Agent-based models of the United States wealth distribution with Ensemble Kalman Filter

Yannick Oswald*[1], Keiran Suchak[2], and Nick Malleson[2]

[1]Institute of Sustainability and Geography, University of Lausanne
[2]School of Geography, University of Leeds

February, 2024

Abstract

The distribution of wealth is central to economic, social, and environmental dynamics. The release of high-frequency distributional data and the rapid pace of the complex global economy makes 'real-time' predictions about the distribution of wealth and income increasingly relevant. For instance, during the COVID-19 pandemic in spring 2020, the stock markets experienced a crash followed by a surge within a brief period, evidently reshaping the wealth distribution in the US. Yet economic data, when first released, can be uncertain and need to be readjusted – again specifically so during crisis moments like the pandemic when information about household consumption and business returns is patchy and drastically different from "business-as-usual". Our motivation here is to develop one way of overcoming the problem of uncertain 'real-time' data and enable economic simulation methods, such as agent-based models, to accurately predict in 'real-time' when combined with newly released data. Therefore, we tested two distinct, parsimonious agent-based models of wealth distribution, calibrated with US data from 1990 to 2019, in conjunction with data assimilation. Data assimilation is essentially applied control theory – a set of algorithms aiming to improve model predictions by integrating 'real-time' observational data into a simulation. The algorithm we employed is the Ensemble Kalman Filter (EnKF), which performs well with computationally expensive methods. Our findings reveal that while the base models already align well historically, the EnKF enables a superior fit to the data.

Keywords: Agent-based modelling, Data assimilation, Ensemble Kalman Filter, Wealth Inequality, Real-time prediction

*corresponding author: y-oswald@web.de

1 Introduction

The distribution of wealth is considerably more unequal than that of income or consumption across the population (Chancel et al. 2022). In the United States wealth inequalities are particularly dramatic when compared to other high-income countries. The top 1% of wealth owners in the population own 35%, the bottom 50% own approximately nothing at all or are even in debt. In a typical European country, say France, the top 1% 'only' own 25% of the wealth and the bottom 50% of the population roughly 5% (World Inequality Database 2024). The circumstances in the USA are in parts due to a high propensity for individuals to participate in the asset markets and a high degree of dynamism in business and the labour market. More than 60% of all adult Americans own stocks (Jones 2023) compared to, for instance, only around 18% in Germany (Aktieninstitut 2023) where wealth inequality is largely shaped through inheritance and low social mobility. This large inequality and dynamism in the US is significant because wealth is an enabler for consumption over time and evidently shapes power-dynamics in business and politics. It also determines environmental outcomes such as CO_2 emissions (Chancel 2022).

The distribution of wealth is shaped by political agendas, economic trends and, increasingly, crises. The most recent example for this is the market crash during the early COVID-19 pandemic and the subsequent stock-rally on technology and renewable energy companies. During the 2008 financial crisis, the impact on the wealth distribution in the USA was even more prominent (see Figure 1). While all groups lost wealth during the financial crisis, the poorest half of the population, measured by wealth, exhibited distinct dynamics and was pushed far below their wealth level of 1976 which for many meant accumulating debt. A lot of poorer and middle-class people defaulted on their mortgages and the stock market crashed too. It took a decade for the Bottom 50% to recover from this. The resilience of the wealth distribution during crisis moments requires improvement and therefore detailed monitoring also on fine-grained time-scales.

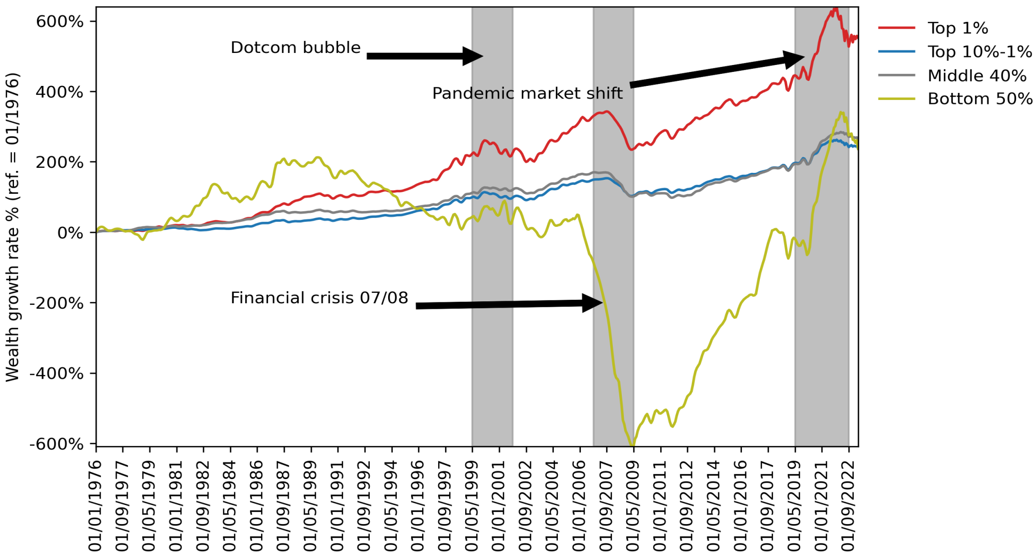

Figure 1: Wealth distribution over time

Note: Data from https://realtimeinequality.org/ with details in Blanchet, Saez, and Zucman (2022)

Historically, the measurement of wealth, and estimation of its distribution, are cumbersome and slow tasks. Piketty assembled the data for his famous work 'Capital in the 21^{st} century' over many years through historical tax records. Typically, the most detailed wealth data have been released or made available on an annual basis. Recently, however, motivated by the fast-paced crisis dynamics of the pandemic, Blanchet, Saez, and Zucman (2022) constructed monthly distributional wealth data – a step towards 'real-time' observation of the wealth distribution. The authors did so based on individual tax accounts and through rescaling of monthly macroeconomic aggregates, such as GDP. Ultimately, due to the increasing volatility and unpredictability of global economic dynamics, it is of interest to economists and policy makers to predict the distribution of wealth at high frequency in order to countersteer shock and negative trends. Moreover, yearly and quarterly GDP data are often quite uncertain when first released and regularly need to be readjusted after months, years, or even decades. On top of that, the definitions of macroeconomic metrics change over time (Semieniuk 2024). Therefore, as with GDP, high-frequency wealth data can be expected to be uncertain and to be readjusted over time. Therefore, observations alone are perhaps not reliable enough on their own to make accurate assessments about the state of the economy, especially so during crisis moments.

At the same time, several computational approaches to economics emerged whose intention is to capture the dynamic complexity of the economy better than traditional equilibrium-focused models do. One such approach is agent-based modelling (ABM). Agent-based models (ABMs) have been identified as one option to go beyond classical analytical economic models in the sense that they are a more natural analog to the discrete and interactive nature of the economy (Axtell and Farmer 2022) and they have been successfully applied across economic themes from housing markets to financial markets and many more. Historically, a lot of ABM work has remained at the conceptual level. Agent-based models are primarily used to describe stylized facts and overall patterns about isolated phenomena such as how a particular distribution of wealth emerges, how business cycles form and so forth. However, with increasing data availability, recently there is a trend to calibrate ABMs to data in detail and also to employ agent-based models for economic forecasting – one example being on credit risk and inequality by Papadopoulos (2019). With respect to COVID-19, infectious-disease-economy models have been employed to predict pandemic impacts on the economy (Silva et al. 2020). Moreover, there is a new trend towards large-scale macroeconomic agent-based models that aim to replicate entire economies, to construct a digital twin of an economy so to speak. Poledna et al. (2023) constructed such a model for Austria for example, with 8.8 million agents mapping the entire Austrian population and a large number of parameters determining the agent behaviour. Poledna et al. (2023) achieve parity with other standard economic forecasting methods such as state-of-the-art dynamic stochastic equilibrium models (DSGE) and econometric methods. However, even such sophisticated models produce forecasts with substantial uncertainty and one central estimate based on several hundred model ensemble members. Although the use of sophisticated ABMs might improve forecasting reliability further, no model will ever be accurate to a degree that is sufficient when facing entirely unforeseen scenarios that are not endogenous to the model, such as the COVID-19 pandemic and its repercussions.

Hence, due to the inherent uncertainties in both economic data and models, in this work we test the combination of economic agent-based models with data assimilation, specifically the Ensemble Kalman Filter algorithm, to overcome these fundamental limitations. Data assimilation is a set of algorithms that feed data in real-time into models and aim to optimally combine new observations with model predictions. It has a considerable history in meteorology and has contributed to the vast improvements in weather predictions over the last decades (Kalnay 2003), with some historic applications in engineering and indeed also in economics. In contrast to calibration, which sets the parameters of a model before a forecasting application, data assimilation updates the internal state of a model based on newly available data during the forecast. In our case we will apply the Ensemble Kalman Filter (EnKF), which is a variant of the Kalman Filter – a popular tool from optimal control – designed to specifically handle ensembles of model simulations. Most economic agent-based models are stochastic in character and produce model ensembles for forecasting purposes, so this type of data assimilation seems intuitively appropriate.

We have two goals in this paper:

(i) we study two distinct, parsimonious agent-based models, one borrowed from the literature, and one developed from scratch, and calibrate them to the USA wealth distribution over time;

(ii) for the first time, we integrate economic agent-based models with data assimilation as a proof-of-concept and in particular test the performance of the Ensemble Kalman Filter (EnKF).

Here we do not make use of an elaborate, high-dimensional model such as the one described in Poledna et al. (2023) for several reasons. First, as we will see, the American wealth distribution can be reproduced by very simple, low-dimensional agent-based models with a respectable degree of accuracy. Second, parsimonious agent-based models still play an important role in economic research and elucidate core principles and mechanisms of complex economies. Indeed one of the most important contributions of agent-based modelling is to explain complex patterns at the meso- and macro-level through very simple micro-foundations and elaborate on mechanisms of emergence. Third, it is reasonable to test and implement a first-use case of data assimilation with rather simple ABMs. First we need need to understand how to optimally integrate the two research strands, and evaluate the performance of the merger, before advancing to more complex models. Of course, it is expected that the trend towards large and high-dimensional agent-based models, including temporal and spatial granularity in the wealth distribution, continues. In the long-term, our approach might offer more advantages the larger a model becomes. Models which approach 'digital twin' character, that is they are more or less direct replicas of entire economies, and are fed with a lot of uncertain high-frequency data, are potentially suited best for data-assimilation-based control.

The remainder of this paper is structured as follows: In Section 2 we review related literature. In Section 3, we introduce our methods including the agent-based models applied, as well as the Ensemble Kalman Filter. In Section 4, we demonstrate results and lastly, in Section 5, we discuss our findings.

2 Related literature

2.1 Agent-based models of wealth dynamics

Besides ABMs, there are many economic modelling approaches that deal with inequality and wealth dynamics in particular. This includes various types of large computable general equilibrium models. Among new work worth mentioning is Blanchet and Martinez-Toledano (2023) and Blanchet, Saez, and Zucman (2022) who apply stochastic differential equations to capture the evolution of wealth. However, ultimately our goal is to integrate an economics ABM with a data assimilation approach (the Ensemble Kalman Filter, EnKF) and therefore we do not discuss in detail the merit of other modelling approaches compared to agent-based models.

In any case, agent-based modelling is a tool that is increasingly used in economics (Axtell and Farmer 2022) and so it is for the analysis of economic inequality in particular. A popular class of agent-based inequality models is from the econophysics literature. Dragulescu and Yakovenko (2000), Drăgulescu and Yakovenko (2001), and Yakovenko and Rosser Jr (2009) show that the exponential-like distribution of income and wealth in the USA can originate from a simple model of agent-exchange. Here agents randomly bump into each other and exchange a unit of 'wealth', much like molecules do exchange energy in an ideal gas. Hence this strand of literature also has been termed 'Statistical mechanics of money'. The sociologist Angle (1986) also developed a similar model 15 years earlier. It has however been pointed out that this class of models fails to represent any meaningful economic ontology, in the sense that they do not consider any actual market mechanisms and so forth, and also fail to reproduce the essential fat tails of the wealth distribution (Lux 2005). We do borrow heavily from this class of model in our own attempt to model wealth inequality since we also implement a simple exchange mechanism. But our approach diverges in an important way – the exchange that two agents have is constraint not only by their own quantity of wealth but also by their position in a social network which immediately gives rise to a more realistic distribution. The interplay between social network structure and the historical as well as contemporary evolution of inequality is an active area of research (Chiang 2015; Mattison et al. 2016). Hence it seems reasonable to assume such a model.

A model we use as a starting point to describe the evolution of wealth and to integrate with data assimilation is by Vallejos, Nutaro, and Perumalla (2018). This model is called agent-based by the authors although individual-based might be a better description. We argue so because the model has little to no agent-interaction and also the agents actually do not apply any 'agency' in the sense that they execute actions according to specified rules. Instead, the core-mechanism of this model is an aggregate economic growth function that, in discrete time, generates a certain 'pot of resources' that is distributed among the agents according to how much wealth they own before already. So why even consider it? The model does fit surprisingly well to the historical time-series of wealth inequality, including its Pareto-tail, and exhibits some computational features of ABMs such as discrete individuals. Therefore it serves well as first test-bed to apply data-assimilation to.

2.2 Data assimilation and economics

Data assimilation is essentially applied control theory. It is a set of algorithms that optimally combine observations with model outputs to predict the future state of any given system. It has been largely developed in the atmospheric sciences and has been hugely successful in making weather predictions better (Kalnay 2003) and also has manifold applications in control engineering.

It is key however to understand that data assimilation, and especially Kalman Filter (KF) approaches, are not new to economics. On the contrary the first applications date back several decades (Athans 1974; Vishwakarma 1970), and many more such applications that have emerged over the following decades (Inglesi-Lotz 2011; Munguia, Davalos, and Urzua 2019; Pasricha 2006; Schneider 1988; R. I. Thamae, L. Z. Thamae, and T. M. Thamae 2015). There are, however, currently none in combination with agent-based models.

Outside of economics, there is some preliminary work that integrates data assimilation with ABMs. These applications focus on urban dynamics and indoor or outdoor pedestrian dynamics as well as transport applications in particular (Clay et al. 2021; Xiaolin Hu 2022; N. Malleson et al. 2020; Tang and Nick Malleson 2022; Ternes et al. 2022; Wang and X. Hu 2015; Ward, Evans, and N. S. Malleson 2016). Urban systems evolve in a fast-paced manner but are monitored by many different sensors such as cameras, air pollution sensors, traffic counters, etc. To make predictions about the system – such as crowd density, traffic congestion or pollution levels – in 'real-time', data assimilation algorithms are designed to optimally combine model predictions with incoming sensor data to use "all the available information, to determine as accurately as possible the state of [the system]" (Talagrand 1991). There is an argument to be made, that with increasingly large models, like for instance Poledna et al. (2023), an increasing number of parameters, an increasingly large numer of possible system states that a model can exhibit, and especially an increasing volume of economically-relevant data available in real-time, now is the time, more than ever, to prepare data assimilation as a method for the optimal control integration between models and data in economics, too. This is where our work begins.

3 Methods

Here we described the different methodological components that make up the overall process. In section 3.1, we provide a brief overview of the data we employ. In sections 3.2 and 3.3 we describe the agent-based models that we use, in Section 3.4, we elaborate on model calibration to historical data, in Section 3.5 we explain the Ensemble Kalman Filter (EnKF). To begin with, however, we first outline the dataset used.

3.1 Data

We employ a monthly time-series of the wealth distribution constructed by Blanchet, Saez, and Zucman (2022) as our observational data. With respect to the evolution of the wealth distribution, the monthly interval can be considered 'high-frequency' and a step towards 'real-time' observation since otherwise only yearly intervals had been available. The authors estimated the time-series based on individual tax accounts and through rescaling of monthly macroeconomic aggregates, such as GDP. This data is essentially a granular extension of the data that Piketty described in "Capital in the 21st century". The data features several wealth groups and we focus on the following disjoint groups: the top 1%, the next 9%, the middle 40% (50th to 90th percentile) and the bottom 50%. While Blanchet, Saez, and Zucman (2022) report respectable accuracy for those groups compared to other estimation methods (meaning the standard deviation of a fit against another method for lesser frequency data is usually within 5% of the mean wealth of a group), the data itself is constructed based on various modelling assumptions and hence can be regarded as inherently uncertain. For example Auten and Splinter (2019) arrive at substantially different estimates of income inequality compared to Blanchet, Saez, and Zucman (2022) but they did not yet account for wealth inequality as well. In any case, in principle large uncertainties are no problem for us, as this exactly why we work on this in the first place and we will represent data uncertainties further down in our hybrid-ABM-data-assimilation exercise. For simplicity we might regard the estimation of Blanchet, Saez, and Zucman (2022) also as a "ground-truth", that is the actual values of wealth inequality in reality which is necessary in order to know how well our model-data-assimilation approximates "reality".

3.2 Agent-Based Model 1

We adapt the model created by Vallejos, Nutaro, and Perumalla (2018). For full details of the model parameters and calibration the interested reader can refer to the original study. Although simple, the model is able to reproduce the major wealth distribution trends as shown extensively in Vallejos, Nutaro, and Perumalla (2018).

At an aggregate level any economy might be represented by its measure of output or wealth and its growth rate. This also holds for individuals. That is if we denote wealth as x and the constant growth rate per unit of time as g, a two-dimensional time-discrete model of any i agent's wealth can be given by (1) and (2):

$$x_{i,t} = x_{i,t-1} + x_{i,t-1} \cdot g_{i,t-1} \tag{1}$$

$$g_{i,t} = f(x_{i,t-1}) \tag{2}$$

where, following Vallejos, Nutaro, and Perumalla (2018), the growth rate of wealth is a function of wealth itself. Here $f(x_{i,t})$ is determined via the probability that an agent receives an additional unit of wealth. This is the defining feature of the entire model. Agents' trajectories are solely determined by a probability to receive more wealth. Vallejos, Nutaro, and Perumalla (2018) define this probability ψ as the share of wealth that an agent owns out of total wealth, where $p(w)_{i,t}$ denotes the wealth of agent

i at time point t:

$$\psi(i,t) = \frac{p(w)_{i,t}}{\sum_{n=1}^{k} p(w)_{i,t}} \tag{3}$$

Note that $p(w) = w^{\beta}$, so the parameter β is an exponent which determines how "social power" to receive additional wealth scales with wealth itself. The final major model assumption is that the economy grows exponentially; $\Omega(t) = \Omega(0)e^{\lambda t}$ where Ω is the wealth of agents summed up and the growth per time step $\Omega(t) - \Omega(t-1)$ is divided into n different discrete pieces of the 'growth-pie' and allocated to the agents via the above outlined probability rule.

3.3 Agent-Based Model 2

The second model, we developed from scratch. It is likewise an agent-based model of wealth distribution, although this new model incorporates agent interactions via a network. As the ability to model interactions is one of the key features of the methodology (Axtell and Farmer 2022; Gilbert 2008), we use this model because it is more characteristic of ABM in general. Although its ability to predict trajectories is slightly lower than the model by Vallejos, Nutaro, and Perumalla (2018), this is no detriment as our principal goal is to use the EnKF in combination with an imperfect ABM whilst still producing output that is close enough to reality to allow an appropriate case study. There is a relatively rich econophysics literature studying the wealth distribution originating from network dynamics (Di Matteo, Aste, and Hyde 2003; M.-B. Hu et al. 2008; Ichinomiya 2012; Lee and Kim 2007; Souma, Fujiwara, and Aoyama 2003), including work that shows that scale-free networks lead to power distributions – a fact we utilize here. Yet it is important to note that our goal is not to deliver a fundamental contribution to economic theory but to conduct a second test for the EnKF. Hence the model assumptions and dynamics may be considered a useful toy model rather than sophisticated economic dynamics.

The agents in the model are located on a Barabási–Albert (BA) network. Thus the model is generated via a preferential attachment algorithm and produces a power law distribution of node degree. The power-law distribution is commonly expressed in the form $p(k) \sim k^{-\gamma}$, where $p(k)$ is the probability of a node in the network having degree k, and γ is the exponent of the power law distribution. In theory $\gamma = 3$ (Albert and Barabási 2002), but given our small network sizes we find $\gamma \approx 2$ when measured empirically from the generated network. The only free parameter in a BA-graph is m which determines the minimum node degree of new nodes added to the network during network construction. We set $m = 2$.

The assumption behind the network is that that economic exchange between agents is constrained by social structure. Abstracting beyond households and firms, in reality, not everyone trades with everyone and some individuals have much more access to trade and resources than others. In the following, we then initialize the agents according to a Power-Lognormal (PLN) distribution and associate the wealthiest agents with the highest node degree. Therefore, the richest agent has the highest node degree, the poorest agent the least.

We assume that every agent's wealth $w_{i,t}$ grows at every time step by a global fixed rate, so $w_{i,t} = w_{i,t-1} * (1+g)$. The core mechanism for wealth distribution dynamics is trade. Each agent trades at every time steps with all its network neighbors. Therefore, there are two trades between agents at each time step. If agent i initiates trade with j, symmetrically j initiates trade with i. The agents have a willingess-to-risk (WTR) which is the the fraction of their own wealth that they offer up to loose in a trade and which we denote η . This parameter is drawn from a uniform distribution in the interval (0,u) where u is maximally equal one. The trade amount between two agents i and j, denoted γ for fraction, is then the minimum of both quantities offered, so:

$$\gamma(i,j,t) = \min(w_{i,t}\eta_i, w_{j,t}\eta_j) \tag{4}$$

This is a reasonable first model assuming rational agents that offer only as much as they can afford but in line with diverse risk preferences. Moreover, the chance for any agent i to win the trade is determined by its network position, that is its node degree D_i subject to an exponent c. The latter lets us control how strong the difference in likelihood between two agents is due to their different network position. But generally, a more central agent is more likely to win, representing some kind of 'social power':

$$P_{win}(i,t) = \left(\frac{D_i}{D_i + D_j}\right)^c + \varepsilon_t \tag{5}$$

Here ε_t is an error term drawn from a truncated normal distribution.

Lastly, and importantly, agents adjust their WTR based on whether they win or lose or the trade. If they lose, they lower their WTR, if they win they increase it. Therefore the agents are adaptive in nature. We call this parameter adaptive sensitivity or s (see Table 2).

This model converges to a steady state or equilibrium because over time when $t \to \infty$, the WTR of an agent who has on average a lower node degree than its neighbors goes towards zero, while the WTR of agents that have on average higher node degree than their neighbors tends towards one. Hence at some point the trade comes to a halt because some agents are not willing to engage in it anymore. The equilibrium prediction on the distribution of wealth is plausible because qualitatively it predicts in line with the US wealth data that the top 10% of the population hold far more than half of all wealth, while the bottom 50% of the population do not hold any wealth. If the adaptive WTR is taken out of the model, in the long-run all wealth ends up with the top 10%, which is less plausible. Section 3.4 illustrates the accuracy of the model in detail.

3.4 Model calibration and benchmarking

First we initiate the model with the appropriate distribution of wealth at time point $t = 0$. We choose the period of January 1990 to December of 2018 as our benchmark period to test the models and the EnKF. This is because there are several interesting crisis moments that occur during this period (e.g. financial crisis 07/08) that provide a useful test for the methods, but at the same time the data are likely more reliable than newer estimates that arose during the pandemic and post-pandemic period and hence more appropriately taken as "ground truth". Initially we fit the agent wealth distribution to the wealth distribution of 1990 to initialize the model. To do so we employ a Power-lognormal distribution, that is a lognormal distribution with pareto tails, from the Python 'scipy' library and calibrate the parameters so it fits the observations in January 1990 optimally. This distribution does not perfectly describe wealth inequality but, as our calibration shows, it does capture the tail and centre of the distribution and allows for a convenient random draw for agent initialization. We initialize 100 agents per simulation which is, admittedly, a rather small scale but we aim for efficiency to test the EnKF algorithm.

Both models simulate the average wealth growth rate in the USA from 1976-2023 which is 2.5% according to Blanchet, Saez, and Zucman (2022) as key parameter for driving dynamics. We list all the key parameter specifications in Table 1 and Table 2. Parameter c in model 2 is set to 1 for simplicity but we also find that this produces some plausible degree of convergence.

Parameter	Unit	Value
Avg. growth rate	$\%yr^{-1}$	2.5
Start yr	yr	1990
Pop. size	dmnl	100
β	dmnl	1.3

Table 1: Model 1 key parameters names and their values

Parameter	Unit	Value
Avg. growth rate	$\%yr^{-1}$	2.5
Start yr	yr	1990
Pop. size	dmnl	100
WTR	dmnl	$\in$ (0,0.1)
s	dmnl	0.02

Table 2: Model 2 key parameters names and their values

In all computational experiments, we employ a mean absolute error metric (MAE), defined in equation (6) to evaluate the performance of the models as well as that of the EnKF with respect to the 'ground truth'. The ground truth is defined as the raw data given by Blanchet, Saez, and Zucman (2022) and assuming no uncertainty in the data. Note that the EnKF creates a number of distinct instances of the model and executes them simultaneously (full details are forthcoming in Section 3.5). Together these model instances are called an 'ensemble' and an 'ensemble member' refers to one of those instances. Therefore in equation (6), m_{ijt} is the prediction of the ith ensemble member for the jth wealth group at time point t, and d_{jt} is the ground truth for the jth wealth group. We take the absolute value difference and sum over the wealth groups and then over the ensemble members and finally average across ensemble members.

$$MAE_t = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \sum_{j=1}^{4} |m_{ijt} - d_{jt}| \tag{6}$$

Figure 2 shows the calibration results. In Figure 2 panel (A) and (B) both models are depicted with one singular run against the data to illustrate their archetypal behaviour. We see that Model one produces more "linear" trajectories while Model two produces non linear trajectories but tending towards the above mentioned equilibrium. Both models exhibit different strengths and weaknesses. Model one in general exhibits a closer fit to the upper end of the distribution but Model two matches the zero percent wealth share of the bottom 50% more accurately. However, in reality, at times adults from the bottom 50% may exhibit negative wealth which in both models does not occur for simplicity. Keep in mind that all the wealth data is real wealth per adult person. In Figure 2 panels (C) and (D) we depict the uncertainty with confidence intervals across 30 simulation runs. In both models, the top 1% are by group exhibiting the most variation across simulations. The high uncertainty in this wealth group corresponds to the high uncertainty of the wealth among the very wealthy. Finally, Figure 2 panel (E) depicts the error as defined in Equation (6). Model 2 exhibits clearly higher error but its error stabilizes in the long run while the Model 1 error is relatively constant over time. For our purposes, testing the EnKF with agent-based modelling in economics, it is of advantage to work with both models so that we can find out whether the EnKF works under both circumstances equally well.

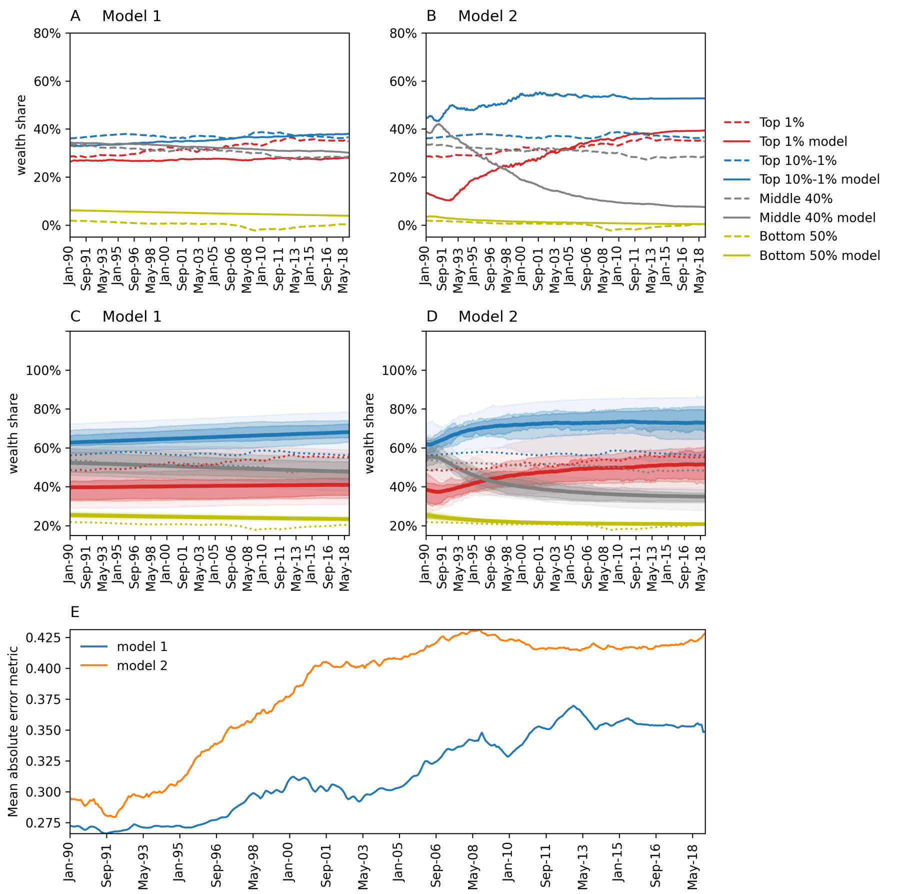

Figure 2: Models' ensemble error compared to data

3.5 Ensemble Kalman Filter

In the following section, we outline the Ensemble Kalman Filter (EnKF) and its application to our agent-based wealth distribution model. Note that a simple example is included at the end of the section to aid understanding of the method. The Ensemble Kalman Filter is a Monte-Carlo variation of the Kalman Filter (Kalman 1960) that is particularly useful in situations where the underlying model is complex and non-linear (Evensen 2003) — as is the case with agent-based models. Compared to the original Kalman Filter, the EnKF has several advantages:

Computational Efficiency: the Ensemble Kalman Filter uses an ensemble of state estimates that can be propagated independently in parallel, providing the ability to derive covariance matrices in a more computationally efficient manner for large state spaces;

Non-linearity: it handles non-linearity more effectively as the state covariance matrix no longer needs to be forecast by applying a linear operator, but instead can simply be generated as a sampling covariance.

The basic idea behind the EnKF is to represent a system and its associated uncertainty using a collection (or "ensemble") of possible state estimates, rather than just a single estimate as would be the case with the original Kalman Filter. These ensemble members represent different possible evolutionary paths of the system, based on different assumptions or uncertainties in the model. At each time step, the EnKF uses the model to predict the state of the system based on the current ensemble of state estimates (termed the 'predict' step). At time steps when observations of the system are provided, it then compares these predictions to actual observations of the system and adjusts the ensemble to better estimate the system state (the 'update' step). The key to the effectiveness of the EnKF (and data assimilation approaches more generally) is that it uses information from both the model and the observations to provide a best estimate of a system. This allows it to handle situations where the model is imperfect or incomplete, and where there may be substantial uncertainties in the observations.

The ensemble of system state estimates, $\mathbf{X}$, that consists of a number of individual realisations of the model, $\mathbf{x}$, is defined as:

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] = [\mathbf{x}_i], \quad \forall i \in \{1, 2, \dots, N\}, \tag{7}$$

where N is the number of ensemble member models. The mean state vector, $\bar{\mathbf{x}}$, can be found by averaging over the ensemble:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i. \tag{8}$$

Similarly, we have a collection of data vectors, $\mathbf{D}$, which are considered to be "nearly-true":

$$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_N] = [\mathbf{d}_i], \quad \forall i \in \{1, 2, \dots, N\}. \tag{9}$$

These "nearly-true" data vectors are generated using a simple stochastic rule; each data vector consists of the sum of the original observation $\mathbf{d}$, and a random vector, ϵ_i:

$$\mathbf{d}_i = \mathbf{d} + \epsilon_i, \quad \forall i \in \{1, 2, \dots, N\}. \tag{10}$$

The random vector is drawn from an unbiased normal distribution:

$$\epsilon \sim \mathcal{N}(0, \mathbf{R}), \tag{11}$$

where $\mathbf{R}$ is the data covariance matrix which is defined by the uncertainty in the observations. As with the model state, the mean data vector, $\bar{\mathbf{d}}$, can be found by averaging over the data ensemble:

$$\bar{\mathbf{d}} = \sum_{i=1}^{N} \mathbf{d}_i. \tag{12}$$

Given the above framework, the data assimilation is made up of the predict-update cycle (outlined in Figure 3), with the calculation of an updated state ensemble, $\hat{\mathbf{X}}$, being undertaken on the basis of the following equation:

$$\hat{\mathbf{X}} = \underbrace{\mathbf{X}}_{\substack{\text{agent}\\\text{states}}} + \underbrace{\mathbf{K}}_{\text{weight}} \underbrace{(\mathbf{D} - \mathbf{H}\mathbf{X})}_{\text{perturbation}}, \tag{13}$$

where $\mathbf{H}$ is the observation operator and $\mathbf{K}$ is the Kalman gain matrix. The role of the observation operator is to transform the state vectors between the form in which we store state variables (in our case, wealth per agent) and the form in which they are represented in observations which is a macroeconomic time-series (namely wealth per wealth group). An important clarification is that based on (13) we only update the state of the model, that is the agent wealth variable, but not parameters determining the model state. In principle one can also use data assimilation to learn the underlying parameters but for our case study we focus on the state variables to keep it relatively simple.

Note here that we do not update our state covariance matrix, instead generating a sample covariance matrix based on the state ensemble, $\mathbf{C}$:

$$\mathbf{C} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

The Kalman gain matrix, $\mathbf{K}$ is given by

$$\mathbf{K} = \mathbf{C}\mathbf{H}^T \left(\mathbf{H}\mathbf{C}\mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{14}$$

in which $\mathbf{C}$ is the sample state covariance (calculated based on the state ensemble, $\mathbf{X}$), and $\mathbf{R}$ is the observation covariance. We can consider $(\mathbf{D} - \mathbf{H}\mathbf{X})$ in Equation 13 to be the proposed perturbation to the ensemble states, and the Kalman gain matrix, $\mathbf{K}$, to be the weight given to this perturbation (just as in a standard Kalman Filter) based on the relative uncertainties in the state ensemble and the observations. When the uncertainty in the observations is low in comparison to the uncertainty in the model state, the gain increases, and consequently the model state receives a larger perturbation from the provided data; conversely, when the uncertainty in the observations is high in comparison to the uncertainty in the model state, the gain decreases, and consequently the model state receives a smaller perturbation from the provided data. Figure 3 provides an overview of the flow of the combined algorithm using the EnKF with the respective agent-based model.

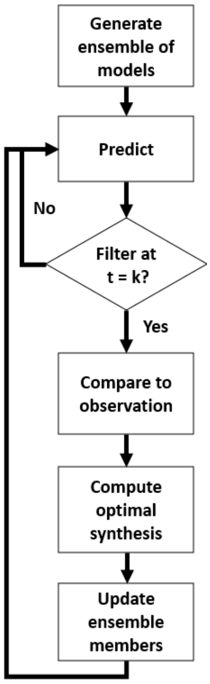

Figure 3: Ensemble Kalman Filter algorithm overview. Note that t is the model timestep and k indicates whether t is an *update* step (i.e. a timestep when the EnKF updates the model state in response to new data).

Before outlining the results, it is useful to illustrate briefly how the agent states, observations and the observation operator are organized in matrix form. Assume a wealth agent-based model with only three agents, one rich one and two poor ones and three ensemble simulations. The exact mechanics of the model do not matter because we are interested in understanding how the EnKF alters the output and the state of the model. Moreover let the observations come in two groups, rich and poor. We also assume particular values, so the numbers in the matrices are the wealth if not specified otherwise. Then the entire ensemble of model states and the observation can be depicted as follows. Here $\mathbf{x}_i$ are the ensemble members and the dashed lines indicate the agent groups that correspond to specific observation wealth groups.

$$\begin{array}{c c} & \begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{matrix} \\ \begin{matrix} \text{Agent}_1 \\ \text{Agent}_2 \\ \text{Agent}_3 \end{matrix} & \begin{pmatrix} 150 & 170 & 130 \\ 10 & 15 & 5 \\ 5 & 5 & 5 \end{pmatrix} \end{array} \tag{15}$$

$$\text{Obs.}_{\cdot i} = \mathbf{d_i} = \begin{pmatrix} \textit{rich} \\ \textit{poor} \end{pmatrix} = \begin{pmatrix} 150 \\ 11 \end{pmatrix} \tag{16}$$

Now the observation operator, $\mathbf{H}$, compresses the 3×3 matrix of agent states to a 2×2 matrix so it matches the observation dimensions of some $\mathbf{d_i}$ and $\mathbf{D}$. If there is more than one agent within a group, as in the case of the poor agents, then $\mathbf{H}$ averages over the state variable. In our case this operator effectively bridges the micro-economic level and the macro-economic level via aggregation.

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix} \tag{17}$$

4 Results

4.1 Experiment 1: Medium filter frequency vs. non-filtering

The first experiment compares the EnKF-supported models with the baseline performance depicted in Figure 2 employing a medium filter parameter setting. The model parameters are set as in Table 1 and Table 2. For the uncertainty in the observations we make pragmatic assumptions as it is not clear how uncertain the data are. In Blanchet, Saez, and Zucman (2022) there is an analysis of predicted vs. actual growth rates when testing their method for estimating real-time wealth inequality against established "slower" wealth accounting methods. They report generally low standard deviations of between 4% and 6% of the mean value depending on the exact wealth group considered. For the sake of simplicity, we assume a standard deviation of 5% of the mean value for observations for each wealth group which is well in line with their estimates. We do test varying uncertainty assumptions in section 4.5, but 5% is our default assumption if not specified otherwise.

For this experiment, we apply the filter every 30^{th} time step using 100 ensemble members and models with 100 agents. We do not use a smaller number of ensemble members for this experiment because as we will find later in, Section 4.3, that this influences the reliability and stability of the results (illustrated in Figure 6). Figure 4 outlines the results. Panel E demonstrates that, as expected, the error between ground truth and model is substantially reduced with the use of the EnKF. Panels A and B demonstrate, for illustrative purposes, single simulations of models 1 and 2, chosen at random, and how the EnKF affects them. Panels C and D illustrates the same but for the ensemble of all 100 members (solid lines show the mean of all ensembles). Considering panels C and D, it is apparent that the first correction step, between Sep-91 and May-93, makes a substantial change to the trajectory of both models. We can observe the following:

- The bottom 50% are corrected down towards their actual true state value which is near 0% of wealth.
- The top 1% are corrected upwards to reflect a higher proportion of wealth ownership than the models predicted.
- Interestingly the proportion of the top 10%–1% and middle 40% of agents are not corrected as well as other groups. The possible causes of this finding will be discussed in Section 5.

The important lesson from this first experiment is that, overall, the error over time remains substantially lower than when using no EnKF. For both models the error, relative to the ground truth, at each time point is reduced around 40% as can be seen in 4 panel E.

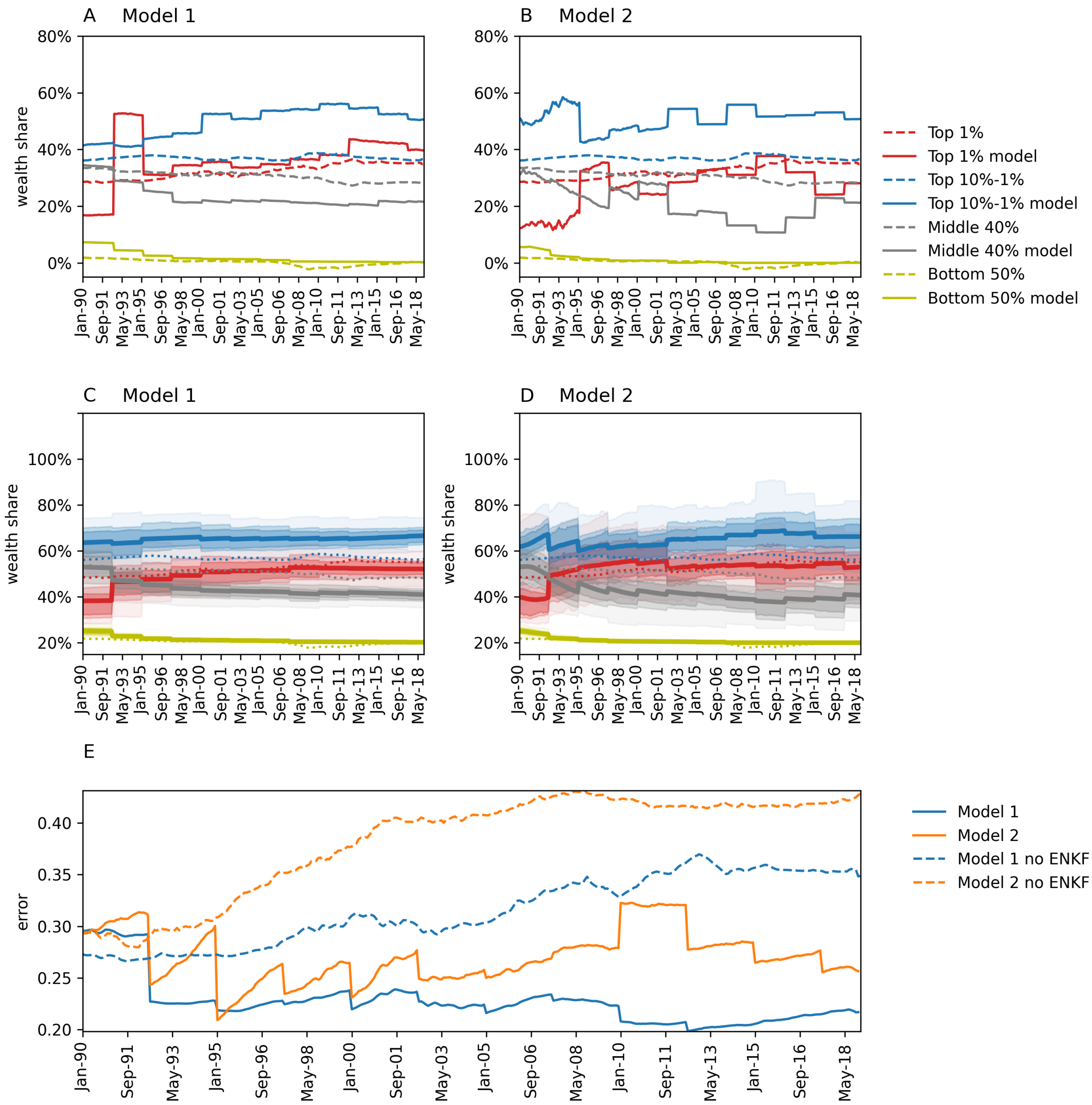

Figure 4: Models' ensemble with EnKF

4.2 Experiment 2: Variation of filter window size

Recall that the filter 'frequency' refers to the number of model time steps (in this case months) between assimilation episodes. Plausibly we would expect that higher filter frequency should result in an overall better estimate of the true state of the system as the models are not required to predict for too long without any support from the data. Here we test this assumption by varying the filter frequency and summing the error over time. As depicted in 4 panel E, this means the area under the curve S or formally adding a third summation term to Equation 6, so $S = \sum_{t=1}^{t} \left(\frac{1}{n}\right) \sum_{i=1}^{n} \sum_{j=1}^{4} |m_{ijt} - d_{jt}|$. To keep this experiment computationally efficient, we limit the time horizon to the first three years, so 1990-1993, again with 100 agents, fix the ensemble size at 10 and repeat each filter frequency 20 times. We test filtering every 2^{nd} up to every 100^{th} time step. We do not test filtering at every time step as the model ensemble variation is reduced to zero and consequently the EnKF always decides that fully following the model estimate is optimal. But actually this way the combined estimate diverges from the ground truth while behaving like a single unadjusted model run.

We quantify two further metrics for this experiment – the elasticity of the error with respect to the filter frequency and the coefficient of variation across experiment repetitions (one experiment parameter setting including all its repetition represents one the box plots in Figure 5). In economics, the elasticity of a variable with respect to another is $E_{x,y} = \frac{\%\Delta x}{\%\Delta y}$, where in our case x is the error and y is the filter frequency. The average error elasticity for model one is ~ 0.42 and for model two ~ 0.33. This means for example for model one that the error increases by 0.42% for every 1% increase in the filter window length. The the coefficient of variation of each experiment is plotted below the corresponding boxplot. Generally it is a little higher for Model 1 at small filter windows and does not diminish as much with filter window size as the mean error.

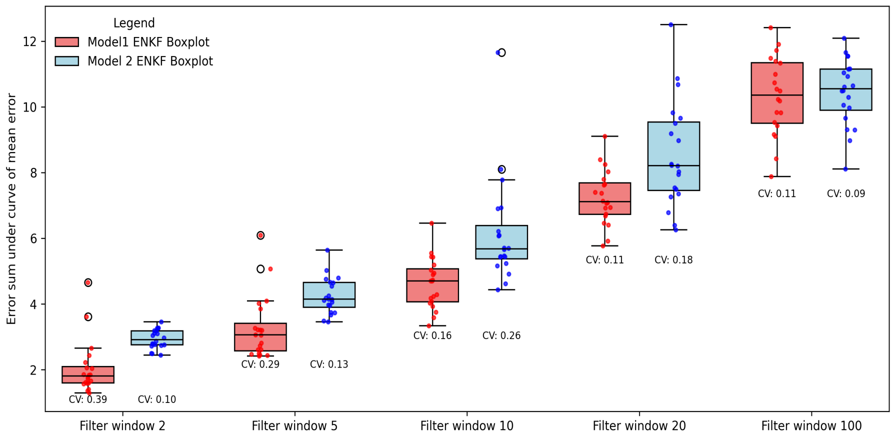

Figure 5: Filter frequency influence on overall error

4.3 Experiment 3: Variation of ensemble size

Next we isolate the effect of the ensemble size on the EnKF-model performance. Again we quantify the error using the error sum under curve as in 4 panel E, choose a short time horizon (1990-1993), and fix the filter frequency at 10 time steps. We estimate the error elasticity with respect to ensemble size, which is approximately -0.2 for both models but with strongly diminishing returns – an ensemble of 30 simulations works nearly as well as one of 100. The main finding here is that smaller ensembles produce substantially less reliable outcomes. The coefficient of variation across experiment repetitions is 0.38 for model one and 0.27 for model two when the ensemble size is 5 and only 0.03 for both models when the ensemble size is 100. At an ensemble size of 30 it is around 0.07. This means the EnKF should not be employed with ensemble sizes smaller than 30 at least with respect to our case study. In other words, larger ensemble sizes in model and observations gauge a more reliable estimator of the true state of the system.

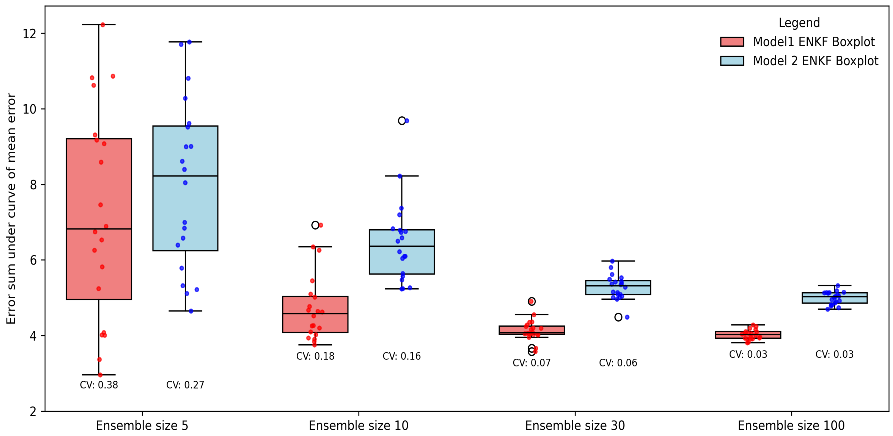

Figure 6: Ensemble size influence on overall error

4.4 Experiment 4: Trade-off between filter window size and ensemble size

In this experiment we combine filter window size and ensemble size to quantify the trade-off between the two. Our variable of interest is again the error under the curve as in Figure 4 panel E. Here we only repeat the experiment for each window-size-ensemble-size combination five times and we only test the behavior over two years – all for computational efficiency. The main finding, depicted in Figure 7, is that the size of the ensemble primarily influences the reliability of an estimate, with only a minor impact on its accuracy but the filter frequency, that is s smaller filter window, improves accuracy substantially. An increase in ensemble size contributes to accuracy enhancement primarily when the filter window is quite small (so filter frequency is high). In terms of standard statistical terminology, ’accuracy’ pertains to the proximity of the mean estimate to the true value, while ’reliability’ relates to the consistency of estimated values around the mean.

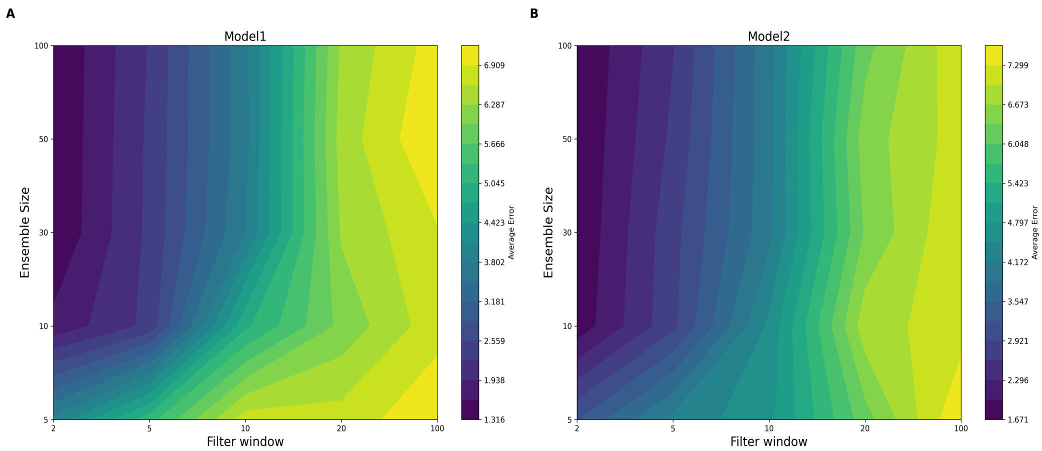

Figure 7: Trade-off filter frequency and ensemble size

4.5 Experiment 5: Variation of assumptions about uncertainty

Experiment number five is critical in the overall evaluation of our setup because here we study the tradeoff between uncertainty in the model and uncertainty in the observations. For this purpose, we define two similar sources of uncertainty – one in each model. For model one we alter the "power" that an agent derives from having wealth by introducing a normally distributed term to the exponent that defines that "power" $p^{\beta+\mathcal{N}(0,a*\beta)}$. The term $a*\beta$ says that the standard deviation of the distribution is dependent on its mean β. For model two, we simply vary the normally distributed error defined in Equation 5. Both normal distributions' standard deviation can be meaningfully varied within the interval $[0,1]$. They do not represent the same underlying variable, but this helps comparability of the EnKF performance across both models. We also define a new performance metric for this experiment because we are not interested in the error relative to the ground truth, but in how sensitive the EnKF is to changes in the uncertainty in model ensemble and observations ensemble and whether it tends towards one of them based on the uncertainty in the other. Therefore we evaluate the posterior difference, T, between the mean model estimate $M_{j,i}$ for wealth group j at time point i and the data ensemble average $D_{j,i}$. This is the average of **D** defined in equation 9. We then sum up across wealth groups and over time as depicted in Equation 18:

$$T = \sum_t \sum_i |M_{j,i} - D_{j,i}| \tag{18}$$

The size of this posterior gauge should decrease as the uncertainty in the observation decreases, so towards the left of the horizontal axes in Figure 8, and should increase when there is greater uncertainty in the model, so upwards the vertical axes in Figure 8. This is because we measure the distance to the observations in equation **??**. This is exactly what we find; Figure 8 illustrates the results for both models. Although it should be noted that, the uncertainty, as we have defined it, in the model matters less than the uncertainty present in the data.

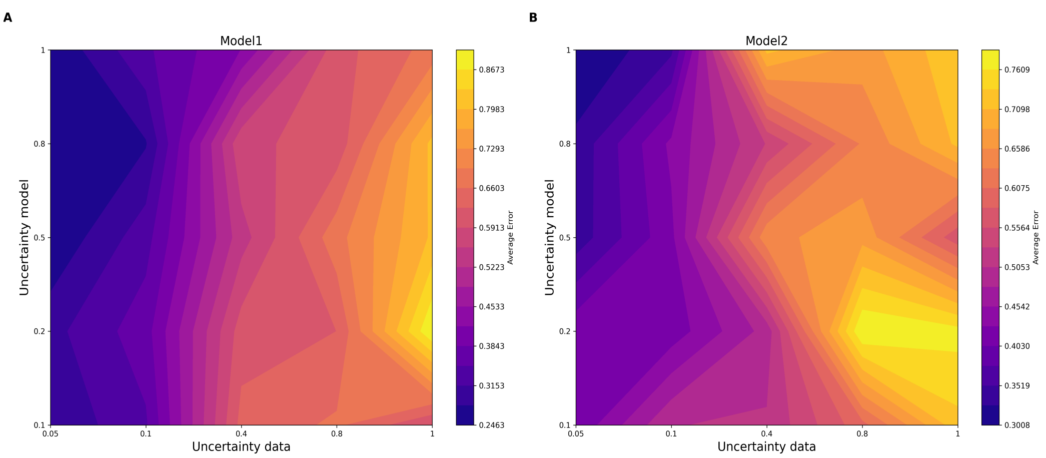

Figure 8: Trade-off uncertainty model and uncertainty observation

5 Discussion and conclusions

In summary, we have shown that the Ensemble Kalman Filter (EnKF) can be successfully combined with macro-economic agent-based models, specifically for the dynamics of wealth inequality. The EnKF consistently improves the fit to the ground truth state of the system, as characterised by the data provided by Blanchet, Saez, and Zucman (2022) (Section 3.1). An important contribution we make is that that our hybrid implementation is specifically designed for the *dynamics* of the wealth distribution and consequently macro-economic dynamics in general. Thereby we go beyond other recent ABM-Bayesian-Update hybrid models like the one by Yang and P. Zhou (2022) who focus on reproducing the shape of the wealth distribution in a static setting.

The combination of data assimilation (DA) and agent-based modelling (ABM) has the potential to significantly enhance some economic modelling approaches. Whilst the benefit of making up-to-the-minute predictions is less important for economics than for fields such as meteorology (Kalnay 2003) or crowd simulation (Clay et al. 2021), the benefits of creating a system to make "live" (Swarup and Mortveit 2020) predictions is still potentially transformative. For example, economists could create a simulation platform that made predictions about future conditions days, months, or years in advance. Using methods like the EnKF – in this case through manipulating the observation operator (H), see Section 3.5, although we note later that this may be non-trivial – the system could take up-to-date, diverse data sets from sources such as official government releases, social media data, new empirical research results, etc., and use the DA-ABM framework as a means to combine them and create accurate predictions using "all the available information" (Talagrand 1991). Predictions made from a model that begins with the most accurate *current* system state will be much more reliable than those that are based solely on historical data. Such a system could be invaluable to policy makers as they navigate complex global events and shifting social attitudes. The benefits of the combined DA-ABM system are that we can not only make the best use of a diverse array of up-to-date data, but can also use the model to better understand the underlying dynamics that are driving current trends. Neither the model nor the data, in isolation, will be able to provide such insight.

An interesting and unexpected observation is that, occasionally, the EnKF makes the prediction worse (e.g. Figure 4 panel E). This can happen due to sampling error in the data ensemble, for instance, when the sampled observations diverges from the ground truth. The Ensemble Transform Kalman Filter (ETKF) is a variant of the EnKF that intends to overcome this problem (T. Zhou et al. 2019); future work could explore this approach as an alternative to the EnKF for economic models. In addition, the specific case of U.S. wealth distribution explored here exhibits unusual features that can be problematic. For example, in 2007 there are negative wealth values for the bottom 50% and, simultaneously, large positive values for the very wealthy groups. This can "confuse" the EnKF and hence the likelihood for bad performance is heightened (meaning it is difficult for to find optimal update weights for either group).

We tested two dynamic models simultaneously for a more general performance evaluation of our proposed method. While both models in an EnKF-ABM setting act relatively similarly, Model Two is somewhat more unstable. This is because the steady-state of the model is further away from the ground-truth data than the average model run in Model One. Moreover, this steady-state is strongly attracting. If the model is corrected towards the ground truth by the EnKF, there is a strong tendency for the model to evolve rapidly toward its characteristic steady-state again, which in turn results into more abrupt changes in the wealth trajectories of the wealth groups. This is an important lesson for further studies combining economic models and data assimilation. Most classical economic models exhibit "transitional dynamics" towards a steady state (Acemoglu 2008). While many agent-based models are constructed for the purpose of increased complexity and instability when compared to classical equilibrium models, they still might exhibit long-run steady states for tractability as well or even emergent, unforeseen, attractors. Our results suggest that if those steady states are rapidly converging and perhaps not exactly matching some more "chaotic" time series, it might be more difficult for data assimilation to learn the true state of a system. The fact that the ground truth state that we track is four dimensional (because there are four wealth groups), with possibly diverging trajectories, also amplifies this problem. For this reason, it is important for the EnKF to intervene particularly early on with respect to Model Two. If the data assimilation window is too wide, and the model has already diverged far from the ground truth, it is more difficult for the model-EnKF hybrid to generate a low-error performance over time.

There are many practical implementation challenges for agent-based models in combination with advanced filtering techniques like the EnKF. We specifically focused on "learning" the state of the wealth dynamics only, so the wealth of agents or the wealth groups, and not any underlying parameters which simplifies the exercise. Even in this simplified procedure, it is a non-trivial question of how the observations relate to the agent states. In our case, we defined the H operator to be an aggregation operation, since agent states and observations both represented the same quantity, that is "wealth" measured in dollars. While this seems like a natural choice for micro-founded macro-economic models this might not always be possible. For instance, one might think about an agent-based model explaining inflation through labour market dynamics and central bank policy. The inflation rate is perhaps the observation but the agents in the model, workers and banks perhaps, do not exhibit "inflation" as an individual characteristic and this cannot be simply be aggregated. Therefore, defining a more general framework of what observation operators for economic agent-based models look like is interesting for further studies. In comparison to work that has been done with the EnKF in urban system studies, and particularly pedestrian trajectory estimation, the H operator for instance simply excludes some variables that appear in the model but not in the observations.

Another challenge with the wealth distribution particularly is that the right-tail of the distribution exhibits extreme values. This makes a smooth update experience difficult. The Kalman Gain weights for distinct wealth groups potentially develop in entirely different directions with very extreme absolute difference. Hence one natural variation of our work might be with log-transformed wealth values instead of untreated values. However for this the mathematics in Section 3.5 do not translate trivially.

Again, our observation operator H is based on summing agent states to wealth groups and averaging them. Log-transformed values cannot be summed and averaged in the same way. A different operation other than basic matrix multiplication would be required for this to work.

In conclusion, this paper suggests that the synergy of data assimilation and agent-based modelling holds significant promise for advancing economic prediction and estimation. Our findings underscore the potential of such models to harness contemporary data for more accurate and timely economic predictions, once challenges around model stability and reconciliation of data-model mappings have been resolved. These findings pave the way for further research, balancing model accuracy with practical application in economic modelling.

Declaration of competing interests

The authors declare no competing interests.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 757455).

Code and data accessibility

All code and data relating to this study is open-access at `https://github.com/yloswald/real_time_ineq_abm`.

References

Acemoglu, Daron (2008). *Introduction to modern economic growth*. Princeton university press, p. 61.

Aktieninstitut, Deutsches (2023). *Shareholder numbers 2022: Germany can share! (in German)?* URL: `https://www.dai.de/en/detail/aktionaerszahlen-2022-trotz-rekordniveau-muss-die-politik-jetzt-liefern/`.

Albert, Réka and Albert-László Barabási (2002). "Statistical mechanics of complex networks". In: *Reviews of modern physics* 74.1, p. 47.

Angle, John (1986). "The surplus theory of social stratification and the size distribution of personal wealth". In: *Social Forces* 65.2, pp. 293–326.

Athans, Michael (1974). "The importance of Kalman filtering methods for economic systems". In: *Annals of Economic and Social Measurement, Volume 3, number 1*. NBER, pp. 49–64.

Auten, Gerald and David Splinter (2019). "Income inequality in the United States: Using tax data to measure long-term trends". In: *Working paper*.

Axtell, Robert L and J Doyne Farmer (2022). "Agent-based modeling in economics and finance: Past, present, and future". In: *Journal of Economic Literature*.

Blanchet, Thomas and Clara Martinez-Toledano (2023). "Wealth inequality dynamics in europe and the united states: Understanding the determinants". In: *Journal of Monetary Economics* 133, pp. 25–43.

Blanchet, Thomas, Emmanuel Saez, and Gabriel Zucman (2022). *Real-time inequality*. Tech. rep. National Bureau of Economic Research.

Chancel, Lucas (2022). "Global carbon inequality over 1990–2019". In: *Nature Sustainability* 5.11, pp. 931–938.

Chancel, Lucas et al. (2022). *World inequality report 2022*. Harvard University Press.

Chiang, Yen-Sheng (2015). "Good samaritans in networks: An experiment on how networks influence egalitarian sharing and the evolution of inequality". In: *PloS one* 10.6, e0128777.

Clay, Robert et al. (2021). "Real-time agent-based crowd simulation with the Reversible Jump Unscented Kalman Filter". In: *Simulation Modelling Practice and Theory* 113, p. 102386. ISSN: 1569-190X. DOI: `https://doi.org/10.1016/j.simpat.2021.102386`. URL: `https://www.sciencedirect.com/science/article/pii/S1569190X21000939`.

Di Matteo, Tiziana, Tomaso Aste, and ST Hyde (2003). "Exchanges in complex networks: income and wealth distributions". In: *arXiv preprint cond-mat/0310544*.

Dragulescu, Adrian and Victor M Yakovenko (2000). "Statistical mechanics of money". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 17, pp. 723–729.

Drăgulescu, Adrian and Victor M Yakovenko (2001). "Evidence for the exponential distribution of income in the USA". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 20, pp. 585–589.

Evensen, Geir (2003). "The ensemble Kalman filter: Theoretical formulation and practical implementation". In: *Ocean dynamics* 53, pp. 343–367.

Gilbert, G. Nigel (2008). *Agent-Based Models*. Quantitative Applications in the Social Sciences no. 07-153. Los Angeles: Sage Publications. ISBN: 978-1-4129-4964-4.

Hu, Mao-Bin et al. (2008). "Properties of wealth distribution in multi-agent systems of a complex network". In: *Physica A: Statistical Mechanics and its Applications* 387.23, pp. 5862–5867.

Hu, Xiaolin (2022). "Data Assimilation For Simulation-Based Real-Time Prediction/Analysis". In: pp. 404–415.

Ichinomiya, Takashi (2012). "Wealth distribution on complex networks". In: *Physical review E* 86.6, p. 066115.

Inglesi-Lotz, Roula (2011). "The evolution of price elasticity of electricity demand in South Africa: A Kalman filter application". In: *Energy Policy* 39.6, pp. 3690–3696.

Jones, Jeffrey M. (2023). *What Percentage of Americans Own Stock?* URL: `https://news.gallup.com/poll/266807/percentage-americans-owns-stock.aspx`.

Kalman, Rudolph Emil (1960). "A new approach to linear filtering and prediction problems". In.

Kalnay, Eugenia (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Lee, Gyemin and Gwang Il Kim (2007). “Degree and wealth distribution in a network induced by wealth”. In: *Physica A: Statistical Mechanics and its Applications* 383.2, pp. 677–686.

Lux, Thomas (2005). “Emergent statistical wealth distributions in simple monetary exchange models: a critical review”. In: *Econophysics of Wealth Distributions: Econophys-Kolkata I*, pp. 51–60.

Malleson, N. et al. (2020). “Simulating crowds in real time with agent-based modelling and a Particle Filter”. In: *Journal of Artificial Societies and Social Simulation* 23.3, p. 3. DOI: `https://doi.org/10.18564/jasss.4266`.

Mattison, Siobhán M et al. (2016). “The evolution of inequality”. In: *Evolutionary Anthropology: Issues, News, and Reviews* 25.4, pp. 184–199.

Munguia, Rodrigo, Jessica Davalos, and Sarquis Urzua (2019). “Estimation of the Solow-Cobb-Douglas economic growth model with a Kalman filter: An observability-based approach”. In: *Heliyon* 5.6.

Papadopoulos, Georgios (2019). “Income inequality, consumption, credit and credit risk in a data-driven agent-based model”. In: *Journal of Economic Dynamics and Control* 104, pp. 39–73.

Pasricha, Gurnain Kaur (2006). “Kalman filter and its economic applications”. In: *Munich Personal RePEc Archive*.

Poledna, Sebastian et al. (2023). “Economic forecasting with an agent-based model”. In: *European Economic Review* 151, p. 104306.

Schneider, Wolfgang (1988). “Analytical uses of Kalman filtering in econometrics—A survey”. In: *Statistical Papers* 29.1, pp. 3–33.

Semieniuk, Gregor (2024). “Inconsistent definitions of GDP: Implications for estimates of decoupling”. In: *Ecological Economics* 215, p. 108000.

Silva, Petrônio CL et al. (2020). “COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions”. In: *Chaos, Solitons & Fractals* 139, p. 110088.

Souma, Wataru, Yoshi Fujiwara, and Hideaki Aoyama (2003). “Wealth Distribution in Scale-Free Networks”. In: *Meeting the Challenge of Social Problems via Agent-Based Simulation: Post-Proceedings of the Second International Workshop on Agent-Based Approaches in Economic and Social Complex Systems*. Springer, pp. 37–49.

Swarup, Samarth and Henning S. Mortveit (2020). “Live Simulations”. In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS ’20. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 1721–1725. ISBN: 978-1-4503-7518-4.

Talagrand, O. (1991). “The use of adjoint equations in numerical modelling of the atmospheric circulation”. In: *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*. Ed. by A. Griewank and G. F. Corliss. Philadelphia, PA: SIAM, pp. 169–180. ISBN: 0-89871-284-X.

Tang, Daniel and Nick Malleson (2022). “Data assimilation with agent-based models using Markov chain sampling”. In: *arXiv preprint arXiv:2205.01616*.

Ternes, Patricia et al. (2022). “Data assimilation and agent-based modelling: towards the incorporation of categorical agent parameters”. In: *Open Research Europe* 1.

Thamae, Retselisitsoe Isaiah, Leboli Zachia Thamae, and Thimothy Molefi Thamae (2015). “Dynamics of electricity demand in Lesotho: A Kalman Filter approach”. In: *Studies in Business and Economics* 10.1, pp. 130–139.

Vallejos, Hunter A, James J Nutaro, and Kalyan S Perumalla (2018). “An agent-based model of the observed distribution of wealth in the United States”. In: *Journal of Economic Interaction and Coordination* 13, pp. 641–656.

Vishwakarma, Keshav P (1970). “Prediction of economic time-series by means of the Kalman filter”. In: *International Journal of Systems Science* 1.1, pp. 25–32.

Wang, M. and X. Hu (2015). “Data assimilation in agent based simulation of smart environments using particle filters”. In: *Simulation Modelling Practice and Theory* 56, pp. 36–54. DOI: https://doi.org/10.1016/j.simpat.2015.05.001.

Ward, Jonathan A, Andrew J Evans, and Nicolas S Malleson (2016). “Dynamic calibration of agent-based models using data assimilation”. In: *Royal Society open science* 3.4, p. 150703.

World Inequality Database (2024). *World Inequality Database*. Accessed: 2024-01-25.

Yakovenko, Victor M and J Barkley Rosser Jr (2009). “Colloquium: Statistical mechanics of money, wealth, and income”. In: *Reviews of modern physics* 81.4, p. 1703.

Yang, Xiaoliang and Peng Zhou (2022). “Wealth inequality and social mobility: A simulation-based modelling approach”. In: *Journal of Economic Behavior & Organization* 196, pp. 307–329.

Zhou, Tengjiao et al. (2019). “Ensemble transform Kalman filter (ETKF) for large-scale wildland fire spread simulation using FARSITE tool and state estimation method”. In: *Fire safety journal* 105, pp. 95–106.