Using Supervised Learning to Estimate Inequality in the Size and Persistence of Income Shocks

David Bruns-Smith
bruns-smith@berkeley.edu
UC Berkeley
California, USA

Avi Feller
afeller@berkeley.edu
UC Berkeley
California, USA

Emi Nakamura
enakamura@berkeley.edu
UC Berkeley
California, USA and Outside Researcher, Central Bank of Iceland

ABSTRACT

Household responses to income shocks are important drivers of financial fragility, the evolution of wealth inequality, and the effectiveness of fiscal and monetary policy. Traditional approaches to measuring the size and persistence of income shocks are based on restrictive econometric models that impose strong homogeneity across households and over time. In this paper, we propose a more flexible, machine learning framework for estimating income shocks that allows for variation across all observable features and time horizons. First, we propose non-parametric estimands for shocks and shock persistence. We then show how to estimate these quantities by using off-the-shelf supervised learning tools to approximate the conditional expectation of future income given present information. We solve this income prediction problem in a large Icelandic administrative dataset, and then use the estimated shocks to document several features of labor income risk in Iceland that are not captured by standard economic income models.

CCS CONCEPTS

• **Applied computing → Economics**; • **Computing methodologies → Model development and analysis**.

KEYWORDS

income inequality, time series forecasting, uncertainty quantification

ACM Reference Format:
David Bruns-Smith, Avi Feller, and Emi Nakamura. 2023. Using Supervised Learning to Estimate Inequality in the Size and Persistence of Income Shocks. In *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), June 12–15, 2023, Chicago, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3593013.3594113

1 INTRODUCTION

Economic hardship and the dynamics of socioeconomic inequality depend crucially on the income shocks that households face. Such shocks come in a variety of forms: positive shocks include promotions and stimulus checks; negative shocks include job loss, illness, and lack of available working hours. A large body of research shows that unexpected income shocks pass through to changes in household spending and saving [45, 53, 59], with wide heterogeneity in responses across household characteristics [8, 24, 27, 36, 38]. Households' responses to shocks are also key drivers of financial fragility [1, 47, 50], the effectiveness of fiscal policy [7, 34, 35], and the evolution of wealth inequality [6, 20, 54].

To study the impacts of shocks on households and the corresponding implications for subsidy allocation or macroeconomic policy, we first need to be able to measure them. While we observe changes in income from one period to the next in many economic datasets, we rarely observe what proportion of those changes were unexpected shocks, or whether those shocks were temporary or will persist long into the future. Consider a household that makes $60,000 dollars annually. In the next year, the household might simultaneously experience a promotion to site manager — a persistent increase of $5,000 a year — and an especially-snowy spring construction season — a temporary decrease of $20,000 a year — for a total observed shock of -$15,000. Does this hypothetical household spend more because their expected income will be higher into the future? Do they (or can they) spend less to weather the larger but temporary negative shock? Do they have a savings buffer to draw upon, or does the unexpected temporary loss of income cause them to miss their mortgage payments? These various considerations are difficult to tease apart based on the observed, overall shock alone.

In the economics literature, the workhorse statistical model for analyzing shocks and their persistence is a panel model where current income is the sum of a random *transient* shock and unobserved *permanent* income that evolves according to an autoregressive process [2, 10, 44]. Statistical estimands of interest, such as the size of transient and persistent shocks, are then defined with respect to the parameters of this model. Importantly, however, this model embeds a series of assumptions about the income process that impose strong homogeneity across households and over time, such as assuming that households across the income distribution face shocks of the same size, severely limiting our ability to understand critical sources of variation.

In this paper, we instead propose to directly estimate income shocks and their persistence from income data. We first propose a non-parametric estimand for income shocks — defined outside of any particular statistical model — in terms of the conditional expectation of future income given the information known in the present. Estimating these conditional expectations for a particular population requires finding the best mean-squared error predictors for data drawn from that population, a task we can perform using off-the-shelf supervised learning tools with strong uniform convergence guarantees. Our procedure outputs estimates of income shocks associated with each income observation along with

FAccT '23, June 12–15, 2023, Chicago, IL, USA

ACM ISBN 979-8-4007-0192-4/23/06.
https://doi.org/10.1145/3593013.3594113

the persistence of those shocks at several horizons into the future. These shocks can then be used in downstream tasks like estimating households' consumption/savings response, calibrating models for the evolution of wealth inequality, or as real-world datasets for studying algorithmic fairness.

Contributions:

- We provide a nonparametric definition of income shocks that relaxes strong functional form assumptions, and allows researchers to assess heterogeneity in the size and persistence of income shocks across observed features.
- As a real-world application, we estimate income shocks in Iceland by predicting labor income at various horizons into the future using a large administrative tax dataset.
- We document several features of the estimated shocks that are not captured by standard economic parametric income models, including: a much larger magnitude of income risk faced by individuals at the bottom of the income distribution; an exponential decay in the persistence of shocks on average over time; wide hetereogeneity in the persistence of shocks across household circumstances; and substantial asymmetry between positive and negative shocks.

We hope to draw attention to an under-utilized role for prediction in the social sciences, where supervised learning models are used as an approximation of a conditional expectation, rather than used to predict future outcomes for new, potentially out-of-distribution observations. We also hope to further connect the parallel research on income shocks in economics and computer science. The large body of research on income uncertainty and household responses in economics can bring valuable insight to the recent literature on algorithmic fairness and inequality. Likewise, the powerful non-parametric modelling and optimization toolboxes from computer science can shed new light on the dynamics of income.

2 RELATED WORK

2.1 Economics

A large literature on economic theory studies household responses to income shocks. The permanent income hypothesis [16, 26] suggests that households will smooth consumption over the lifecycle, and predicts very small responses to temporary shocks but large responses to permanent shocks. A literature on precautionary savings explores why consumption seems to track income so closely in the data with an emphasis on uninsurable idiosyncratic income risk [12, 13, 29, 32, 35].

Most closely related to the current work is the literature on transient-persistent income process models. Linear transient-persistent autoregressive models have been widely used [2, 10, 44], and we will compare these parametric models to our non-parametric estimands in Section 3.2.

Several recent papers have critiqued these models for failing to match key stylized facts documented in real income data. For example, Guvenen et al [30] emphasize the substantially higher skewness and kurtosis exhibited by US income data that linear panel models have difficulty reproducing. Other papers address the assumptions about the persistence of shocks embedded in standard models. For example, De Nardi et al [21] discretize income into buckets, and then non-parametrically estimates a first-order Markov chain for transitions between these discrete states. They find evidence for heterogeneity in the persistence of shocks and substantial deviations from the AR(1) process in [10]. Arellano et al [4, 5] propose a Bayesian approach for estimating the posterior of permanent income (defined as a latent variable) using expectation-maximization. Straub [63] creates proxies for permanent income in real data by averaging together several past and future income observations — a procedure that could be seen as a very simple version of predicting income at several future horizons.

2.2 Computer Science and Machine Learning

Recently, a growing literature in computer science and algorithmic fairness has also emphasized the role of income shocks. This includes algorithm and mechanism design research for subsidy allocation where the level of household income and wealth as well as their susceptibility to shocks play central roles [1, 50, 52]. Other work studies the dynamics of income inequality over time [31, 56], including their implications for policy interventions. In many ways, this literature has close connections to the macroeconomic consumption and inequality literature. In fact, Nokhiz et al [50] solves and simulates from a macroeconomic consumption model with incomplete markets and precautionary savings motives in the style of Hubbard et al [32] or Gourinchas and Parker [29].

Most of the related work in computer science has used simulated income data. For example, Abebe et al [1] simulate shocks as arriving via a Poisson process. Nokhiz et al simulate income with a first-order Markov chain over discretized income states. On the other hand, D'Amour et al [18] have an *implicit* model for income shocks in their simulated model for loan repayment. In their model, the probability of repayment is a deterministic function of credit score, embedding a number of assumptions about credit score calculations and the income risk faced by households that necessarily partially determines their ability to repay. Our work is complementary to the work above and provides an alternative to simulation, measuring the degree of labor income risk directly in real-world data.

Reader et al [56] suggest modelling the evolution of income inequality as a linear dynamical system, with policy interventions and feedback loops modelled as a PID controller. Our method similarly has connections to the controls literature. As we will discuss, the transient-persistent models for income can be formulated as partially-observed dynamical systems, finite-sample estimation of which has featured in recent research on system identification [39, 40, 46, 61].

An adjacent literature studies the dynamics of income *between* generations, mostly focused on interventions in university admissions [3, 31]. This work complements a large body of work on inter-generational mobility in economics [14, 17, 19]. Extending our predictive estimands and the corresponding measures of income risk to an inter-generational context would be an interesting direction for future work.

2.3 Limits of Prediction in Social Science

Another important literature emphasizes the limits of predictability of future life outcomes. Narayanan [49] called predicting social outcomes "fundamentally dubious". A large-scale prediction competition, the Fragile Families Challenge, found that predictive accuracy across a variety of social outcomes and algorithms was low across the board [60]. Flexible machine learning models hardly performed better than linear regression on a handful of features. More broadly, Liao et al [41] and Raji et al [55] outline a large taxonomy of basic ML functionality failures in real-world deployments.

We instead emphasize a potentially under-utilized role for prediction in social science: approximating a conditional expectation. This follows the exhortation in Lundberg et al [42] regarding prediction in sociology: to clearly state the statistical estimand of interest. We would like to characterize the distribution of prediction errors around the conditional expectation of future income and how it evolves from one period to the next. If the absolute size of these errors for the best possible predictor given the feature set is large, then this is not a functionality failure, but an accurate statement of income risk in the population. Likewise, if the model has substantially larger prediction errors for one sub-group compared to another, then the relative distribution of these residuals tell us about the inequality in income risk across groups.

2.4 Predicting Future Income

In this work, we solve prediction problems for income h periods into the future conditional on current and past income and other covariates. Surprisingly, we have found very few published papers that solve this kind of income forecasting problem. The only such example to our knowledge is Gerardi et al [28], who use unpenalized linear regression to predict future income conditional on current income, housing wealth, and demographic variables. See Section 4.3 for a discussion of the performance of linear regression in our setting.

A very large literature in machine learning considers income prediction problems framed as classification tasks. Most of this work is centered on the Adult dataset [37], first used to assess the performance of tree-based ensembles [9]. See Ding et al [23] for a review of recent research using this dataset, especially on algorithmic fairness, and several associated limitations. The standard task on Adult is classifying whether or not income falls below or above \$50,000. In contrast, we consider income prediction as a regression problem, and introduce a dynamic dimension by forecasting future income conditional on current and past income. Furthermore, the emphasis of our work is different but complementary to the fairness literature using Adult — if our prediction algorithms have larger prediction errors for certain sub-populations we interpret this as a substantive result about the relative income risk faced by those sub-populations.

3 DEFINING INCOME SHOCKS

Let y_{it} denote log income of individual i at time t and let x_{it} denote covariates such as age, education, calendar year, and wealth. We assume that we observe N i.i.d. samples of the trajectories $\tau_i := \{(y_{it}, x_{it})\}_{t=1}^{T_i}$ from the same joint distribution — we make no assumptions within a trajectory on the relationship between y_{it} and x_{it} or their evolution over time. The τ_i can either be interpreted as draws from an underlying joint distribution or as samples from some finite population of individuals. In what follows, we omit the i subscripts when clear from context. In practice, draws across individuals are unlikely to be entirely independent and we consider some common violations, such as domestic partners, in the Appendix.

We define an *income shock* at time t to be the difference between observed income y_t and expected income given all information available before time t. Define the information set $\mathcal{I}_{t-1} = \{y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ...\}$. Then the income shock at time t is:

$$\Delta_t := y_t - \mathbb{E}[y_t|\mathcal{I}_{t-1}]. \tag{1}$$

We define the persistence of the time t income shock as the change in expected future income upon adding the new information (y_t, x_t) into the information set. We write the horizon-h persistence of the shock Δ_t for all $h \geq 1$ as:

$$\phi_{t,h} := \mathbb{E}[y_{t+h}|\mathcal{I}_t] - \mathbb{E}[y_{t+h-1}|\mathcal{I}_{t-1}]. \tag{2}$$

As a concrete example, let $\mathbb{E}[y_t|\mathcal{I}_{t-1}] = 1.0$ and realized income $y_t = 2.0$. Then the total shock at time t is $\Delta_t = 1.0$. Now we use the realized y_t (and x_t) to update the expectations for the future to measure how long the shock lasts. If the updated conditional expectation $\mathbb{E}[y_{t+1}|\mathcal{I}_t] = 1.5$, then the portion of the total shock Δ_t that is expected to remain after one period is $\phi_{t,1} = \mathbb{E}[y_{t+1}|\mathcal{I}_t] - \mathbb{E}[y_t|\mathcal{I}_{t-1}] = 0.5$. If the original and updated 2-step-ahead conditional expectations are $\mathbb{E}[y_{t+2-1}|\mathcal{I}_{t-1}] = 1.0$ and $\mathbb{E}[y_{t+2}|\mathcal{I}_t] = 1.1$, then the amount of the total shock expected to persist two periods into the future is $\phi_{t,2} = 0.1$. So in this example, while the time t unexpected change in income was large, only half of the shock is expected to persist one period into the future, and only 10% of the shock is expected to persist two periods into the future. See Figure 1 for an illustration.

The quantities Δ_t and $\phi_{t,h}$ are our *non-parametric estimands*. They are not observed directly, and we would like to estimate them from data. However, first we briefly justify our choice of these quantities.

3.1 Theoretical justification

Simple theoretical models for household responses to income shocks usually imply that consumption depends on the expected present value of future income, sometimes called *permanent income*. Given a discount rate γ, permanent income at time t is:[1]

$$y_t^{\text{perm}} := \mathbb{E}\left[\sum_{k=t}^{\infty} \gamma^{k-t} y_k \middle| \mathcal{I}_t\right]. \tag{3}$$

Then the unexpected change to permanent income at time t is

$$y_t^{\text{perm}} - \mathbb{E}[y_t^{\text{perm}}|\mathcal{I}_{t-1}] = \Delta_t + \sum_{h=1}^{\infty} \gamma^h \phi_{t,h}. \tag{4}$$

So in this sense, the objects Δ_t and $\phi_{t,h}$ are precisely the relevant theoretical objects for studying a household's response to income shocks. Equation (4) suggests one way to summarize these shocks in a single measurement. Indeed, the framework of updating future expectations as new information arrives is exactly the motivation for the definition of permanent income in Flavin [25].

[1] Typically, permanent income would be discounted by $1/r$, where r is the rate of return on assets and therefore represents the relative value of money now versus money in the future. In the simplest models, $\gamma = 1/r$ in equilibrium.

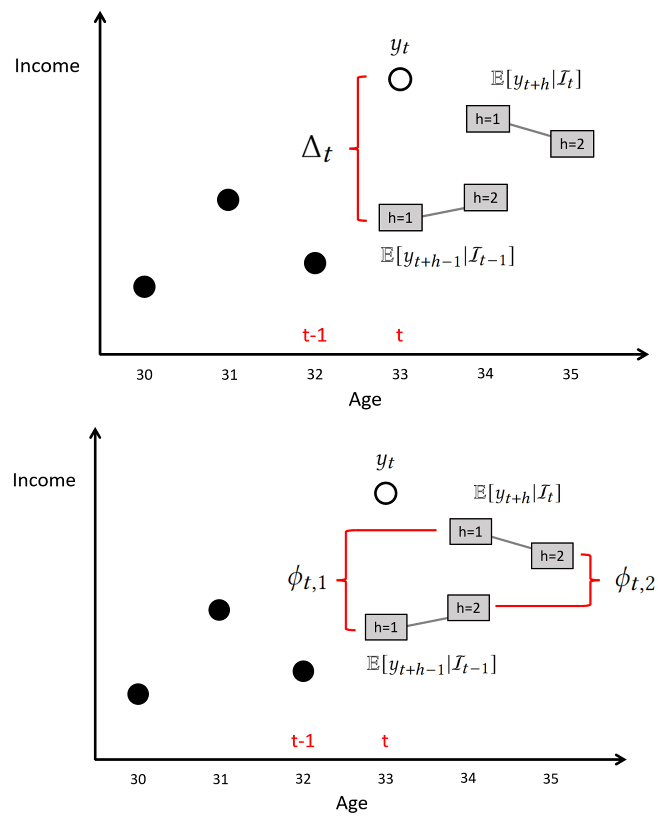

Figure 1: An illustration of our estimands. The black dots represent the conditioning set $\mathcal{I}_{t-1}$. The grey boxes represent the predictions one and two periods ahead. At time t, we observe the new observation (shown as an open circle) y_t. The difference between the new observation and the previous prediction is Δ_t as shown in the upper diagram. When we add y_t to the conditioning set and update the predictions, we get the persistence at each horizon as shown in the lower diagram.

More sophisticated economic models suggest that current consumption choices may depend on the whole conditional distribution of $P(y_{t+h}|\mathcal{I}_{t-1})$ rather than just the conditional mean. This suggests a straightforward extension of our procedure using conformalized quantile regression [57] that we hope to pursue in future work.

3.2 Comparison to parametric estimands

It is helpful to compare our non-parametric estimands to the commonly used transient-persistent model [4, 10]. This model imposes the following structural assumptions on the income process:

$$y_t = p_t + \epsilon_t,$$
$$p_t = f(p_{t-1}) + \eta_t,$$

for some measurable function f and where $\mathbb{E}[\epsilon|p_t] = 0$ and $\mathbb{E}[\eta_t|p_{t-1}] = 0$. First, notice that the expected system transitions in the implied autoregressive model, obtained via the standard trick of re-writing the partially-observed non-linear system as an infinite-order autoregressive model, is exactly the conditional expectation $\mathbb{E}[y_t|\mathcal{I}_{t-1}]$. See for example [40] for discussion in the controls setting.

The classical parametric model used in [10] makes the additional functional form assumption:

$$p_t = p_{t-1} + \eta_t.$$

where the variance of η and ϵ are independent of p_t. Note that in this case, our non-parametric definition for shock persistence exactly corresponds to the persistent shock in the model; $\phi_{t,h} = p_t - p_{t-1}$ because $\mathbb{E}[y_{t+h}|p_t] = p_t$. However, the classical model imposes several additional testable implications: that households across the income distribution face shocks of the same size; that there are no interactions between age, demographics and shock size or persistence; that persistent shocks are perfectly-persistent into the future; and that there is no asymmetry in the persistence of positive and negative shocks. Using our non-parametric estimands that avoid making such strong assumptions, we will demonstrate substantial deviations from this simple model for labor income in Iceland.

3.3 Limitations of any particular estimator

The exact interpretation of our non-parametric estimand depends entirely on our dataset, our definition of the relevant random variables, and the contents of the conditioning set, $\mathcal{I}_t$. For example, as we discuss in Section 4.1, we have to choose the definition of the periods t; are these yearly or monthly shocks? Monthly shocks have to account for seasonal variation, whereas for yearly shocks the business cycle becomes a central object of concern. Likewise, we have to choose a definition of income; does y_t represent labor income or total family income? Any of these choices may not be inherently right or wrong, but will have different implications for the relevant downstream economic analysis.

More generally, as commonly done in the economics literature [2, 4, 10, 20], we measure shocks in terms of a *statistical* expectation. The relevant theoretical objects of interest in Section 3.1, are *household* expectations because a household's behavioral responses to income risk depend on their own beliefs about the future. This can introduce substantial measurement error along across at least two dimensions. First, we do not have access to important private information — for example an individual's plan to leave their job next year to go back to school. Second, our predictions are formed using hundreds of thousands of observations from across the entire population of Iceland, information that any given individual might not have. This discrepancy must be kept in mind when performing any later economic analysis using our estimated shocks. For example, we may try to impose some structural assumptions on the nature of this measurement error, and *partially identify* behavioral estimands to account for the additional uncertainty. Or otherwise, we have to more narrowly interpret our estimated income shocks as a measure of *aggregate* labor income risk across Iceland, capturing heterogeneity across observables in the tax data, rather than the personal uncertainty about future income faced by any particular individual.

4 THE INCOME PREDICTION PROBLEM

Our non-parametric estimands for income shocks, Δ_t, $\phi_{t,h}$, can all be computed given the conditional expectations $\mathbb{E}[y_{t+h}|\mathcal{I}_t]$ for all t and all $h \geq 1$. The conditional expectation is equivalent to the best mean-squared error predictor of y_{t+h} over all possible functions of the features in the set $\mathcal{I}_t$. Therefore, we have reduced the problem of estimating income shocks and their persistence to a series of prediction problems for which we can use off-the-shelf supervised learning tools. In this section, we discuss how we solve these prediction problems in practice with a large administrative tax record dataset from Iceland.

4.1 Data and sample selection

We use income measurements from Icelandic income tax data, made available to us through collaboration with Statistics Iceland. In our Icelandic tax data, the period t is measured in years, and for every individual in every year from 1981-2018 we observe (log) labor income y_t and a collection of other demographic and financial variables. We transform all income observations to 2018 US dollars, adjusting for Icelandic CPI [33] and the exchange rate between the dollar and the Icelandic Króna [51]. While we observe nearly the entire population of Iceland during this timeframe, we restrict our sample to reflect *in-employment* labor income risk. This involves three sample selection steps: (1) we only include individual-year observations with labor income strictly greater than zero; (2) we only include those individual-year observations for which we observe non-zero income for at least six consecutive periods before and at least twelve consecutive periods after; and (3) we only include observations for individuals aged thirty and older (to avoid income changes due to switching in and out of higher education). This leaves 508,235 individual-year observations across 62,387 individuals.

Note that the choice to study in-employment labor risk and unemployment risk separately is common [10, 22, 48, 62]. Our choice to focus on in-employment risk is mostly for purposes of presentation and for comparison with [10]. Unemployment risk is also of central interest, and re-estimating income shocks with unemployment will be the object of future work. Furthermore, the particular choice of zero for the minimum threshold might include individuals who are unemployed for part of the year; for discussions on alternative choices of the minimum threshold see Nakajima and Smirnyagin [48]. Likewise, the choice to study *labor* income as opposed to total income after taxes and transfers has consequences for the interpretation of our estimand. In general, these choices for sample selection and the definition of income do not change the non-parametric estimands outlined in Section 3 but are enormously important for substantive economic analysis of the results.

For covariates x_t, we include age, education, gender, total assets (net of debt), and housing wealth.[2] Education is binned into five categories: incomplete compulsory education, compulsory education only, upper secondary only, undergraduate only, and beyond undergraduate. With no essential loss of generality, instead of fitting a separate model for each t, we will fit a single predictor, but additionally condition on calendar year. Thus our complete feature set includes dummies for calender year t, current income and covariates (y_t, x_t), and six lags of income and covariates, $\{(y_{t-\ell}, x_{t-\ell})\}_{\ell=1}^{6}$. Note that this is only an approximation of the information set $\mathcal{I}_t$, which should include as many lags as are available. However, we can justify this theoretically with relatively mild assumptions on the mixing of the stochastic process for income. In practice, we also found that including more lags does not improve mean-squared error in cross-validation.

[2] Notably missing from the tax data is information on race or ethnicity, presumably due to extremely low rates of immigration. During the timeframe of our dataset, more than 92% of the population were ethnically Icelandic.

4.2 Training

As we would like to use highly-flexible regularized function classes for prediction, we leverage both sample-splitting and cross-validation to prevent over-fitting. Note that while each training sample corresponds to an individual-year observation, the observations within an individual trajectory are highly-correlated. Therefore, we perform all sample-splitting and cross-validation at the *individual level*. First, we randomly divide the full population of individuals into two halves. Within each half, we train models predicting y_{t+h} for $h = \{1, ..., 12\}$, using the feature set described above. Prior to training, all features were shifted and scaled to have mean zero and standard deviation one. We considered a variety of regularized linear models, random forests, and gradient-boosted tree regressors, over a range of hyperparameter values as described in the Appendix. We chose the best performing model using 5-fold cross-validation. Gradient-boosted trees consistently performed the best in cross-validation across all horizons.

The output of this process is our best approximation of the conditional expectations $\mathbb{E}[y_{t+h}|\mathcal{I}_t]$ for all h from each of the two halves. We then compute the estimated income shocks Δ_t and persistence profiles ϕ_{th} by applying the models trained in one half to the individual-year observations in the opposite half. This way, the income shock for each observation is estimated using a model that was never previously trained using that observation.

Remark: We claim that this process gives the best approximation of the conditional expectation in the population. This does not mean that our trained models are the best predictors of future income on never-before-seen observations from a different population. Applying the predictors outside our dataset could face significant distribution shift, perhaps most notably the massive impact of COVID-19 in 2020 and onward. Instead, we rely on the fact that we *randomly* split individuals into two halves from a known population. This means the strong uniform convergence guarantees that come from the i.i.d. assumption in supervised learning apply exactly. As a result, however, any insights about income risk from our procedure are only guaranteed to describe the population of Iceland during the timeframe of our sample — extrapolating outside this population would require additional statistical assumptions.

4.3 Model assessment

Before presenting our results on income shocks, we first assess our models' predictive performance. First, we emphasize the advantage of using a highly flexible model class by comparing our final gradient-boosted tree models to two simpler benchmarks: a simple random-walk baseline that always outputs most-recent income and

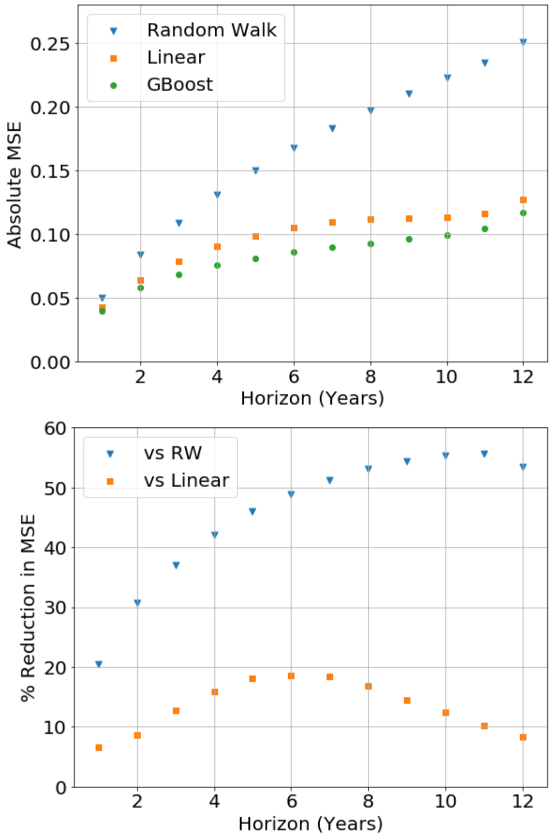

Figure 2: Mean-squared error of hold-out predictions. The top diagram plots the mean squared error of the gradient-boosted trees model, the linear model, and the random walk baseline for horizons 1 to 12. The bottom diagramplots the percent reduction in mean-squared error achieved by the flexible gradient-boosted tree model relative to the linear model and random walk baseline.

ordinary-least-squares linear regression. We are inspired to include the random-walk baseline by a famous macroeconomics result that a random walk beat existing models for predicting exchange rate out-of-sample [43], and the linear regression model due to its strong performance in the Fragile Families Challenge [60].

Figure 2 compares the mean-squared error of the best performing gradient-boosting model and the two baselines. In particular, we plot the MSE of predictions on data points that were not used in training; for each data point, we make predictions for all horizons h, using the models trained in the opposite split. The gradient-boosted trees model perform much better than the random walk, and modestly better than the linear model, achieving between 7 to 19% reduction in MSE. Note that the magnitude of the average prediction errors across the whole population is quite large. For one year ahead, the MSE suggests that the average magnitude of prediction errors is around 0.2 in logs. In levels, this corresponds to an error of about 22% of income. For twelve years ahead, even the best performing model has average prediction errors of around 40% of income. Recall that prediction error one period ahead is exactly the definition of the income shock Δ_t and so, assuming that we have a good approximation of the conditional expectation, the large absolute mean-squared error indicates a fairly substantial amount of income risk. However, the *average* squared-error can be misleading, and we will show later that the largest prediction errors are concentrated at the bottom of the income distribution.

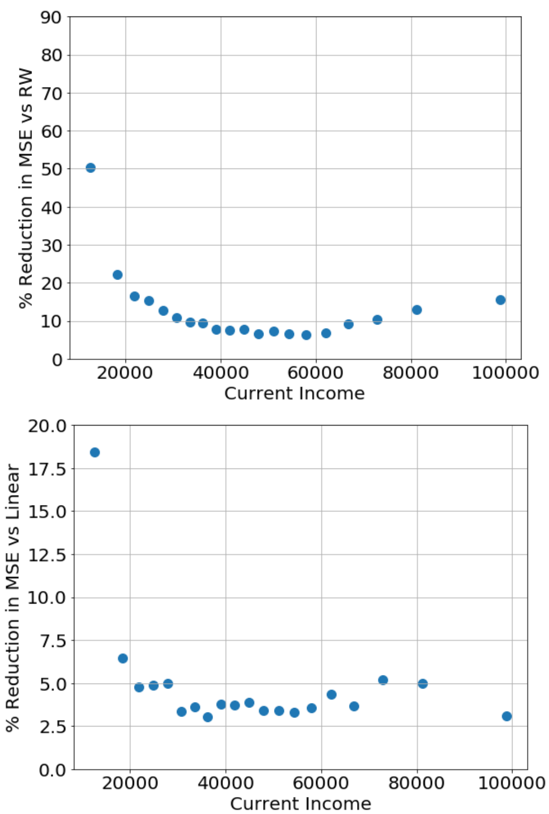

Figure 3: Percentage reduction in mean-squared error for predictions $h = 1$ year ahead across bins of current income. The top and bottom plots compare the gradient-boosted model to the random walk baseline and linear model respectively.

The importance of flexible models becomes more clear in Figures 3 and 4. These figures plot the relative improvement of our gradient-boosted model against the linear and random walk baselines across the distribution of current income. We proceed by binning: we split the observations into 20 equally-sized bins based on quantiles (from 5% to 95%) of log current income. Each dot in these figures corresponds to one of these bins. The x-axis is average current income within bins, with values in levels — recall that this represents inflation-adjusted income in 2018 US dollars. Figure 3 plots the reduction in MSE achieved by the gradient-boosted model compared to the two baselines for predictions $h = 1$ year ahead. Figure 4 plots the reduction in MSE achieved by the gradient-boosted model compared to the two baselines for predictions $h = 10$ years ahead. Note that the flexible model is especially important when predicting future income for houses at the bottom of the income distribution.

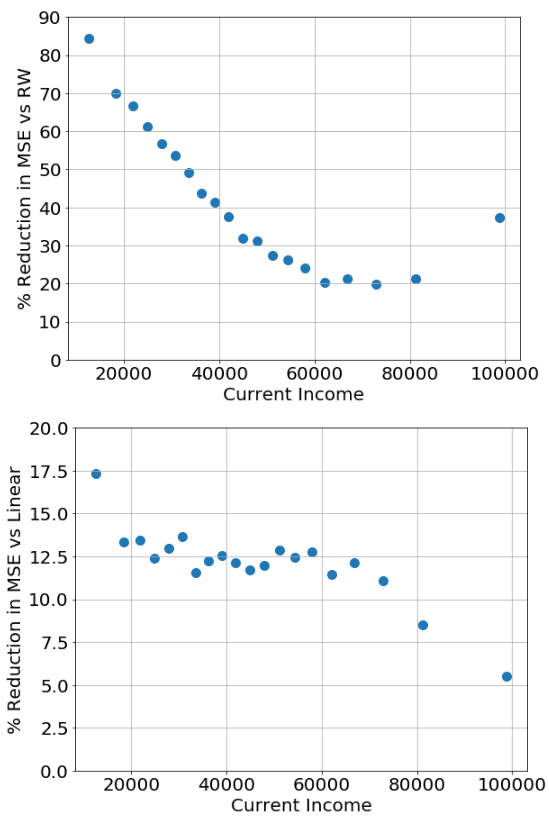

Figure 4: Percentage reduction in mean-squared error for predictions $h = 10$ year ahead across bins of current income. The top and bottom plots compare the gradient-boosted model to the random walk baseline and linear model respectively.

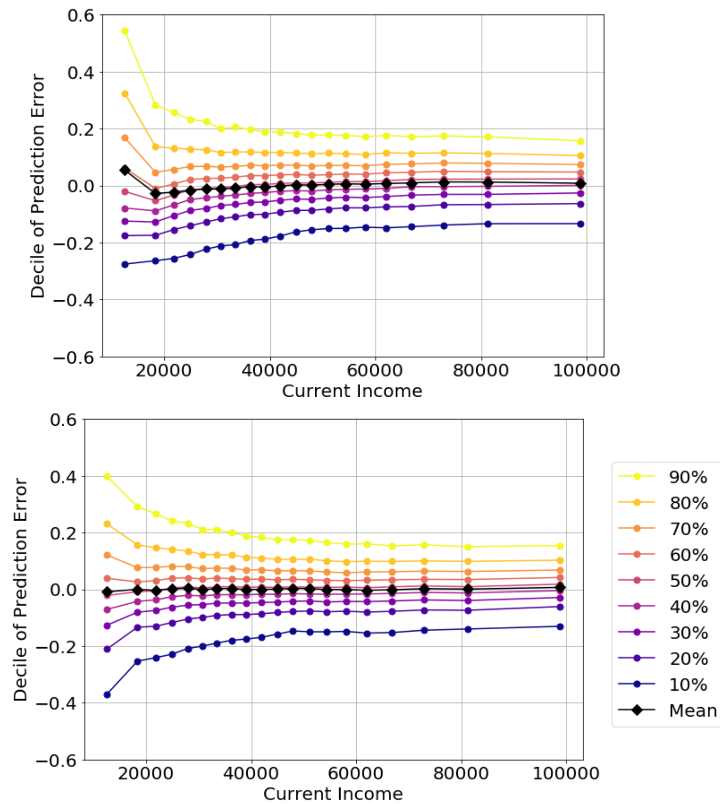

Figure 5: Deciles and mean of the distribution of prediction errors within bins of current income. The upper and lower diagrams plot the distribution of errors for the linear model and gradient-boosted model respectively.

The flexible model achieves nearly 20% improvement versus the linear model for individuals who make less than \$20,000 (2018 US dollars) a year.

By definition, conditional on any value of the features, our prediction errors should be mean-zero if we have achieved a good approximation of the conditional expectation.[3] We explore this in Figure 5, where we compare the distribution of prediction errors across current income for both the linear and gradient-boosted trees models. We use the same buckets of current income, but the y-axis now plots the deciles and mean of prediction error within each bucket. Note that the distribution of prediction errors for the linear model in the upper plot is asymmetric with non-zero mean, and with the most substantial deviation for households with current income less than \$20,000. The 90-10 interquantile range is only slightly smaller in the lower plot, but most importantly the distribution has approximately mean-zero everywhere, further validating our approximation of the conditional expectation.

[3]To see this, note that $\mathbb{E}[\, y_t - \mathbb{E}[\, y_t \,|\, \mathcal{I}_{t-1}] \,|\, \mathcal{I}_{t-1}] = \mathbb{E}[\, y_t \,|\, \mathcal{I}_{t-1}] - \mathbb{E}[\, y_t \,|\, \mathcal{I}_{t-1}] = 0$. Or more intuitively: if the prediction errors were not mean-zero conditional on a particular input, we could always improve the MSE by shifting all predictions for that input.

5 SHOCKS

With our approximations of the conditional expectations $\mathbb{E}[y_{t+h}|\mathcal{I}_t]$ in hand for all h and t, we can estimate the shocks $\Delta_t := y_t - \mathbb{E}[y_t|\mathcal{I}_{t-1}]$ and their persistence $\phi_{t,h} := \mathbb{E}[y_{t+h}|\mathcal{I}_t] - \mathbb{E}[y_{t+h-1}|\mathcal{I}_{t-1}]$ for every individual in every year of our sample. This produces a concrete artifact as output: income shock estimates attached to every observation that can be used in downstream economic research tasks. In this section, we use the shock data to provide an initial characterization of labor income risk in Iceland.

The distribution of total shocks Δ_t, is exactly equal to the distribution of prediction errors, but now we analyze them *substantively* instead of as a diagnostic tool for model fitting. From the bottom diagram in Figure 5, we can see that low-income households face a much wider distribution of shocks. The 10-30% quantile shocks and the 70-90% quantile shocks are all at least twice as large for individuals at the bottom of the income distribution compared to the middle and top. This is a substantial deviation from the classical autoregressive model discussed in Section 3.2 that predicts an equal amount of income risk across the income distribution. Furthermore, notice that if we had used linear regression for prediction, then from the upper diagram in Figure 5 we would have incorrectly concluded that low income individuals face much larger positive shocks than negative shocks.

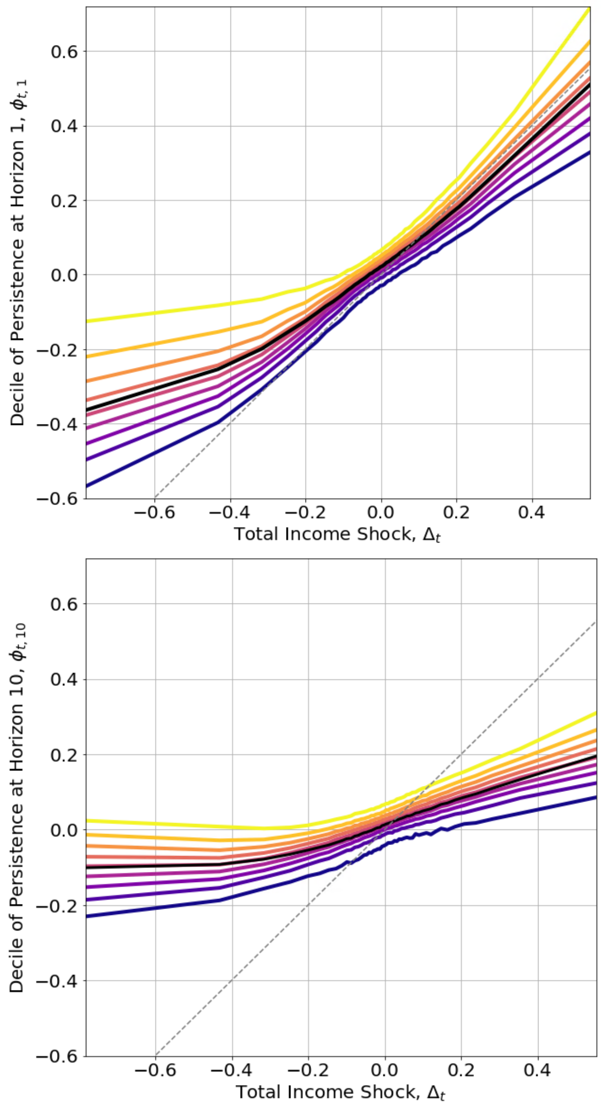

Figure 6: The upper and lower diagrams compare Δ_t and $\phi_{t,h}$ for the estimated income shocks computed using the predictions from our model on held-out samples for $h = 1$ and $h = 10$ respectively. We divide the obvservations into 50 bins according to Δ_t; the x-axis plots the mean value within each of these bins. The y-axis gives the 10% through 90% deciles of $\phi_{t,h}$ within these bins, each as a different line. We plot $y = x$ as a dashed line for reference to indicate perfect persistence.

Using our methodology, we can also assess how these shocks persist over time. Figure 6 plots the persistent shocks $\phi_{t,h}$ as a function of the total income shock Δ_t. The upper diagram shows the results for $h = 1$, and the lower for $h = 10$. We begin by summarizing some observations for the $h = 1$ case. First, notice the asymmetry between positive and negative income shocks, a result that mirrors recent findings of asymmetry in consumption responses [15]. For positive total income shocks, there is a clear and roughly linear relationship between the total shock size and the persistence one period ahead. A substantial and fairly consistent proportion of total income shocks are persistent. Negative income shocks, on the other hand, are typically less persistent on average and the heterogeneity in persistence for negative shocks (e.g. as represented by 90-10 interquantile range) also appears to be much larger. That is, the degree of persistence of negative income shocks varies more — especially for the lowest income individuals. Furthermore, as the total income shock becomes more negative, the relationship between the shock size and persistence appears less linear.

The degree of persistence drops off rapidly at longer horizons. The lower diagram of Figure 6 plots the deciles for $h = 10$ case; the y-axis now corresponds to the change in expected income 10 years into the future upon receiving the shock, $\phi_{t,10} = \mathbb{E}[y_{t+10}|I_t] - \mathbb{E}[y_{t+10-1}|I_{t-1}]$. Notice that the relationship between the total shock size and persistence 10 years into the future is much flatter, although still noticeably asymmetric.

These results contrast sharply with the AR(1) income process specification from Section 3.2, in which permanent income is perfectly persistent and so we should not see any drop over time. The classical model also does not predict a gap in persistence between positive and negative income shocks, nor does it predict any heterogeneity in the degree of persistence across the distribution of shocks. Each of these features of our shock series is a substantively interesting fact about labor income risk in Iceland that cannot be explained by income processes predominantly adopted in macroeconomic structural models.

6 DISCUSSION

6.1 Roles for Prediction in Social Science

Our work emphasizes an under-utilized role for prediction in the social sciences. While machine learning models cannot predict future life outcomes with high accuracy [49, 60], they can instead be used to approximate conditional expectations, and the distribution of prediction errors can be of scientific interest in its own right. In this sense, we join Lundberg et al. [42] in stressing the importance of clearly defining a statistical estimand.

To estimate a conditional expectation, we need to select the best predictor relatively from among all functions of the input features, a task for which supervised learning algorithms together with sample-splitting and cross-validation are well-suited. We can at least partially validate our model by checking for conditionally mean-zero prediction errors in held out data. Here flexible regressors like gradient-boosted trees play an important role, as simpler prediction models like linear regressions are observably mis-specified in our setting, as we illustrated in Figure 5. Our narrowly-scoped usage of these predictors contrasts with typical applied settings, where a predictor is trained from historical data, and then deployed in real-time on newly collected data that will not generally be drawn from the same distribution as the training data. We hope to have demonstrated the utility of our methodology by illustrating the substantial inequality in the size and persistence of labor income shocks in Iceland, especially for low income individuals.

6.2 What can we do with this shock series?

One benefit of our procedure is that we produce a concrete research artifact: a series of income shocks and their persistence for every individual in every year of our sample. These shocks are interesting in their own right for studying labor income risk, see our discussion above about the distribution of shocks over quantiles of current income, and the asymmetry and heterogeneity of shock persistence. However, the principle goal is to use these shocks in downstream scientific tasks. In this section, we briefly highlight directions for future work.

First, there is a large literature on scarring during business cycles. In the United States, individuals who first entered the labor market during or immediately before the Great Recession faced worse outcomes that persisted even after the economy recovered [58]. Because we condition on both age and year, we can directly assess the size and persistence profile of shocks that occur in the Great Recession in Iceland, which could provide valuable additional evidence on scarring.

Second, we can estimate the response of household consumption to these shocks. One approach to studying the consumption response would be to estimate the average derivative of consumption with respect to these shocks and their persistence over time. Furthermore, since our shocks are computed using the full heterogeneity across observables, we would be able to break down how these consumption responses differ across income, age, education, assets, etc.

Finally, structural macroeconomic models typically use a simple autoregressive model or first-order discrete Markov chain for the income process when modelling household behavior. Typically, the parameters of this income process are estimated separately, and then the macroeconomic model is calibrated using simulated draws. Our estimates of expected future income give us a way to potentially test these macroeconomic models directly with data, subject to the limitations described above on the difference between our predictions formed with tax data, and the private future expectations of individuals

ACKNOWLEDGMENTS

We thank Margrét K. Indriðadóttir, Hrafnhildur Arnkelsdóttir, Hildur Sigurðardóttir, and Jósef Sigurdsson at the Central Bank of Iceland for their extensive assistance with the tax data. DBS and EN were supported by the Alfred P. Sloan Foundation, grant G-2019-11255.

REFERENCES

[1] Rediet Abebe, Jon Kleinberg, and S Matthew Weinberg. 2020. Subsidy allocations in the presence of income shocks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 7032–7039.
[2] John M Abowd and David Card. 1986. On the covariance structure of earnings and hours changes.
[3] Krishna Acharya, Eshwar Ram Arunachaleswaran, Sampath Kannan, Aaron Roth, and Juba Ziani. 2022. Wealth dynamics over generations: Analysis and interventions. *arXiv preprint arXiv:2209.07375* (2022).
[4] Manuel Arellano, Richard Blundell, and Stéphane Bonhomme. 2017. Earnings and consumption dynamics: a nonlinear panel data framework. *Econometrica* 85, 3 (2017), 693–734.
[5] Manuel Arellano, Richard Blundell, Stéphane Bonhomme, and Jack Light. 2021. *Heterogeneity of Consumption Responses to Income Shocks in the Presence of Nonlinear Persistence.* Technical Report. Mimeo, University of Chicago.
[6] Hero Ashman and Seth Neumuller. 2020. Can income differences explain the racial wealth gap? A quantitative analysis. *Review of Economic Dynamics* 35 (2020), 220–239.
[7] Adrien Auclert and Matthew Rognlie. 2018. *Inequality and aggregate demand.* Technical Report. National Bureau of Economic Research.
[8] Scott R Baker. 2018. Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy* 126, 4 (2018), 1504–1557.
[9] Eric Bauer and Ron Kohavi. 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning* 36 (1999), 105–139.
[10] Richard Blundell, Luigi Pistaferri, and Ian Preston. 2008. Consumption inequality and partial insurance. *American Economic Review* 98, 5 (2008), 1887–1921.
[11] David Card, Jörg Heining, and Patrick Kline. 2013. Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly journal of economics* 128, 3 (2013), 967–1015.
[12] Christopher D Carroll. 1994. How does future income affect current consumption? *The Quarterly Journal of Economics* 109, 1 (1994), 111–147.
[13] Christopher D Carroll. 2001. A theory of the consumption function, with and without liquidity constraints. *Journal of Economic perspectives* 15, 3 (2001), 23–45.
[14] Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *American Economic Review* 104, 5 (2014), 141–147.
[15] Dimitris Christelis, Dimitris Georgarakos, Tullio Jappelli, Luigi Pistaferri, and Maarten Van Rooij. 2019. Asymmetric consumption effects of transitory income shocks. *The Economic Journal* 129, 622 (2019), 2322–2341.
[16] Lawrence Christiano, Martin S Eichenbaum, and David Marshall. 1987. The permanent income hypothesis revisited.
[17] Miles Corak. 2013. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives* 27, 3 (2013), 79–102.
[18] Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, David Sculley, and Yoni Halpern. 2020. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* 525–534.
[19] Mariacristina De Nardi. 2004. Wealth inequality and intergenerational links. *The Review of Economic Studies* 71, 3 (2004), 743–768.
[20] Mariacristina De Nardi and Giulio Fella. 2017. Saving and wealth inequality. *Review of Economic Dynamics* 26 (2017), 280–300.
[21] Mariacristina De Nardi, Giulio Fella, and Gonzalo Paz Pardo. 2016. *The implications of richer earnings dynamics for consumption and wealth.* Technical Report. National Bureau of Economic Research.
[22] Angus Deaton and Christina Paxson. 1994. Intertemporal choice and inequality. *Journal of political economy* 102, 3 (1994), 437–467.
[23] Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. 2021. Retiring adult: New datasets for fair machine learning. *Advances in neural information processing systems* 34 (2021), 6478–6490.
[24] Andreas Fagereng, Martin B Holm, and Gisle J Natvik. 2021. MPC heterogeneity and household balance sheets. *American Economic Journal: Macroeconomics* 13, 4 (2021), 1–54.
[25] Marjorie A Flavin. 1981. The adjustment of consumption to changing expectations about future income. *Journal of political economy* 89, 5 (1981), 974–1009.
[26] Milton Friedman. 1957. The permanent income hypothesis. In *A theory of the consumption function.* Princeton University Press, 20–37.
[27] Peter Ganong, Damon Jones, Pascal J Noel, Fiona E Greig, Diana Farrell, and Chris Wheat. 2020. *Wealth, race, and consumption smoothing of typical income shocks.* Technical Report. National Bureau of Economic Research.
[28] Kristopher S Gerardi, Harvey S Rosen, and Paul S Willen. 2010. The impact of deregulation and financial innovation on consumers: The case of the mortgage market. *The Journal of Finance* 65, 1 (2010), 333–360.
[29] Pierre-Olivier Gourinchas and Jonathan A Parker. 2002. Consumption over the life cycle. *Econometrica* 70, 1 (2002), 47–89.
[30] Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. 2015. *What do data on millions of US workers reveal about life-cycle earnings risk?* Technical Report. National Bureau of Economic Research.
[31] Hoda Heidari and Jon Kleinberg. 2021. Allocating opportunities in a dynamic model of intergenerational mobility. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* 15–25.
[32] R Glenn Hubbard, Jonathan Skinner, and Stephen P Zeldes. 1995. Precautionary saving and social insurance. *Journal of political Economy* 103, 2 (1995), 360–399.
[33] Statistics Iceland. 2023. Consumer price index. https://www.statice.is/statistics/economy/prices/consumer-price-index/
[34] Tullio Jappelli and Luigi Pistaferri. 2014. Fiscal policy and MPC heterogeneity. *American Economic Journal: Macroeconomics* 6, 4 (2014), 107–136.
[35] Greg Kaplan and Giovanni L Violante. 2014. A model of the consumption response to fiscal stimulus payments. *Econometrica* 82, 4 (2014), 1199–1239.
[36] Greg Kaplan, Giovanni L Violante, and Justin Weidner. 2014. *The wealthy hand-to-mouth.* Technical Report. National Bureau of Economic Research.
[37] Ronny Kohavi and Barry Becker. 1996. Adult data set. *UCI machine learning repository* 5 (1996), 2093.
[38] Lorenz Kueng. 2018. Excess sensitivity of high-income consumers. *The Quarterly Journal of Economics* 133, 4 (2018), 1693–1751.

[39] Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. 2020. Logarithmic regret bound in partially observable linear dynamical systems. *Advances in Neural Information Processing Systems* 33 (2020), 20876–20888.
[40] Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. 2021. Model learning predictive control in nonlinear dynamical systems. In *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 757–762.
[41] Thomas Liao, Rohan Taori, Inioluwa Deborah Raji, and Ludwig Schmidt. 2021. Are we learning yet? a meta review of evaluation failures across machine learning. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.
[42] Ian Lundberg, Rebecca Johnson, and Brandon M Stewart. 2021. What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review* 86, 3 (2021), 532–565.
[43] Richard A Meese and Kenneth Rogoff. 1983. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics* 14, 1-2 (1983), 3–24.
[44] Costas Meghir and Luigi Pistaferri. 2004. Income variance dynamics and heterogeneity. *Econometrica* 72, 1 (2004), 1–32.
[45] Costas Meghir and Luigi Pistaferri. 2011. Earnings, consumption and life cycle choices. In *Handbook of labor economics*. Vol. 4. Elsevier, 773–854.
[46] Zakaria Mhammedi, Dylan J Foster, Max Simchowitz, Dipendra Misra, Wen Sun, Akshay Krishnamurthy, Alexander Rakhlin, and John Langford. 2020. Learning the linear quadratic regulator from nonlinear observations. *Advances in Neural Information Processing Systems* 33 (2020), 14532–14543.
[47] Jonathan Morduch and Rachel Schneider. 2017. *The financial diaries: How American families cope in a world of uncertainty*. Princeton University Press.
[48] Makoto Nakajima and Vladimir Smirnyagin. 2019. Cyclical labor income risk. *Available at SSRN 3432213* (2019).
[49] Arvind Narayanan. 2019. How to recognize AI snake oil. *Arthur Miller Lecture on Science and Ethics* (2019).
[50] Pegah Nokhiz, Aravinda Kanchana Ruwanpathirana, Neal Patwari, and Suresh Venkatasubramanian. 2021. Precarity: Modeling the Long Term Effects of Compounded Decisions on Individual Instability. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 199–208.
[51] OECD. 2023. Conversion rates - exchange rates - OECD data. https://data.oecd.org/conversion/exchange-rates.htm
[52] Marios Papachristou and Jon Kleinberg. 2022. Allocating stimulus checks in times of crisis. In *Proceedings of the ACM Web Conference 2022*. 16–26.
[53] Jonathan A Parker, Nicholas S Souleles, David S Johnson, and Robert McClelland. 2013. Consumer spending and the economic stimulus payments of 2008. *American Economic Review* 103, 6 (2013), 2530–2553.
[54] Fabian T Pfeffer, Sheldon Danziger, and Robert F Schoeni. 2013. Wealth disparities before and after the Great Recession. *The Annals of the American Academy of Political and Social Science* 650, 1 (2013), 98–123.
[55] Inioluwa Deborah Raji, I Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst. 2022. The fallacy of AI functionality. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 959–972.
[56] Lydia Reader, Pegah Nokhiz, Cathleen Power, Neal Patwari, Suresh Venkatasubramanian, and Sorelle Friedler. 2022. Models for understanding and quantifying feedback in societal systems. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1765–1775.
[57] Yaniv Romano, Evan Patterson, and Emmanuel Candes. 2019. Conformalized quantile regression. *Advances in neural information processing systems* 32 (2019).
[58] Jesse Rothstein. 2021. The lost generation? Labor market outcomes for post great recession entrants. *Journal of Human Resources* (2021), 0920–11206R1.
[59] Claudia R Sahm, Matthew D Shapiro, and Joel Slemrod. 2010. Household response to the 2008 tax rebate: Survey evidence and aggregate implications. *Tax Policy and the Economy* 24, 1 (2010), 69–110.
[60] Matthew J Salganik, Ian Lundberg, Alexander T Kindel, Caitlin E Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M Altschul, Jennie E Brand, Nicole Bohme Carnegie, Ryan James Compton, et al. 2020. Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences* 117, 15 (2020), 8398–8403.
[61] Max Simchowitz, Ross Boczar, and Benjamin Recht. 2019. Learning linear dynamical systems with semi-parametric least squares. In *Conference on Learning Theory*. PMLR, 2714–2802.
[62] Kjetil Storesletten, Chris I Telmer, and Amir Yaron. 2004. Cyclical dynamics in idiosyncratic labor market risk. *Journal of political Economy* 112, 3 (2004), 695–717.
[63] Ludwig Straub. 2019. Consumption, savings, and the distribution of permanent income. *Unpublished manuscript, Harvard University* (2019).

A CROSS-VALIDATION AND MODEL SELECTION

As discussed in the text, we performed model selection and hyperparameter tuning using 5-fold cross-validation. We compared a variety of linear and tree ensemble models. For linear models, we compared elastic net regressions, varying the degree of regularization and the relative weight for the ℓ_2 and ℓ_1 norms. Interestingly, across all linear models, un-regularized linear regression consistently performed best, which probably reflects our large sample size relative to the number of features. For random forests and gradient-boosted trees, we varied the subsampling strategy, the number of trees, the maximum tree depth, the maximum leaf nodes, and the maximum number of features. For gradient-boosted trees, we also varied the optimization learning rate. We also considered training the same models but using only the first four and first five lags — these models performed worse than the tree-based models with six lags, but not by much. Across all horizons, gradient-boosted trees had the best mean-squared error in cross-validation.

Due to the high degree of dependence between observations on the same individual, we perform cross-validation at the individual level. However, some dependences are likely to remain across individuals. This is not necessarily a problem for our statistical procedure. We split individuals randomly into two subsamples and if we interpret the conditional expectation as being with respect to this sampling variation from a larger finite population, then dependence between individuals may not be an issue. However, this perspective might make interpretation more difficult, so we consider common depedences between individuals here. One obvious concern is domestic partners, whose labor income is very likely correlated. In future work, we plan on using tax identifiers to collect households into single units to ameliorate this problem. Another interesting source of dependence in labor income is when individuals work for the same firm. For example, see the discussion of firm fixed effects in Card et al [11]. We have matched firm data for individuals in Iceland, but only for the years 2000 and onward. We hope to explore the correlations between individuals working for the same firm in future work.