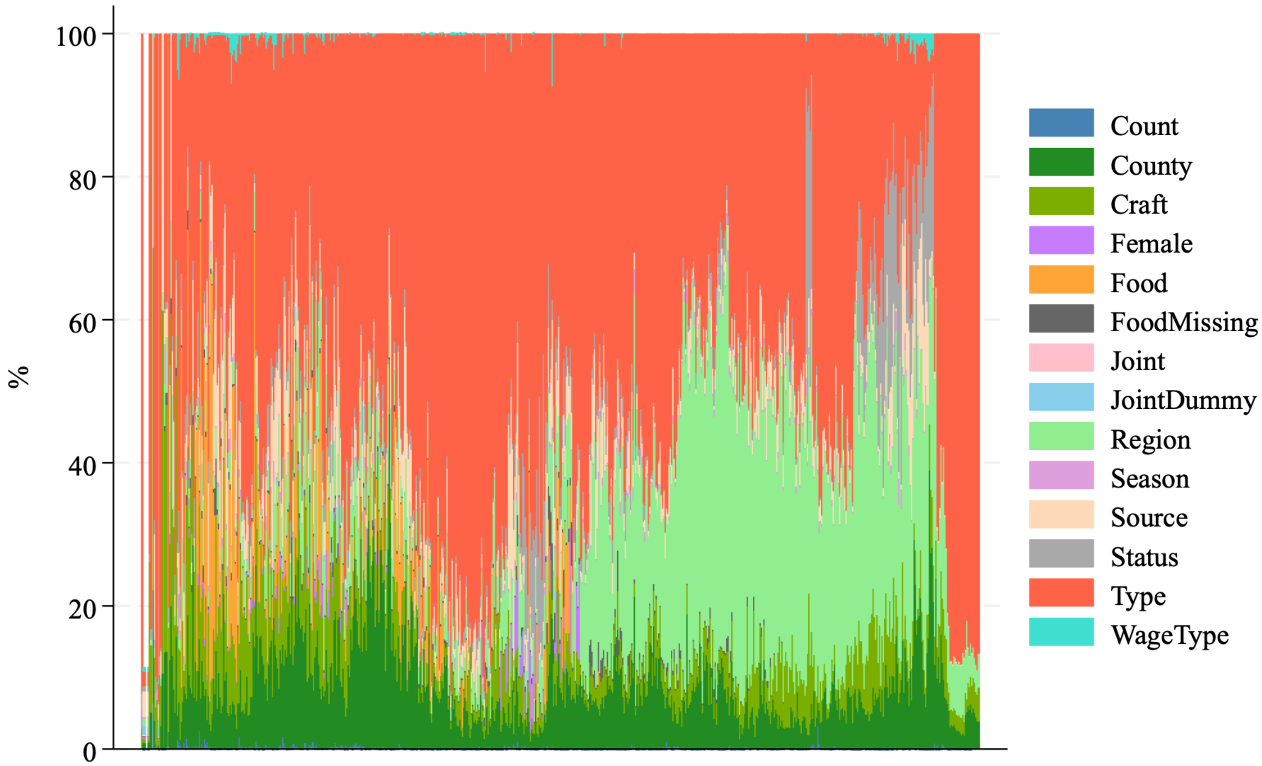

FIGURE A3: FEATURE IMPORTANCE, 1290-1900.

E Application to Japanese Wages

As an additional robustness check, we explore whether we see similar improvements in precision for an entirely different setting — male servant wages in agriculture in Japan from 1610–1890. Kumon (2022) uses a regression in the style of Clark (2005) to reconstruct an annual series of farm wage estimates for Japan, which are then further analyzed using welfare ratios to explore the strength of Malthusian forces during this period. We select this paper for an additional robustness check because it is one of few papers in Table 1 that provides complete and detailed replication files including the raw wage observations underlying the regression analysis.

First we replicate the analysis in Kumon (2022) as a regression model. The model includes dummies for decades or half-decades, dummies for if it was a loan or hereditary servant contract, the duration of the contract, and regions interacted with a dummy for after 1750. We take the OLS model as specified in the replication files. We draw 100 bootstrap samples, estimate this model, and then compute the bootstrap standard error as the standard deviation of the predictions of that model.

Then we implement the past predictive modeling framework for 1610–1890, again computing bootstrap standard errors. Our first prediction is for 1850, using 40 years of data through 1890 to make that prediction. This is because the data are very scarce for the recent years, so 40 years are required to have the about 300 observations that are minimally required to run LightGBM. We use LightGBM with its default hyperparameters. Because we do not use the validation dataset to tune the hyperparameters, our estimates below are an underestimate of the total possible gain from using the past predictive modeling framework.

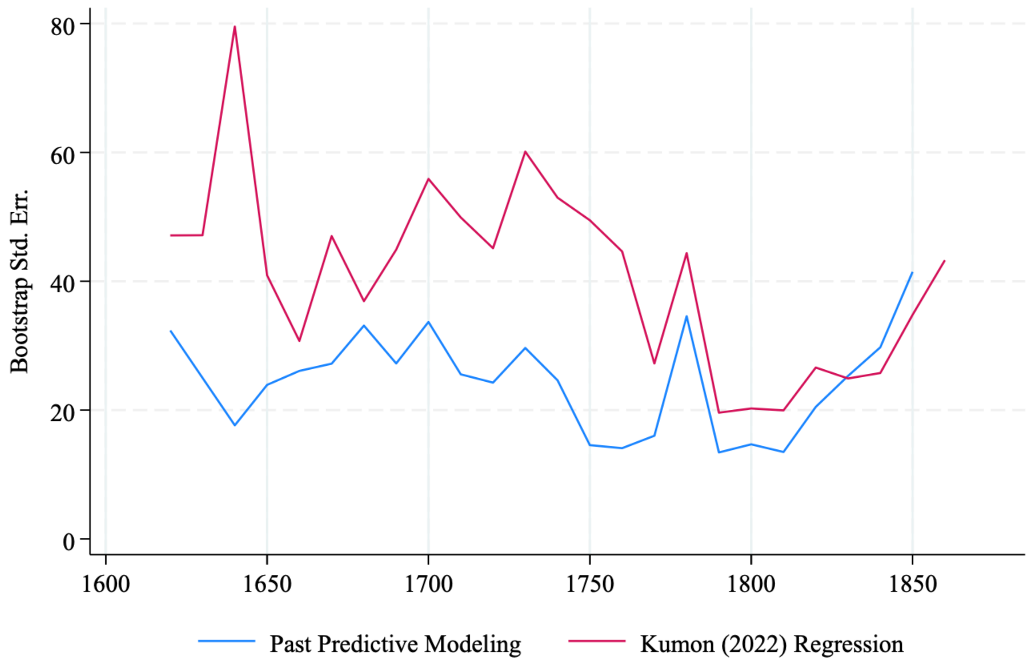

FIGURE A4: BOOTSTRAP STANDARD ERRORS,
KUMON (2022) REGRESSION VS. PAST PREDICTIVE MODELING

Figure 4 presents the results of the bootstrap analysis for the Kumon (2022) regression model (in pink) and the past predictive modeling approach implemented with LightGBM (in blue), plotted as decadal averages. A higher standard error indicates more uncertainty in the prediction. On the whole, the past predictive modeling framework produces more precise estimates in almost every decade. The average bootstrap standard error 1610–1850 is reduced from 38.62 to 23.07 using the past predictive modeling framework. This is a reduction in uncertainty of 40.25%. The fact that the past predictive modeling framework reduces bootstrapped standard errors for farm servant laborers in Japan shows that the approach can reduce overfitting and improve precision generally across many settings.

F Robustness exercises for validation of the framework

First, we check to see if the differences between the new past predictive modeling series and the standard wage estimates from Clark (2005) are greatest in periods where data are most scarce. Figure A5 relates the level difference between the two wage series and the sample size of the prediction sample. The correlation is negative, -0.0908 for laborers and -0.1061 for craftsmen, indicating that as the sample size increases, the difference between the two series decreases. This is the relationship we would expect to see if the new method is indeed improving the predictions in periods where data are more scarce.

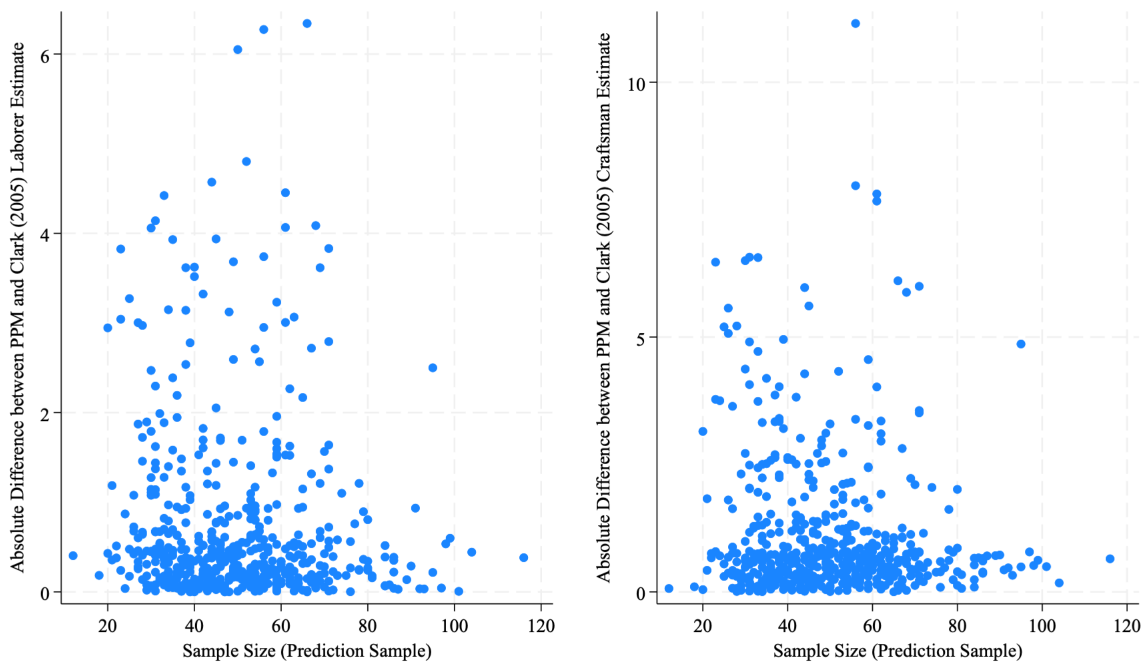

FIGURE A5: DIFFERENCE BETWEEN PAST PREDICTIVE MODELING AND CLARK (2005) SERIES, AGAINST SAMPLE SIZE

Our next robustness check is to compare the correlation of our laborers' (unskilled) wage estimates with wheat prices. Wage prices and grain prices should be related through both income and in wage effects for a period in which agriculture is the dominant sector (Broadberry et al. 2015) and consumption baskets are dominated by grain-based foods (Clark 2005, p. 1328). To explore this, we compute the first differences regression of the laborer's wage on the wheat price prior to 1600:

$$\Delta LaborerWage_t = \beta_0 + \beta_1 \Delta WheatPrice_t + \epsilon_i \tag{1}$$

Where $LaborerWage_t$ is the estimate of the average laborer's wage in year t, $WheatPrice_t$ is the estimate of the average wheat price in year t from (Clark 2005), and β_1 is the impact of changes in the wheat price on changes in wages. An estimate nearer to 1 indicates that wages and wheat prices change at a more similar rate.

The results are given in Table A6. Columns (1) and (2) give the results of the correlation with the wheat prices from Clark for 1300-1600, and the remaining columns restrict the sample to 1300-1400, 1400-1500, and 1500-1600, respectively. The correlation with the new predicted laborer wages from the past predictive modeling framework is given first, followed by the correlation with Clark's laborer wages. In all periods, there is a more positive correlation with the new past predictive modeling estimate, signaling that changes in the wheat price are more closely associated with changes in the wage. However, we note that this analysis is limited by a lack of statistical significance.

TABLE A6: CORRELATIONS OF LABORER WAGES AND WHEAT PRICE, PAST PREDICTIVE MODELING VS. CLARK (2005)

	(1) PPM 1300-1600	(2) Clark (2005) 1300-1600	(3) PPM 1300-1400	(4) Clark (2005) 1300-1400	(5) PPM 1400-1500	(6) Clark (2005) 1400-1500	(7) PPM 1500-1600	(8) Clark (2005) 1500-1600
Wheat	0.103* (1.75)	0.0626 (1.25)	0.142 (1.54)	0.0869 (0.79)	0.135 (0.97)	-0.0582 (-0.47)	0.0885 (0.90)	0.0658 (0.88)
Constant	0.0228 (1.10)	0.0189 (1.07)	0.0187 (0.76)	0.0169 (0.57)	0.00626 (0.27)	0.00304 (0.14)	0.0417 (0.83)	0.0324 (0.85)
Observations	282	282	81	81	100	100	100	100

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G Comparisons with other wage evidence

Clark provides the main wage series for England (Clark 2005). However, other series have been created, for regions or specific locations. In this section, we compare the wage predictions from past predictive modeling to these.

Figure A6 compares our prediction of the laborer's wage gainst Allen's alternate influential wage (Allen 2001, 2009). It shows that our estimates of the broad trends for labourer's day rates broadly accord with Allen's for the three regions he discusses, the North and South of England and London.

Figure A7 presents a range of town-level wage estimates from the late seventeenth and eighteenth century generated by Woodward (1995) in his detailed study of urban construction in Northern England.

It shows that our prediction of northern construction laborers' wages is very close to these different urban estimates, generally sitting around the middle of the range.

Finally, Figure A8 gives a comparison between our northern construction laborer wage and a set of wage estimates for agricultural laborers constructed by Hunt (1986) for several counties. It also includes the agricultural wage estimate from Allen (2001). As we would expect, urban construction workers received a wage premium in the eighteenth century. The convergence of urban and rural unskilled wages in Yorkshire and Lancashire the 1790s may reflect the impact of growing demand from industry in these areas.

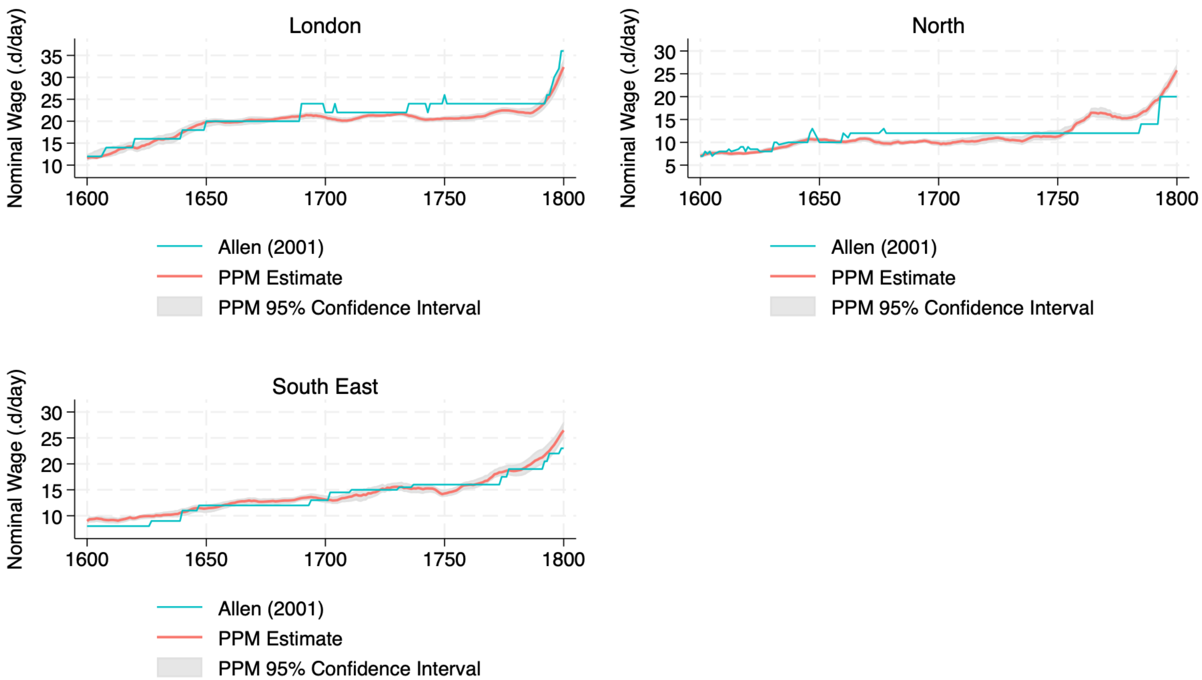

Figure A6: Comparison to Allen's Regional wage patterns, 1500 to 1840

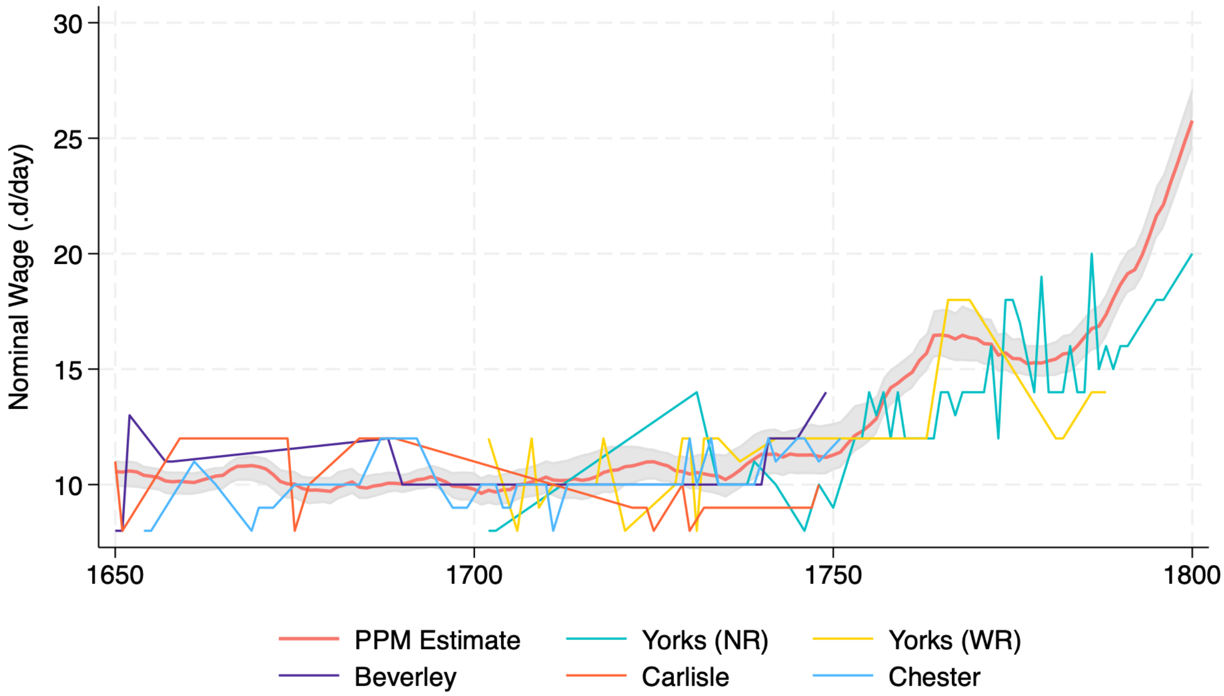

FIGURE A7: COMPARISON TO URBAN CONSTRUCTION WAGES, 1650-1800

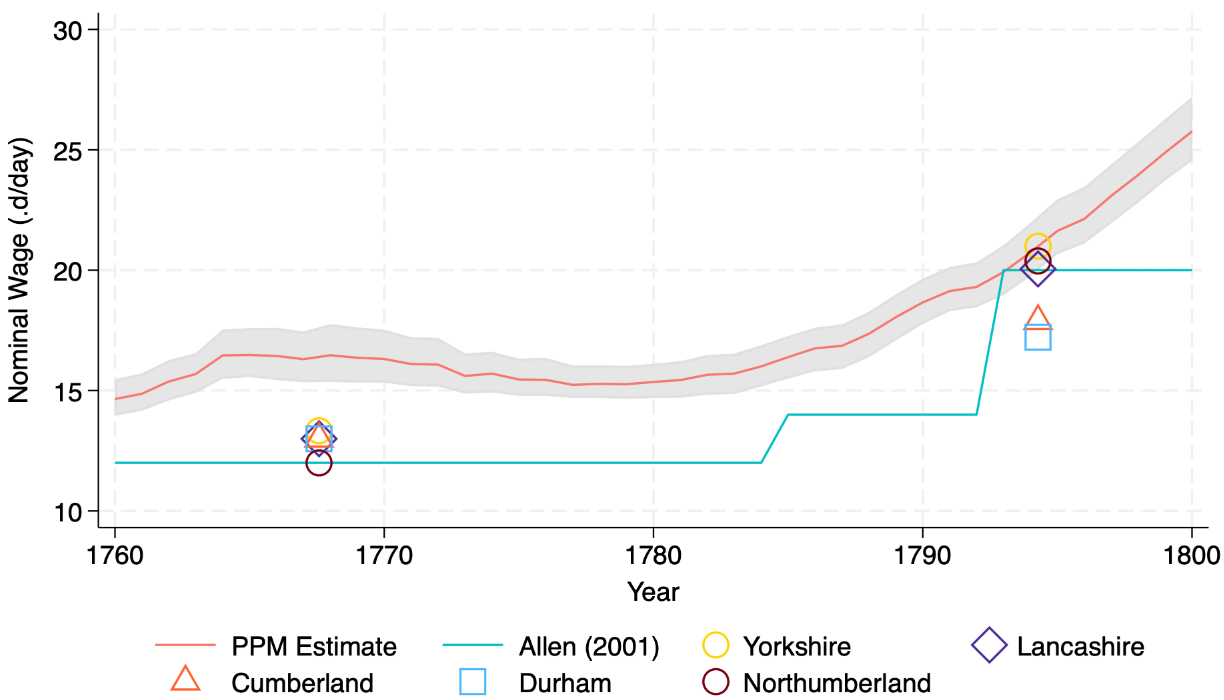

FIGURE A8: COMPARISON TO AGRICULTURAL WAGES 1760-1800

H Weighting Estimates

In most large historical datasets of day wages or similar variables, the underlying observations survive in varying amounts from the different regions in different periods of time. In most cases it is also true that during the period in question population was distributed unevenly across those different regions. The changing shares of available data that comes from different geographical regions over time mean that generating national estimates requires us to pay attention to the underlying sample structure of the data.

This might be an obvious point, but it has not been taken into account in previous estimates of nominal wage series with the exception of Humphries and Weisdorf (2019). The potential scale of the issue is illustrated clearly in the main text in Figure 9 where the deviation between our weighted and unweighted estimates of laborers and craftsmens' wages become large and economically meaningful in the decades around 1700. It is an issue that cannot simply be resolved by incorporating geographical or regional controls within a regression model. That provides an estimation of the size of regional variations, but it does not provide guidance on how to aggregate from the regional to the national level.

In the specific case we explore here, the deviation between the weighted and unweighted series is driven by a period in which a disproportionate share of the observations come from London, the English city with the highest wages, as illustrated by A9

We aggregate by weighting regional wage estimates by the population of the region in each year to make the average wage more representative geographically (as in Humphries and Weisdorf (2019, 2872 n.11). For long-run estimates this inevitably adds a further challenge, as it requires sub-national population estimates for periods of time when even national population figures are disputed. We address that by focusing on regional population shares, and interpolating between benchmark periods using the growth rate of the national population (Broadberry et al. 2015; Wrigley 2007, 2009). Regional population shares change relatively slowly before the modern period, due to the slower pace of migration and economic development.

We argue that this is more appropriate than the two main alternatives. The first alternative approach would involve allowing the sample to define the weight based on the number of observations in each period. This is evidently potentially a source of substantial bias if the underlying data is not randomly generated, which is the case in most studies.

The second alternative is an unweighted average of regional series. In the absence of a guide to re-

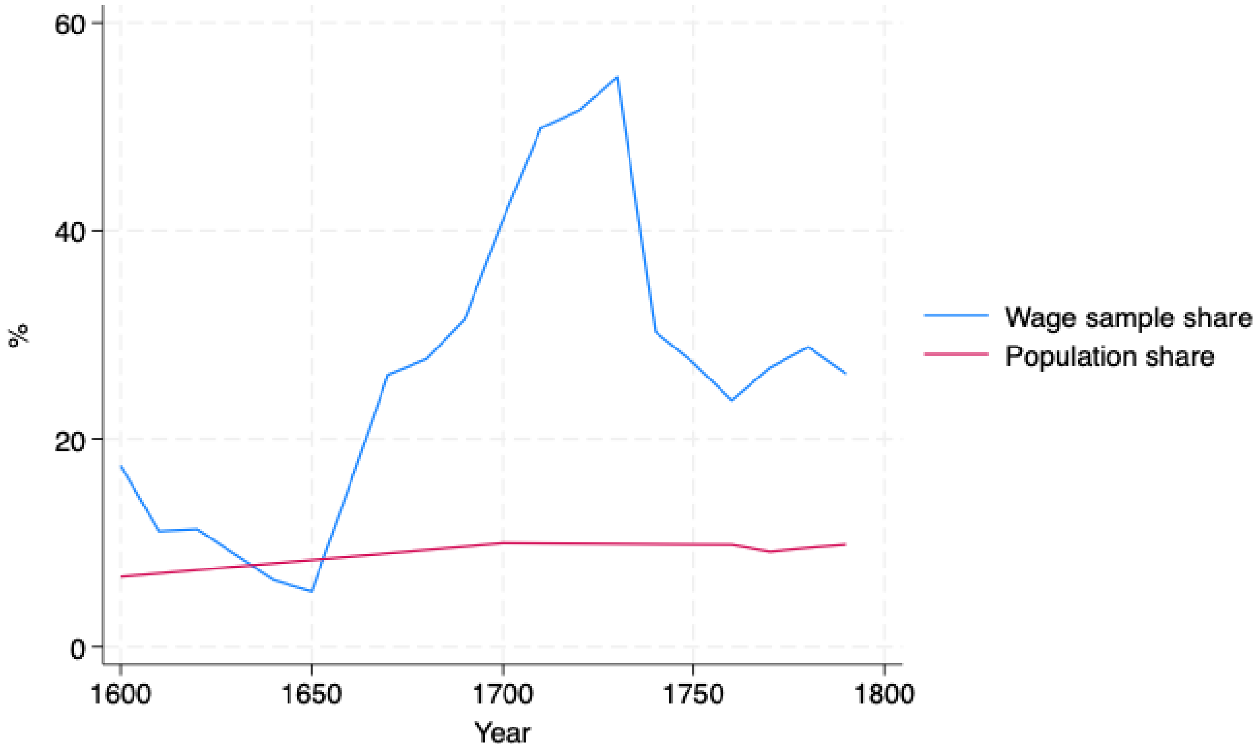

FIGURE A9: SHARE OF OBSERVATIONS AND POPULATION FROM LONDON

Note: The figure reports the share of wage observations in London per decade, and estimates of the share of the English population in London. Source: see text.

gional population shares, this is perhaps unavoidable. However, in these settings, the aggregate estimate may lock in substantial bias if regional shares do change meaningfully. In the period we study, for example, London's share of the English population rises from c. 1.7% in 1300 to c. 13% in 1820.

I Productivity replication

In Section 6.4, we report the results of replicating the recent analysis of long-run productivity growth in England by Bouscasse, Nakamura and Steinsson (Bouscasse et al. 2025). Our replication involved substituting construction laborers' and craftsmen's wages predicted using past predictive modeling for the series from Clark (2005) that is used as the main wage data in their paper. We reused their exemplary code and replication package without any other alterations, as our concern was to identify any implications from using the past predictive modeling wage series.

The only adjustment that we make in methodological terms relates to the time period used to calculate the average wage in each period. The wages that Bouscasse et al. (2025) use are decadal averages of nominal wages deflated by Clark's Cost of Living index. Thay take the forward average of each decade's wages, starting with the first year of the decade. This means that the wage for 1340 is an average of 1340-49. We use an alternate estimate centered on the first year of the decade, making 1340 the average of 1335-1344. We do this because the past predictive modelling framework produces elevated wage estimates for 1345 onwards as it cannot 'see' the arrival of the Black Death until that enters the dataset. By using a centered average, we avoid polluting our estimate of the average wage in the period before the arrival of the pandemic with the effects of the plague.[27]

As Table A7 shows, the trend in our centered average is much closer to that in Clark (2005), while a forward average of the past predictive modeling wage already begins to increase in the 1340-49 estimate. Keeping the 1340 estimate free of the impact of the Black Death is critical for two of Bouscasse et al's models which rely on wages in the decades 1340 and 1360 for identification: "Specifically, we estimate α as the ratio of the change in real wages and the change in the population between 1340 and 1360" (Bouscasse et al. 2025, 11).

	Clark		**PPM**		**PPM**	
Year	**Labourer**	**Craftsmen**	**Labourer**	**Craftsmen**	**Labourer**	**Craftsmen**
	Forward	*Forward*	*Centred*	*Centred*	*Forward*	*Forward*
1300	1.6	3.0	1.6	2.8	1.6	2.9
1310	1.7	3.3	1.7	3.1	1.8	3.1
1320	1.7	3.2	1.8	3.0	1.8	3.1
1330	1.7	3.3	1.9	3.1	1.9	2.9
1340	1.6	2.9	1.9	2.8	2.2	3.2
1350	2.3	4.1	2.8	3.9	2.9	4.0
1360	2.8	4.5	3.0	4.1	3.3	4.3
1370	2.9	4.7	3.5	4.5	3.4	4.6
1380	2.9	4.6	3.3	4.5	3.4	4.5
1390	2.9	4.6	3.4	4.4	3.5	4.5

TABLE A7: AVERAGE WAGE ESTIMATES (1300–1390)

Bouscasse et al. discuss a rich array of models to account for different issues. We concentrate on three of these that are core to their main findings in order to establish the effect of using the new wage series. First, their "simple" model, estimated from the Black Death shock. Second, their "baseline" model, which is their preferred structural model, and which also bases the slope of the labor demand curve on the Black Death. Third, their "structural" model in which the labor demand curve is conditional on the entire

[27] This is a minor concern with the way in which their approach works currently. Clark's annual wage series starts to rise in 1349, the year after the start of the epidemic, particularly for craftsmen. However, as this only provides 10% of the 1340 sample in Bouscasse et al. (2025), it has little impact.

sample, rather than the Black Death. This last model is particularly important as it is largely independent of the decisions we make about how to treat wages in the 1340s.

In the main text, we focus on the changes we observe in the estimates of productivity during the centuries after the Black Death that are obtained from the baseline model. Here we show the estimates of productivity obtained using the other models, to substantiate that result. We also briefly present results that confirms their other main findings on the break point in productivity growth and rates of productvity.

Figure A10 and Figure A11 report new estimates of productivity using the simple and the structural models. The simple model provides an even stronger case for a rise and decline of productivity in the thirteenth to sixteenth centuries. It suggests that productivity may not have fully returned to pre-Black Death levels at the end of this cycle, and that it was not until *c*.1700 that England overtook the productivity seen around 1450. The results of the structural models are more similar, in part because the rise and fall is more evident in the original estimates. Still, we again see higher estimates of productivity in the fifteenth century from the past predictive modeling wages.

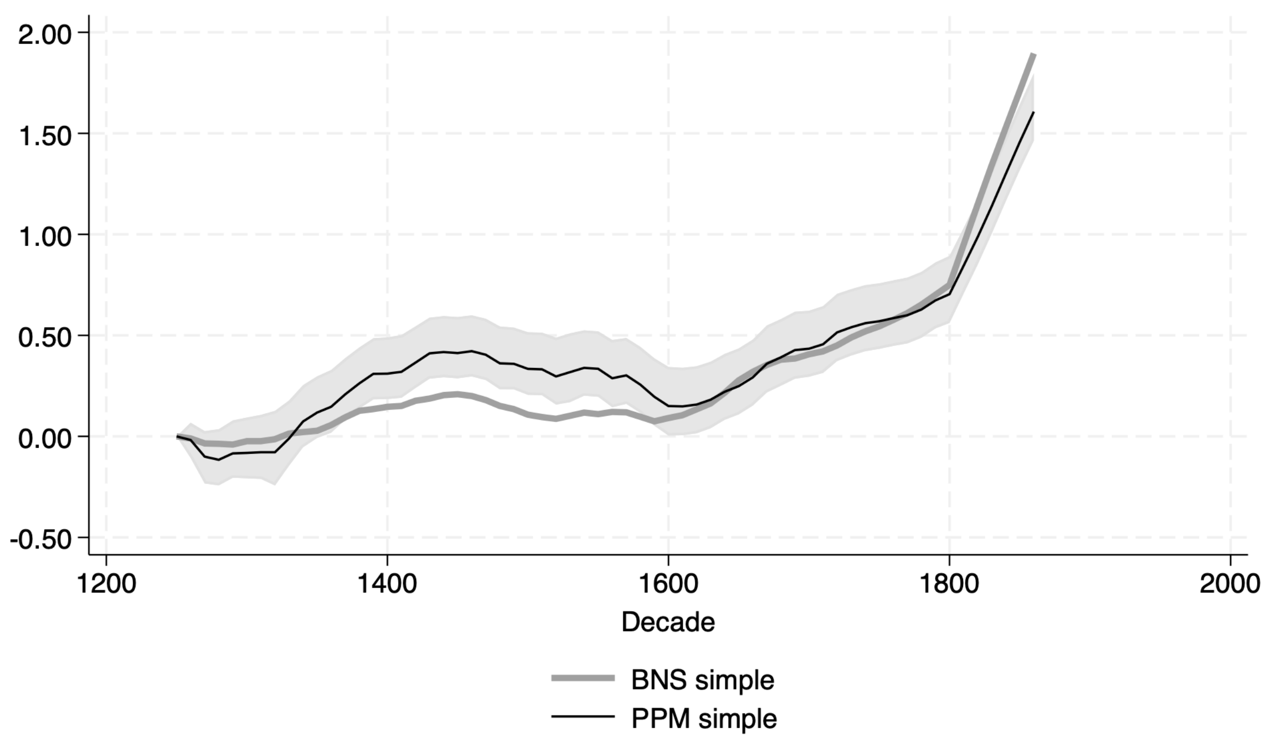

FIGURE A10: ESTIMATES OF PRODUCTIVITY IN ENGLAND, SIMPLE MODEL

Note: Each series is the natural log of productivity, as reported in (Bouscasse et al. 2025, Fig. I). We compare their estimates using Clark's wages ("BNS simple") and the results of the same model using the past predictive modeling wages ("PPM simple").

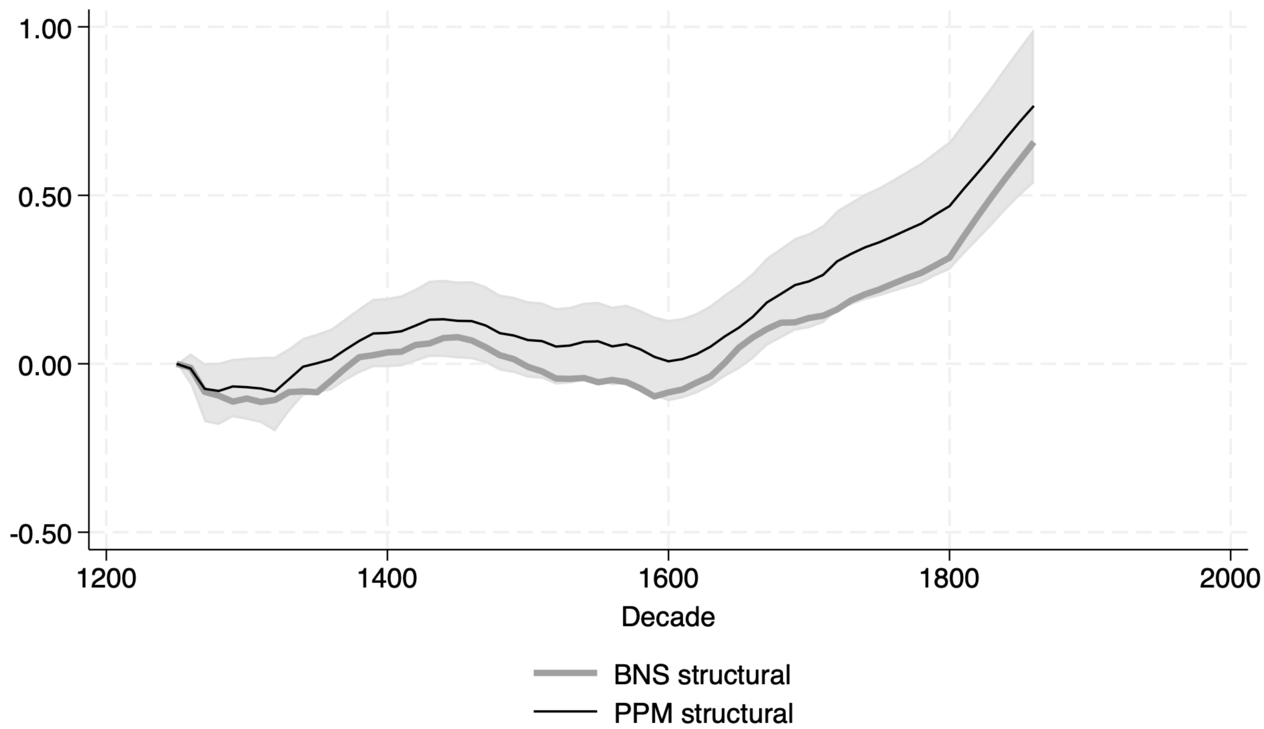

FIGURE A11: ESTIMATES OF PRODUCTIVITY IN ENGLAND, STRUCTURAL MODEL

Note: Each series is the natural log of productivity, as reported in (Bouscasse et al. 2025, Fig. I). We compare their estimates using Clark's wages ("BNS structural") and the results of the same model using the past predictive modeling wages ("PPM structural".

To turn from this question to their main findings, first, Bouscasse et al. find that productivity growth began around 1600, in line with other recent estimates (Wallis et al. 2018). The timing of the break in productivity that we observe using the past predictive modeling wages is the same. Figure A12 reproduces the estimates of the probability that a structural break occurred in the decades between 1550 and 1800 that we obtain using their main models alongside the probability from their simple model using Clark (2005), which is reported in Figure V in their paper. All share the same prediction, that 1600 is the most likely break point, confirming that element of their analysis.

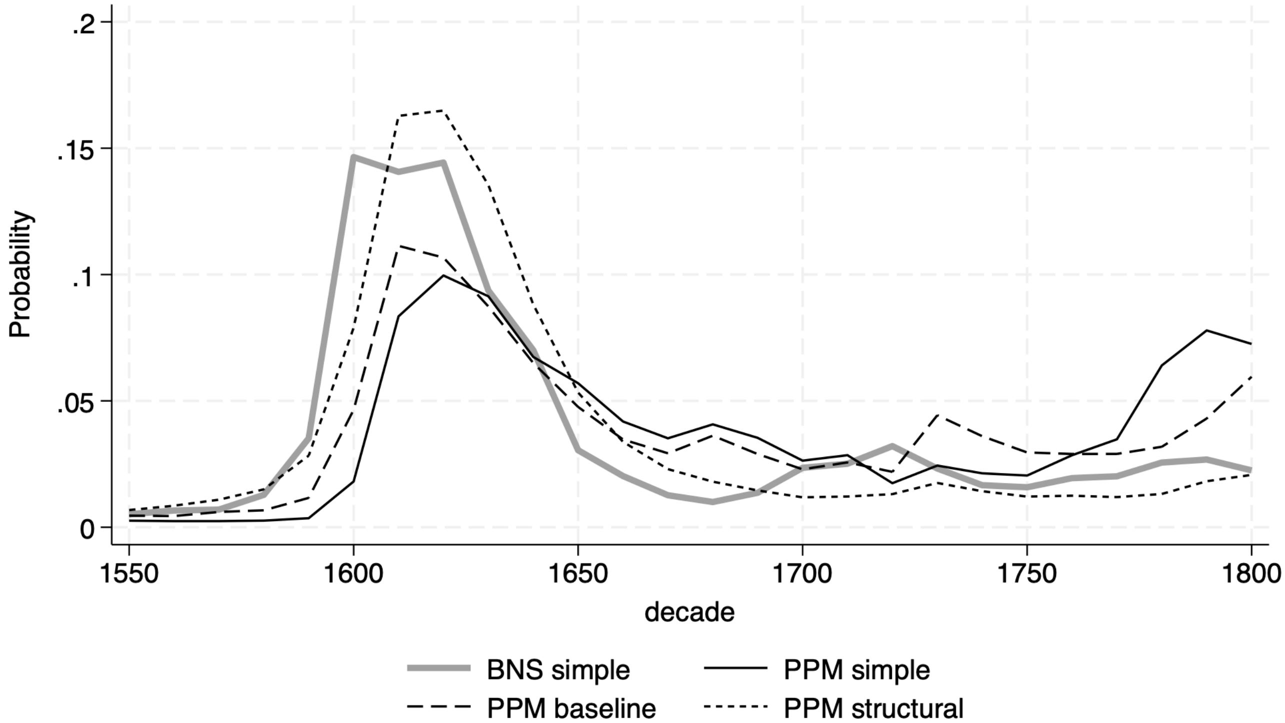

FIGURE A12: PROBABILITY OF DIFFERENT PRODUCTIVITY GROWTH BREAK DATES

Note: The figure plots estimates of the probability that a structural break occurred in different decades between 1550 and 1800, as in Bouscasse et al. (2025, 55). It reports the results from their simple model using Clark's wages ("BNS simple") and three models using the past predictive modeling wages.

Second, Bouscasse et al. provide new estimates of the rates of productivity growth in the long-run, across the three regimes they identify: before 1600, 1600-1800, and 1800-60. We report estimates from the three different models run using past predictive modeling wages in Table A8. The estimates for each period are very similar to those they present, while we report alongside the new estimates for comparison. The baseline model using the new wage series gives the same overall mean productivity growth in each of the regimes, with the only change being wider bounds for the distribution, indicating a greater range of possible values, as we would expect for the period before 1600 particularly. The simple model suggests a positive but very small average rate of productivity growth before 1600, but the bounds cross zero. It also reports somewhat lower rates of growth in both subsequent periods, without challenging the overall narrative.

TABLE A8: PRODUCTIVITY GROWTH

	PPM wages				Clark (2005) wages			
	Mean	St Dev	2.5%	97.5%	Mean	St Dev	2.5%	97.5%
				Simple Model				
$\mu_{a,1}$	0.01	0.01	-0.01	0.03	0.00	0.01	-0.01	0.02
$\mu_{a,2}$	0.03	0.03	-0.01	0.10	0.04	0.02	0.02	0.10
$\mu_{a,3}$	0.15	0.01	0.12	0.18	0.19	0.01	0.17	0.22
				Baseline Model				
$\mu_{a,1}$	0.00	0.01	-0.01	0.02	-0.00	0.01	-0.01	0.01
$\mu_{a,2}$	0.02	0.02	-0.02	0.05	0.02	0.01	0.01	0.04
$\mu_{a,3}$	0.05	0.01	0.03	0.08	0.05	0.01	0.03	0.08

Note: The table reports the mean, standard deviation, 2.5% quantile, and 97.5% quantile of the posterior distribution in the same form given in (Bouscasse et al. 2025, 871). The three regimes cover the years 1250-1590, 1600-1800, 1810-1860.