Bayesian estimation of a large-scale macroeconomic policy agent-based model

Sylvain Barde[a]

[a]*School of Economics, University of Kent*
Sibson Building, Park Wood Road, Canterbury CT2 7FS, UK, email: `s.barde@kent.ac.uk`

Abstract

Empirical parameter estimation of large scale agent-based models has long been recognised as computationally challenging. Their bottom-up nature imposes the use of non-parametric or indirect inference methods, which in turn typically requires a significant amount simulated data. However, their high computational requirements makes the application of these estimation methodologies unfeasible in practice. This hurdle can limit the applicability of agent based models for quantitative policy advice, for example in scenario analysis, in cases where the parameter calibration cannot be checked against empirical data. We show how this problem can be overcome by estimating the Dosi et al. (2015) agent-based model. This extends the original 'Keynes meets Schumpeter' model of Dosi et al. (2010) allowing for the interaction of fiscal and monetary policy. 18 free parameters of the model are estimated on 10 standard macroeconomic US variables, using annual and quarterly data, and the estimates obtained are shown to improve the fit of the baseline model on the data. A model-selection exercise is carried out, investigating impact of changing the expectation-formation mechanism. Finally, the original policy experiments are replicated using the new empirical estimates, showing that pushing the model into a low-growth regime leads to several key differences relative to the original conclusions. Overall, the exercise establishes the feasibility and relevance of macro-economic empirical parameter estimation and mechanism selection to ABM designers for improving the fit of scenario analyses.

JEL classification: C15, C52, C63.

Keywords: Agent-based models, Model comparison, Calibration methods

---

1. Introduction

The bottom-up nature of agent-based models (ABMs) offers both a unique opportunity and challenge for macroeconomics. On the one hand, as argued for example by Haldane and Turrell (2018a,b) their flexibility enables the integration of non-standard mechanisms that are difficult to work with analytically, such as bounded rationality or emergence from interaction, into economic models. Similarly, their reliance on simulation as their main quantitative mechanism means that they are well suited to scenario analysis, and therefore have the potential to contribute to understanding the impact of economic policies.

The flip side of this bottom-up, simulation-based, flexibility is a corresponding difficulty in validating these models empirically, for instance estimating deep parameters from data and selecting amongst different alternative behaviour rules or model mechanisms. This difficulty, referred to as the validation problem in the ABM literature (Fagiolo et al., 2019; Delli Gatti and Grazzini, 2020) is particularly acute for policy-oriented ABMs, for two distinct reasons. First of all, the ability for a model to inform policy arguably requires that it be a reasonable representation of the phenomenon of interest. If this is not the case, then there will necessarily be doubts on whether the findings in the ABM generalise to reality. Second, in order to achieve this reasonable representation of reality, policy ABM models are often large in size and posses a high dimensional parameter space. This in turn requires a larger number of simulation steps *ceteris paribus*, and correspondingly larger computational requirements, which often lead to slower simulation speeds compared to more basic ABMs, complicating the validation problem.

This paper serves a dual purpose. First, the aim is to demonstrate a feasible workflow for empirical validation of a relatively large macroeconomic ABM from macroeconomic data. Similar to Barde (2024), the paper will rely on a surrogate Bayesian methodology to estimate the parameters of the model, however it will additionally discuss the choice of empirical data and parameters to estimate, as well as investigate the role of model uncertainty by carrying out a model selection exercise, thus demonstrating how to identify the most effective expectation mechanism out of the several candidates available. Second, in doing so, the aim is to also contribute to a better understanding of the features of a key policy-oriented ABM, identify how key mechanisms affect empirical fit, in order to make recommendations on model design, and thus improve its ability to provide relevant recommendations. Central to this will be the replication of the policy simulations of the original model using the parameter estimates obtained, in order to assess how the policy predictions of the fitted model might differ from those of the original calibration.

The model selected as the basis of this estimation exercise is the financial version of the 'Keynes meets Schumpeter' (K+S) model (Dosi et al., 2015), which extends the original model of Dosi et al. (2006, 2010) by including a financial sector where banks provide loans to firms. This banking sector is overseen by a central bank, providing scope for a monetary policy which can interact with the various fiscal policies enacted using the government budget. Two main considerations motivate this choice of model as the basis of the analysis. First, this model is part of a wider and mature family of K+S models, which has been used extensively in a range of policy applications. As an example, Dosi et al. (2013) examine the impact of consumer credit policies on income distribution, Dosi et al. (2017b) looks at impact of flexible labour market policies, while Dosi et al. (2017a) investigate how innovation and industrial policies affect the Schumpeterian side of the model. More recently, the Dystopian Schumpeter-Keynes (DSK) model has augmented the basic K+S framework with an integrated assessment climate model, in order to analysis the impact of environmental policies when interacted with an economic model (Lamperti et al., 2018a, 2019, 2020, 2021). A second motivation for this specific choice of K+S version is its relevance for macroeconomic policy decision, as the financial version of the model was specifically designed to examine the interplay of fiscal and monetary policies. Providing clarification as to the sensitivity of these policy recommendations when confronted with empirical data is important.

In terms of methodologies available, a range of approaches have been developed to address the issue of ABM validation, in particular for calibrating or estimating ABMs from macro and microeconomic data.[1] The specific choice made here is to carry out esti-

[1] See for instance Gilli and Winker (2003); Grazzini and Richiardi (2015); Grazzini et al. (2017); Kukacka and Barunik (2017); Lamperti et al. (2018b); Delli Gatti and Grazzini (2020) for key methodological contributions that are relevant to the approach chosen here.

mation using macroeconomic data only, which clearly ignores the possibility of calibrating the behaviour and initial state of the ABM's agents using richer microeconomic dataset. Given that ABMs typically model heterogeneous agents, the ability to calibrate the rules governing agent behaviour and the distributions of firms and households directly from microeconomic data offers great potential. A recent example of this is Poledna et al. (2023), who use census data, business surveys and disaggregated input-output tables to finely calibrate an ABM that can outperform VARs and DSGE models when forecasting the Austrian economy. This approach was recently extended in Wiese et al. (2024), drawing the model's heterogenous agent populations from demographic and socio-economic data, resulting in an improvement of the model's forecasting ability.

Unlike these data-driven approaches, the validation exercise carried out here will focus exclusively on macroeconomic data, for several reasons. First of all, direct calibration from microeconomic data imposes that the structure of the ABM be representative of the real economy. The Poledna et al. (2023) model, for instance, contains 62 distinct industries, which are then linked using the Austrian input-output tables. By contrast, the K+S model is much more stylised, using only two vertically linked sectors. In this case, even assuming the availability of the required microeconomic data, it is not clear how one would map the real-life economic structure to the model. The second motivation for this choice is the fact that for policy models, even in the case where the individual agent behaviours and the initial condition of the model can be set using micro-level data, there might be a need to fine-tune the model to ensure that the overall macroeconomic aggregates generated by the model match the macro-economic observables. In particular, this is likely to be the case for the rules of behaviour for the macroeconomic policy-makers such as the government and the central bank, and the parameters that govern their interaction with the rest of the economy.

The remainder of the paper is organised as follows: the version of the K+S model used in the analysis is presented in section 2, with the methodologies, empirical dataset and choice of estimated parameters presented in section 3. The parameter estimates and goodness-of-fit comparisons for the various specifications estimated are discussed in section 4, while the policy experiments of the original paper are re-examined in light of estimates in section 5. Section 6 then concludes.

2. The financial K+S model

2.1. Key features of the K+S model

The origin of the K+S family of models is rooted in Dosi et al. (2006) and Dosi et al. (2010). All the subsequent additions mentioned above, which extend the base model to a range of policy-relevant issues, share the same set of core features. The first is the Schumpeterian innovation process, consisting of a two-sector productive sector, where upstream firms produce capital goods, which are then purchased by downstream firms to produce a final consumption good. Each upstream firm invests in R&D which stochastically results in improvements in the productivity of their capital good, either through radical innovation, or imitation of competitor technology. The purchasing decisions of the downstream firms then depend on the relative productivities of the range of upstream capital goods, as well as the average productivity of their existing stock of capital vintages. This results in a productive structure where productivity is directly embodied in capital goods, and random innovations disseminate organically through the economy as the less innovative capital goods producers are either competed away by the more innovative ones, or are able to catch up through imitation.

The second feature is the Keynesian structure of demand in the economy, which manifests in several ways. First is the structure of the labour market, where labour is inelastically supplied by households and the institutional wage setting, while responsive to unemployment, does not necessarily clear the market, creating involuntary unemployment. This is complemented by unemployment benefits funded from general taxation and the issuance of bonds. The second important Keynesian flavour resides in the bounded rationality of agents, particularly consumption good firms, that form expectations on future demand using myopic heuristics, thus deviating from rational expectations and modelling an 'animal spirits' behaviour. These heuristics form part of the selection exercise and are discussed further below.

The overall result from the interaction of these two components is a model that is able to replicate several important stylised facts in the economy, particularly the high volatility of investment relative to output, combined with a lower volatility of consumption relative to output, the skewed distribution of firm sizes, as well as the granularity of firm-level investment and associated diffusion of innovation.

The subsequent extensions presented in the introduction typically investigate how this central setting is affected by the combination of various policies and/or additional markets. In the case of the financial model of Dosi et al. (2015), this comes in the form of a banking sector, which provides loans to consumption-good firms in order to finance their purchases of capital goods. Banks have to finance themselves and are allowed to fail, creating a role for the central bank to regulate the interest rate and bailout failing banks. Combined with the potential for automatic stabilisers in the form of government deficits during downturns, this provides a setting where the interaction of monetary and fiscal policies with the Keynesian/Schumpeterian structures outlined above can be examined. The specific details around the fiscal and monetary policies included in the analysis are provided in section 5.

2.2. Expectation mechanisms

Part of the empirical exercise carried out involves verifying the sensitivity of the parameter estimates and policy experiments to the choice of expectation mechanism used by consumer goods firms when forecasting future demand, which governs their production plans and therefore their demand for inputs, primarily labour. This analysis is done for three reasons. First of all, by performing model selection over these various expectation mechanisms, this extends the analysis of Barde (2024) and provides a more complete validation workflow, dealing with both parameter and model uncertainty.

Beyond this demonstration of model selection, a second rationale for evaluating the behaviour of the empirically fitted model under a range of expectation mechanisms specifically is the fact that the original designers of K+S report that under their calibration, the model predictions do not substantially change under a range of expectation mechanisms. Dosi et al. (2006), for instance, present a range of expectation mechanisms in addition to their baseline specification, and report that the model's ability to replicate stylised facts is not affected. Similarly, Dosi et al. (2015, footnote 5) state that "the simulation results do not significantly change when consumption-good firms follow more sophisticated expectation formation rules". Dosi et al. (2020) provide a deeper analysis, allowing agents to switch rules depending on their relative forecasts losses, and again report that all rules coexist in the statistical equilibrium, suggesting none dominates in practice. This robustness appears to be a key feature of the model, to the point that many of the extensions discussed in the introduction simply adopt the baseline expectation mechanism as a simplifying assumption. Given this, it is important to verify if this property remains when the model is fitted to empirical data.

In the baseline version of the model, which is the one used in Dosi et al. (2015), firms have purely myopic expectations, with a firm i setting its expected level of demand equal to the observed demand in the previous period:

$$E_{t-1}[D_{i,t}] = D_{i,t-1} \tag{1}$$

The original version of the model, Dosi et al. (2006) provides 4 alternatives to this mechanism. The first models expected demand using an AR(4) process on observed demand, where $0 < \beta_{1,2,3,4} < 1$:

$$E_{t-1}[D_{i,t}] = \beta_1 D_{i,t-1} + \beta_2 D_{i,t-2} + \beta_3 D_{i,t-3} + \beta_4 D_{i,t-4} \tag{2}$$

Under accelerative expectations firms adjust the myopic expectation mechanism (1) by the growth rate of observed demand, with $0 < \beta_5 < 1$:

$$E_{t-1}[D_{i,t}] = (1 + \beta_5 \Delta_{\%} D_{i,t-1})\, D_{i,t-1} \tag{3}$$

The adaptive expectations is similar in spirit, firms adjust their expectations proportionally to the difference between the previous period's expected and observed demands, by a factor $0 < \beta_6$:

$$E_{t-1}[D_{i,t}] = E_{t-2}[D_{i,t-1}] + \beta_6 \left(D_{i,t-1} - E_{t-2}[D_{i,t-1}]\right) \tag{4}$$

The final expectation mechanism, referred to as micro-macro expectations in Dosi et al. (2006) and 'anchor and adjust' (A-A) in Dosi et al. (2020), is similar to the accelerative expectations (3), but the growth correction to the myopic expectation (1) is a weighted average of the firm-level growth in demand and the economy-wide growth in GDP, controlled by $0 < \beta_7 < 1$.[2]

$$E_{t-1}[D_{i,t}] = \left(1 + \beta_7 \Delta_{\%} D_{i,t-1} + (1-\beta_7)\Delta_{\%} Y_{t-1}\right) D_{i,t-1} \tag{5}$$

3. Empirical strategy

This section presents the key decisions involved in validating an ABM, namely what methodologies to use to carry out the validation, what empirical data to use, and which parameters to estimate. These are covered in the following subsections, however it is important to emphasise that these decisions are not necessarily sequential, and are often made jointly.

3.1. Estimation and selection methodologies

Three main methodologies are used in the estimation workflow, in order to (i) estimate the parameters, (ii) run diagnostics on the estimation method and simulated data generated by the K+S model and (iii) help select a preferred specification from the various expectation mechanisms used.[3]

As already stated, the methodology used to estimate the parameters is the Bayesian estimation with gaussian regression surrogates (BEGRS) approach of Barde (2024), previously used on the stock-flow consistent model of Caiani et al. (2016). This approach

[2]Note that in Dosi et al. (2006), the micro-macro expectations use two free parameters $0 < \beta_7, \beta_8$. This implies that in Dosi et al. (2020) the anchor and adjust mechanism is a restricted version where $\beta_7 + \beta_8 = 1$.

[3]Replication files for the work carried out here can be found in the following repository: `https://github.com/Sylvain-Barde/begrs_ks`.

combines the insights of Lamperti (2018) and Grazzini et al. (2017) and provides a full Bayesian estimation workflow for computationally demanding models, by relying on a surrogate of the simulated model to generate the likelihood function. This greatly reduces the amount of simulations of the underlying ABM that need to be run as part of the estimation process, as time-consuming simulations are only required in the training phase and not in the Monte-Carlo Bayesian estimation phase, when drawing samples from the posterior distribution.

Each version of the K+S expectation mechanisms outlined in section 2.2 was estimated using its own Gaussian process (GP) surrogate. In each case the training data consisted of 4000 distinct simulations, each generated using a unique parametrisation drawn from a multivariate Sobol sequence, with the parameter space intervals provided in table 3. In most cases these intervals are centered on the baseline parameter value, ensuring that when combined with a flat prior, the prior mean corresponds to that baseline. Each simulation generates 300 time-series observations[4] for the 10 empirical variables. This results in a pool of 1,200,000 observations for training the corresponding GP surrogate. Each GP was trained using variational inference, using 7 latent observable variables and 250 inducing points.

Once the surrogate is trained, the estimation itself is carried out using 10,000 iterations of the No U-turn Sampler (NUTS) of Hoffman et al. (2014), using the minimal prior of Barde (2024) and the parameter boundaries in table 3. This minimal prior is a convex relaxation of an uninformative, uniform prior which ensures that the gradient of the posterior is defined at the boundary of the parameter space itself, thus enabling efficient Hamiltonian Monte Carlo methods.

Because BEGRS is a Bayesian methodology, it is possible to use simulation-based calibration (SBC) (Talts et al., 2018) to evaluate the convergence of the surrogate posterior to the true, but unobserved K+S model posterior. The key insight of this methodology is that when using any Bayesian estimation method, the expectation of the posterior distribution of parameters taken with respect to data drawn from the joint distribution is

[4]In practice each simulation consists of 601 periods, allowing for 300 burn-in observations and one lag.

equal to the prior distribution. Such a validation dataset can be obtained by drawing a set of candidate parameters from the prior distribution and running them through the data-generating process to obtain simulated datasets. Any deviation between the resulting data averaged posterior and the prior indicates a problem with the methodology. In practice this is carried out using a histogram of rank statistics for the parameter estimates, as this provides a convenient statistical test: under the null hypothesis that the data-averaged posterior matches the prior, the histogram should follow a uniform distribution.

In practice, the SBC test is a more rigorous generalisation of a common parameter recovery exercise, where data from a known data-generating process is estimated in an attempt to check that the true parameter values are recovered. The key improvement is that the procedure is run on a large and randomised number of such known parameter recovery exercises, allowing a formal statistical verification of the recovery property of the posterior. The SBC analysis presented in section 4.1 is run on a testing set generated from an additional 1000 parameterisations, each providing 400 simulated observations, drawn from further on in the same Sobol sequence as the one used to generate the BEGRS training data. Because the BEGRS estimation uses a uniform prior, the use of a Sobol is appropriate, as it ensures a uniform distribution of prior draws.

A key benefit of the SBC approach is that in the context of the BEGRS surrogate model, it provides a joint diagnostic of the surrogate posterior and the statistical properties of the underlying simulation model. Specifically, in a case where the training data used to train the surrogate does not satisfy the assumptions required for the GP to converge (in particular bounded support, which can be satisfied by stationary data), then this will be picked up in the SBC diagnostic. The downside of the SBC methodology is that should a Bayesian method fail the SBC diagnostic, the method does not by itself indicate what the root cause of the problem is.

Finally, the method used to verify the goodness-of-fit of the various K+S specifications estimated is the Markov information criterion (MIC) of Barde (2020), which provides an unbiased estimate of the cross entropy between a model and the empirical data, in a generalisation of the more traditional Akaike (1974) information criterion (AIC). Because BEGRS performs Bayesian estimation, in principle one would want to compare the marginal densities obtained during the estimation to select amongst the various candidate specifications. In practice, however, each specification is estimated using a separate surrogate, which only provides an approximation to the true model. Because each surrogate will have converged differently to the true model, a direct comparison of the posterior densities does not allow for model selection, due to the presence of model-specific approximation errors. As a result, it is preferable to simulate the K+S model using the estimated parameter values for each specification, and use the MIC to obtain an information criterion on the empirical US data. This method relies on a minimum description length approach to measuring the cross entropy between the simulated and empirical data (see Grünewald, 2007), providing a generalisation of the AIC to simulation models. As is the case for the AIC, differences of the MIC across candidate specifications measure differences in the Kullback-Leibler divergence between the different models and the empirical data. In practice, the MIC achieves this by using context tree weighting to compress a binary discretisation of the empirical data, as this provides an analytical bias correction procedure, providing unbiased estimates of the cross entropy. The context tree, which encodes the conditional probability structure of the models, is itself trained on the same binary discretisation of the simulated data.

In practice, the MIC goodness of-fit analysis in section 4.3 relies on the simulated dataset generated for the policy experiments in section 5, using the default settings of no fiscal policy and a dual-mandate Taylor rule, as this reduces amount of simulation runs required for the overall validation exercise. The analysis uses 1000 replications of 300 observations, providing 300,000 training observations for the MIC algorithm, which run with the variables discretised to 6 bits, using one lag of the variables and 32 bits of context memory.[5]

	Variable	FRED series
Δy	log diff. of p.c. real GDP	GDPC1
Δc	log diff. of p.c. real Consumption	PCEC
Δi	log diff. of p.c. real Investment	FPI
ΔL	log diff. of p.c. real Commercial & industrial loans	ACILACB, CILACBQ158SBOG†
Δnw	log diff. of p.c. real non-financial corporate net worth	TNWMVBSNNCB
Δw	log diff. of hourly non-farm real compensation	COMPRNFB
u	Unemployment rate	CE16OV, CLF16OV
r	Central bank policy rate	FEDFUNDS, FFWSJLOW†, FFWSJHIGH†
π	Inflation (log diff. of GDP deflator)	GDPDEF
lr	loss ratio, Net charge offs to total loans & leases	QBPLNTLNNTCGOFFR, FDIC CB-8 & CB-9†

Note: All empirical series were obtained from `https://fred.stlouisfed.org/`, except from the FDIC data, which is taken from FDIC (1997), accessed via `https://fraser.stlouisfed.org/title/statistics-banking-125`. 'p.c.' stands for per capita, the population series used for per-capita calculations is CNP16OV and the deflator used to obtain real values is GDPDEF. † indicates data used to extend the annual dataset before 1984 for the loans 'ΔL' and loss ratio 'lr', and before 1954 for the policy rate 'r'.

Table 1: Empirical observable variables

3.2. Empirical datasets

A key consideration for the empirical estimation is whether to use annual or quarterly data. This decision involves a number of trade-offs linked to several important issues. First is the design of the K+S model itself where the decision rules of agents, most notably expectation formation or investment decisions, are based on an annual frequency, thus making that setting more natural for estimation. Estimation using quarterly data, however, has the key advantage of providing more observations over a shorter time-frame relative to annual data. Both aspects are important: on top of the reduction in degrees of freedom relative to quarterly series, an annual data series long enough to provide enough observations is likely to contain structural breaks, as the structure of economy changes.[6] Estimating deep model parameters using annual data that spans several decades might therefore be problematic, and in itself create identification problems as a result of such breaks, a problem that is less severe with quarterly data.

[5]For robustness the MIC was also run with 2 lags of the variables, providing similar results which are provided in the supplementary material.

[6]See in particular Nakamura and Steinsson (2018) for an interesting discussion of the challenge of achieving identification in macroeconomic estimation.

In order to be able to illustrate the implications of this data frequency tradeoff, both the estimation exercise and policy experiment replications are carried out on both annual and quarterly data. The macroeconomic variables included in the estimation are presented in table 1. This includes six of the seven macroeconomic variables from Smets and Wouters (2007), prepared in line with their definitions. These are the log differences in real per-capita GDP, consumption and investment; the log-differences in real hourly wages and the GDP deflator, as well as the central bank policy rate. The only variable from the Smets and Wouters (2007) dataset that is not included is labour hours relative to long term average, as labour is inelastically supplied in the K+S model. This is replaced instead by the unemployment rate, which has a simulated counterpart in K+S. In addition, in order to allow for evaluation of the financial side of the K+S model and facilitate the identification of monetary policy parameters, the dataset includes the log difference in per capita net worth of the non-financial sector and aggregate loans, as well as the charge-off rate of loans. As will be explained below it is the inclusion of these financial variables, required for identifying the financial side of the Dosi et al. (2015) version of K+S that creates data availability issues.

The quarterly dataset consists of 165 quarters (1984-Q1 to 2025-Q2), built directly from the Federal Reserve Economic Data (FRED) quarterly series listed in table 1. Obtaining an annual dataset is more complicated. Most of the quarterly macroeconomic series used in table 1 start in 1947 or 1948, making it straightforward to obtain annualised versions of the same variables all the way back to the post WWII era. Unfortunately, the quarterly data series for the financial variables, such as commercial loans and loan defaults, are only available from 1984-Q1 onwards, following the US banking crises of the early 1980s and associated regulatory changes. This is not overly restrictive for the quarterly dataset, which retains a reasonable time dimension of 165 quarters. An annualised dataset would only have a time dimension of 41 observations, insufficient to properly identify most parameters.

The solution to this problem is to augment or replace the annualised quarterly data with historical annual data, allowing us to start the financial series as early as possible

Parameters		Calibration	
		Original	US
Mark-up of capital-good firms	μ_1	0.04	0.15
Labour force growth rate	η	0	0.0128 (a) - 0.002 (q)
Tax rates (labour, profits, banks)	tr	0.1	0.25
Central mark-down on reserves	μ_{CB}	0.33	0
Inverse of bond maturity	Θ_B	0.025	0.1714 (a) - 0.04348 (q)
Pareto parameter for bank size distr.	α_B	0.8	0.7
Interest rate spread per credit class	k	0.1	0.07
Technical lifetime of machines	A	20	20 (a) - 80 (q)

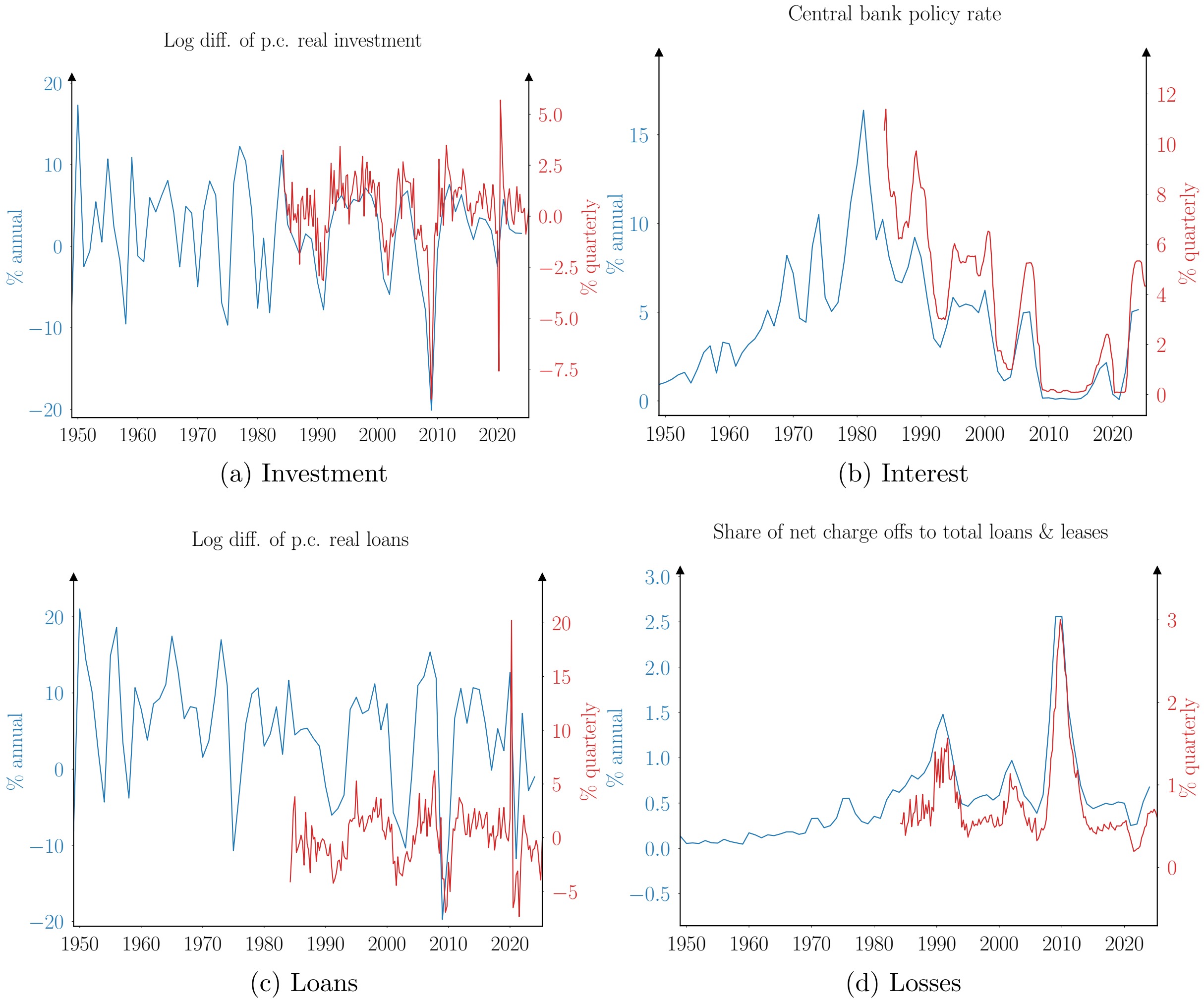
Note: The (a) and (q) mentions after a parameter value indicates the annual and quarterly values respectively.

Table 2: US-calibrated parameter values

and increasing the time-dimension of the annual dataset. Three series are affected. First, between 1949 and 1954 the federal funds policy rate is available as a set of daily high and low values, which were averaged to annual mid-points. The second series is the commercial loans data, where the quarterly level series for commercial and industrial loans is replaced entirely by a separate series, starting in 1947, which directly gives the annual percentage change and only requires deflation and per-capita normalisation. Finally, the most affected series is the net charge-off rate, where the pre-1984 annual rate is calculated from the historical levels of charge-offs and total loans taken from FDIC (1997). This publication provides data until 1996, and the annualised rates from the quarterly series are in good agreement with this historical data from 1984 to 1996. The resulting annual dataset covers 76 years, from 1949-2024. Figure 1 illustrates both datasets for those financial variable 3 variables, as well as the investment variable. The latter plot illustrates an important feature that will become relevant for the policy experiment, which is that while the direction of movement in the annual and quarterly data match, the magnitudes of the growth rates are by construction higher in the annual data.

3.3. Parameter selection and calibration

The second important decision to make is the set of parameters to include in the estimation exercise. As explained above, this is not separate from the question of what empirical data to use. The amount of data available will limit the number of parameters that can be estimated, as identification will require more data the larger the number of



(a) Investment (b) Interest (c) Loans (d) Losses

Figure 1: Empirical variables plot

parameters to estimate. Given the relatively large number of parameters used by the K+S model, a full estimation is infeasible, and thus a choice must be made as to which parameters need to be either estimated or calibrated directly from data.

As stated in the introduction, the purpose of this estimation exercises is ultimately to evaluate the impact of fitting the financial version of the K+S model to empirical data on the resulting fiscal and monetary policy experiments carried out in Dosi et al. (2015). As a result, parameters chosen to be included in the estimation are those that are likely to play a role in the impact of fiscal and monetary policy, in line with the focus of the original Dosi et al. (2015) on exploring the interplay of these two policies. A secondary, but nevertheless significant focus is to also target parameters that can close the gaps between the empirical moments and the moments of the simulated K+S data. As will be shown below, the original calibration of the K+S model produces average growth rates that are significantly above their empirical counterparts, therefore fitting the data will require changes to the parameterisation that underpins the growth process in the simulated economy.

Table 8, in the appendix, lists those parameters whose values are left unchanged for the estimation. This includes most of the initialisation parameters and distributional supports, ensuring that the core behaviour of the firms and households of the K+S model remains broadly similar to that of Dosi et al. (2015). Where possible, US data is used to directly calibrate parameter values. This process, detailed in appendix A, is used to either confirm the existing values of certain behavioural and policy parameters in table 8, as well as directly set the values of the parameters in table 2 to match the US data.

The remaining parameters, listed in table 3, form part of the estimation exercise. These fall into three broad categories, each of which justifies estimating their values from data. First are the deep parameters that govern the stochastic processes which set the trend level of growth in the economy, such as the parameters of the beta distributions that control the innovation process. Next are deep behavioural parameters or linkages between variables in the model that play a role in channelling the effects of policy interventions. These are particularly relevant given the policy focus of the estimation exercise. Examples of the former are the parameters entering the pricing and investment decisions of firms, such as the shares of inventories and capacity utilization targets. For the latter, these are the various elasticities and sensitivities which control the strength of agent responses to price signals, for instance the elasticity of wages to productivity and unemployment, the sensitivity of mark-up adjustment or of the replicator dynamics. Finally, for obvious reasons, the third set of parameters that needs to be estimated contains the monetary policy parameters, which govern the response of the central bank to aggregate fluctuations in prices and unemployment. Note that while single and dual mandate Taylor rules are investigated in the policy experiments of section 5, by default the K+S simulations are run with a dual-mandate Taylor rule, in keeping with the institutional setting of the US Federal Reserve. This is different from the default setting of Dosi et al. (2015), which uses a single-mandate Taylor rule, and has implications for the policy experiment replication in section 5.

In addition, two sets of parameters are transformed before being estimated, either to reduce the overall number of parameters to estimate, reducing in turn the amount of simulation data required for training, or to enforce parameter restrictions. The first case involves the parameters governing the distributions of innovation and imitation draws, both beta distributions shifted to a $[-0.15, 0.15]$ support. In Dosi et al. (2015) the beta distribution for innovation draws is symmetric, implying that a firm's R&D effort is as likely to be successful as it is to be a failure. Letting α_1 and β_1 denote the two parameters of the beta distribution, this is equivalent to assuming that $\alpha_1 = \beta_1$, thus a single parameter governs the entire distribution. For the beta distribution governing imitation draws, in order to reflect the fact that imitation is less valuable than true innovation, Dosi et al. (2015) impose the restriction that $\beta_2 > \alpha_2$. This gives a positive skew to the beta distribution of imitation draws, ensuring that productivity gains from innovation stochastically dominate gains from imitation. This parameter restriction is imposed by estimating α_2 freely and estimating a separate parameter β_2^+ while imposing $\beta_2^+ > 0$ in the prior, and defining $\beta_2 = \alpha_2 + \beta_2^+$ for the purposes of running model simulations.

The second case relates to the AR(4) expectations model (2), where the number of parameters to estimate is reduced by assuming that $\beta_{1,2,3,4}$ follow a geometric progression:

$$E_{t-1}[D_{i,t}] = \beta_1 D_{i,t-1} + \beta_1 \beta_2' D_{i,t-2} + \beta_1 (\beta_2')^2 D_{i,t-3} + \beta_1 (\beta_2')^3 D_{i,t-4} \tag{6}$$

Where $0 < \beta_2' < 1$. This parameter restriction captures the typically decreasing memory of autoregressive processes, while halving the number of parameters required to model the expectations forming process.

4. Estimation results

Three sets of results are covered in this section, corresponding to the 3 methodologies presented in section 3.1. First is the SBC diagnostic carried out on the GP surrogates

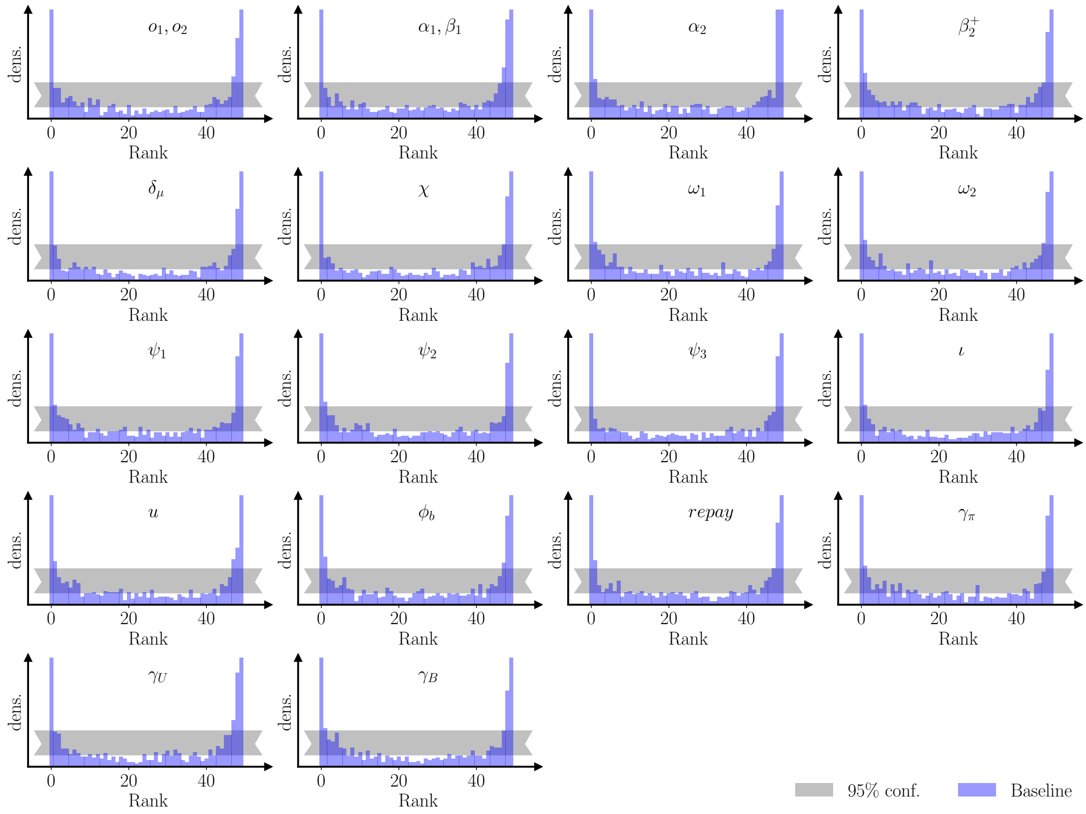


Figure 2: Simulated Bayesian Calibration of baseline expectations model, annual calibration

trained on the simulated K+S models for the 5 different expectation mechanisms (1)-(5), then the presentation of the corresponding parameter estimates obtained on the two empirical datasets, followed by a goodness-of fit of these expectations mechanisms on the data, in order to identify a preferred model specification.

4.1. SBC diagnostic

The results of the SBC analysis for the baseline expectations model using the annual calibrations from table 2 are shown in figure 2. The results for the four additional expectation mechanism as well as the quarterly calibration are very similar, and thus provided in the supplementary material. Under the null hypothesis that the surrogate posterior matches the true, but unobserved posterior, the rank distribution for each parameter should be a uniform distribution. Clearly, in all cases the rank plots exhibit a U-shaped distribution which exceeds the confidence bounds for a uniform distribution, indicating a mismatch between the surrogate and true posteriors.

As explained in section 3.1, the SBC test is useful for verifying the consistency of a

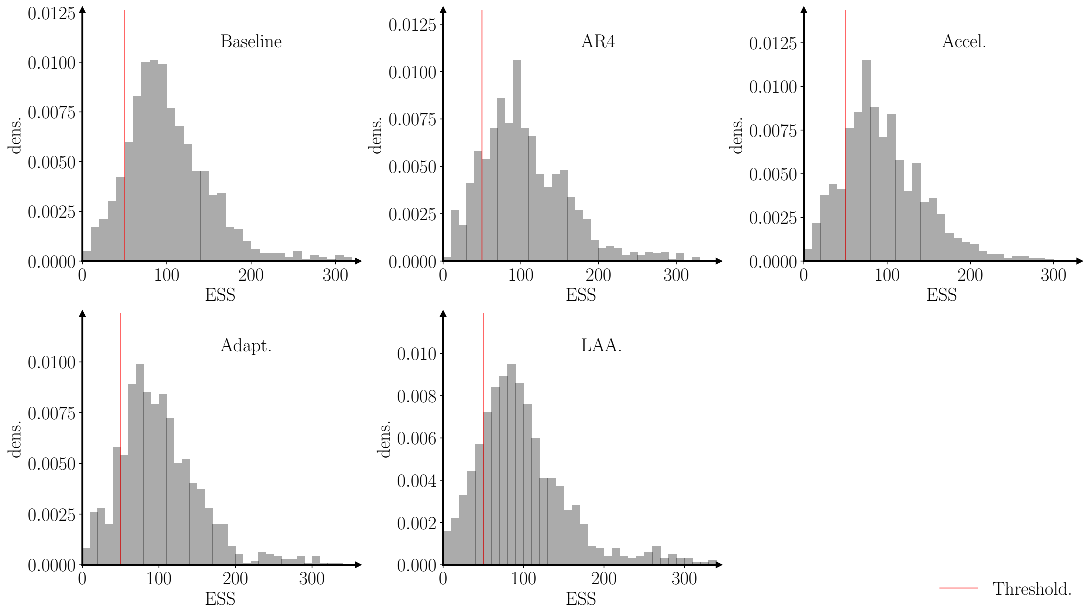


Figure 3: Effective sample sizes for simulation-based calibration, annual calibration, $N_0 = 400$

Bayesian estimation procedure, however when the null is rejected it does not directly point to a cause for this. One possible cause for the U-shape distribution, identified in Talts et al. (2018), comes from serial correlation in the posterior samples. The effective sample sizes (ESS) for each of the 5 K+S specifications, shown in figure 3, confirms some level of serial correlation in the posterior samples, as in all cases the ESS falls below the nominal sample size of 400. Importantly, however, the great majority of SBC samples have an ESS larger than the 50 required in the rank plots. This implies that once thinned, most posterior samples will not suffer from serial correlation. As a result, serial correlation of the posterior samples cannot be the only explanation for the U-shape of the rank plots.

Instead, the U-shape is probably stems from the same misspecification highlighted by Barde (2024) in the VARMA case, examined as part of their benchmarking of the methodology. The GP surrogate in the BEGRS methodology relies on a Vecchia approximation for the simulated data, i.e. it assumes that the simulated data used for training the surrogate GP follows a first-order Markov process. If the data do not satisfy this assumption, then by construction the surrogate will deviate slightly from the truth.

It is likely that this assumption does not hold for the K+S model, as is the case for a VARMA model, which explains the U-shaped SBC rank distributions. As pointed out by Talts et al. (2018), their symmetric shape suggests the surrogate posterior is correctly centred on the true posterior, but the U-shape indicates that it narrower than the true posterior. Crucially, the U-shape of the rank distributions is similar to those in the VARMA case of Barde (2024), yet they show that in the VARMA case the GP surrogate was still able to provide reasonable parameter estimates. This provides confidence that while the distributions for the K+S parameter estimates obtained using BEGRS might suffer from the misspecifications, mean estimates should be reliable.

4.2. BEGRS parameter estimates

The BEGRS estimates of the posterior distributions for all 18 parameters of the baseline and the anchor and adjust (A-A) expectations models (1) are presented in figure 4 for both empirical settings, while the posterior means for the A-A model parameters are provided in the last two columns of table 3. The focus on the A-A model results in addition to the baseline expectations model stems from the fact that as will be discussed here and in section 4.3, this specification seems to outperform the others on multiple dimensions.

The posterior distributions in figure 4 show broad distributions for most parameters, covering the full support of the parameter space. This is not necessarily a surprise, given the relatively small amount of data available and the use of a flat uninformative prior over the parameter range. As is common in macroeconomic estimation, this indicates that parameters are often not entirely identifiable from data alone. This is also consistent with the fact that the parameters estimates obtained on the quarterly data, in red, are sharper than those obtained on the annual dataset, reflecting the larger amount of data available in the former case. The overall implication is that in a lot of cases the original K+S calibrated values of the parameters, indicated on the plots by a vertical red line, are in good agreement with the posterior distributions.

As a result, of this, it is important to focus on those cases where the estimated values of the parameters are clearly different from the original K+S calibrated values. The first case relates to the elasticities of innovation and imitation o_1, o_2, assumed to be equal, which govern the probability that R&D investment results in an innovation.[7] In the quarterly case the distribution is sharp and much lower in expectation than the calibrated value of

[7]The size and sign of the impact on productivity is governed by the beta distribution.

0.3. In the annual case the expected value is also below the calibrated value, although the parameter does not seem well identified in either expectation specification. This can be explained by the use of quarterly data, which imposes that the fitted model produce correspondingly lower rates of growth. These in turn require a slowing down of the imitation/innovation engine, which is achieved by a reduction of the probability of R&D activity resulting in a successful outcome.

Related to this is the estimate of α_1, β_1, which parametrises the distribution of productive innovations, constrained to be symmetric around 0. This estimate is significantly higher than the original value in for both datasets and expectation parameters. As is the case for the o_1, o_2, higher values of α_1 and β_1 ensure that the distribution of productivity improvements is more tightly concentrated around 0, slowing down the rate of productivity (and thus economic) growth exhibited by the model.

The estimates for the sensitivity of mark-up adjustment δ_u are much higher than the calibrated values. This implies stronger market power in consumer good firms, as increases in a firm's market share more easily result in increases mark-ups and prices. In the A-A expectations model, this is compounded by the fact that the replicator dynamics are practically switched off, as the χ parameter which adjusts market shares over time is estimated to be close to 0. The estimated value of the desired inventory level ι also shows a significant increase, especially for the A-A model on the quarterly dataset. This can be explained by the fact, discussed further below in the goodness of fit, that the estimated models experience much higher aggregate volatility than the calibrated model, especially in the quarterly case. In this context, a higher level of inventories and market shares that adjust more slowly help to insulate firms from that volatility.

Figure 5 overlays the posterior densities obtained on the both the annual and quarterly dataset for the different expectation formation mechanisms presented in section 2.2, in order to illustrate the impact of the various expectation specifications on the parameter estimates. A first important observation to make is that again, the posterior densities are stable across a large proportion of parameters. Again, this implies that the most interesting findings are on those areas of divergence across specifications.

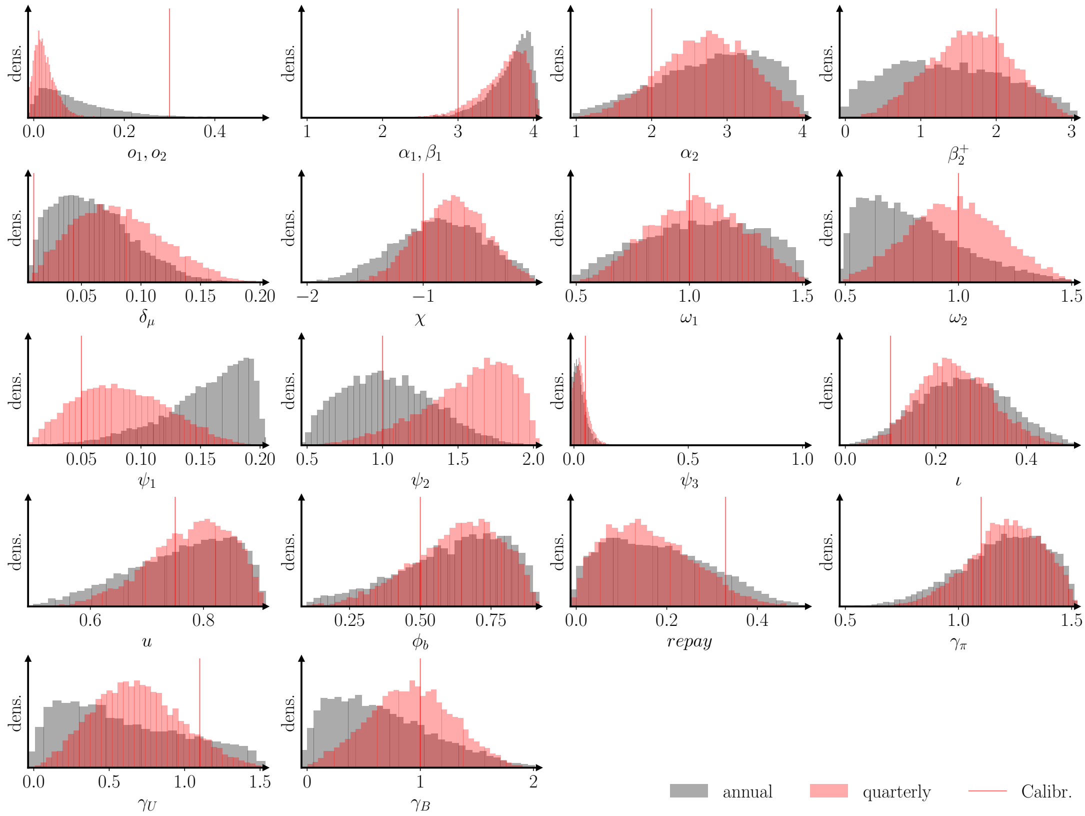
Parameter		Baseline	Prior Range	Posterior Mean Annual	Quartely
Elasticity of innov./imit. w.r.t R&D	o_1, o_2	0.3	0 - 0.5	0.181	0.015
Beta dist. parameters for innov. draws	α_1, β_1	3	1 - 4	3.666	3.553
Beta dist. parameters for imit. draws	α_2	2	1 - 4	2.585	2.735
Beta dist. parameters for imit. draws	β_2^+	2	0 - 3	1.731	1.711
Sensitivity of mark-up adjustment	δ_μ	0.01	0.01 - 0.2	0.050	0.040
Replicator dynamics coefficient	χ	-1	-2 - -0.05	-0.206	-0.215
Competitiveness weight of price	ω_1	1	0.5 - 1.5	1.127	0.942
Competitiveness weight of unfilled demand	ω_2	1	0.5 - 1.5	0.857	0.892
Share of inflation passed to wages	ψ_1	0.05	0.01 - 0.2	0.109	0.094
Elasticity of wages to productivity	ψ_2	1	0.5 - 2	0.833	1.103
Elasticity of wages to unemployment	ψ_3	0.05	0.01 - 1	0.049	0.044
Share of exp. demand held in inventory	ι	0.1	0 - 0.5	0.228	0.300
Planned utilization of machinery	u	0.75	0.5 - 0.9	0.755	0.711
Lending sens. to net worth vs. turnover	ϕ_b	0.5	0.1 - 0.9	0.524	0.596
Desired share of debt to pay back	$repay$	0.33	0 - 0.5	0.195	0.368
Taylor rule sensitivity to target infl.	γ_π	1.1	0.5 - 1.5	1.012	1.023
Taylor rule sensitivity to target unemp.	γ_U	1.1	0 - 1.5	0.601	0.743
Bank sensitivity to financial fragility	γ_B	1	0 - 2	0.760	0.495

Note: The 'Baseline' column indicates parameter values used in Dosi et al. (2015).

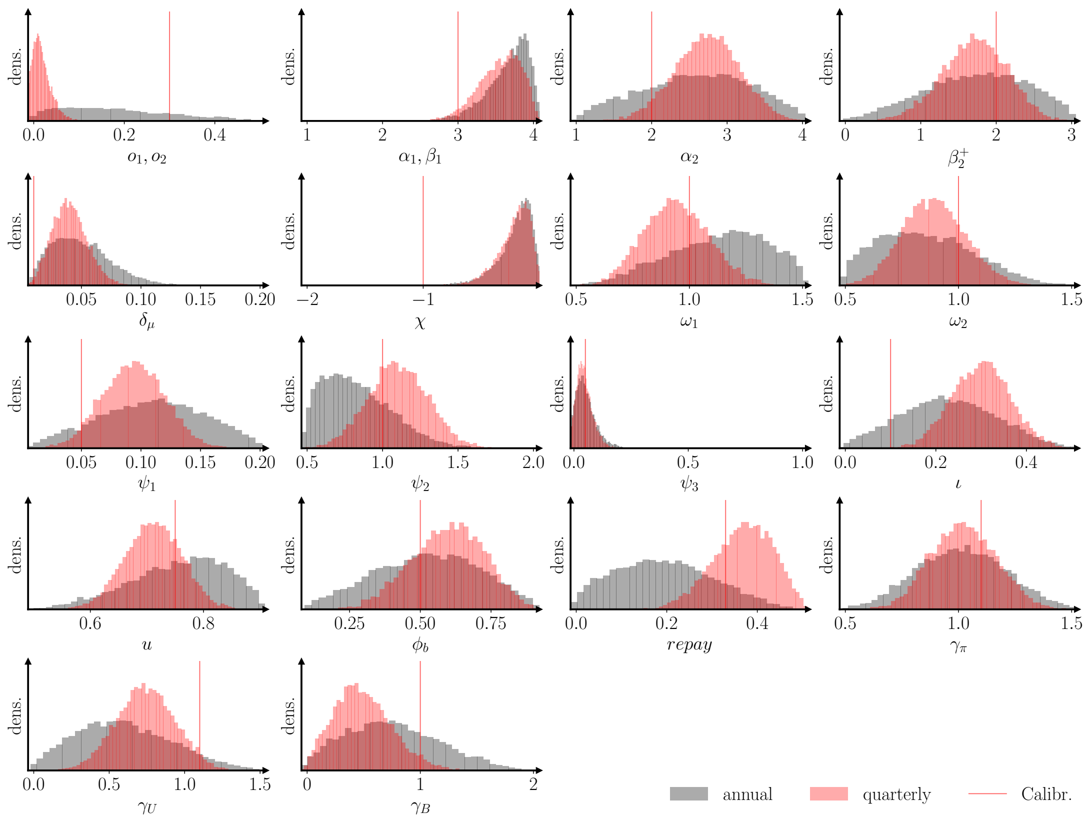
Table 3: Estimated parameters

The first and most obvious feature is the fact that the posterior densities for the A-A expectations mechanism (5) tend to be narrower than for the other mechanisms. While this effect is weaker on the annual data, where as stated above the posterior densities are wider, it is particularly pronounced for the quarterly dataset. Given that both the SBC analysis and the ESS plots in the supplementary material show no major difference between the A-A expectations specification and the other four, this suggests that the A-A specification is better identified given the data available, even in the annual case. The second observation is that again there are several parameters that are sensitive to the choice of expectation mechanism, and display clear differences in posterior densities across expectations specification. Most notably, two parameters from the wage-setting equation ψ_1 and ψ_2, that link unemployment and productivity to wage adjustment are affected, as well as the desired repayment share *repay* and the monetary policy parameters. In the latter case this is not entirely surprising, as the discussion around the Taylor principle that $\gamma_\pi > 1$ in order for monetary policy to be able to ensure macroeconomic stability depends in part on the specification of the expectation-generating process.

Finally, an important point is that the densities in figure 5 are obtained by estimating

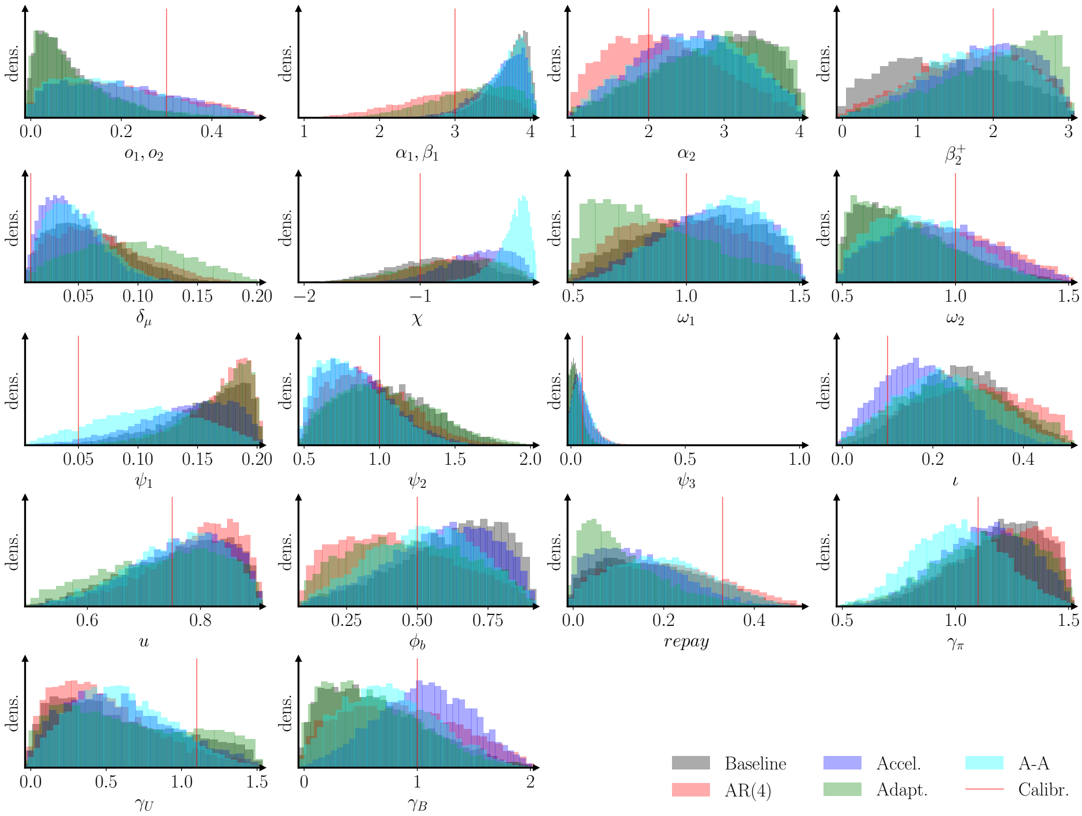


(a) Baseline expectations model

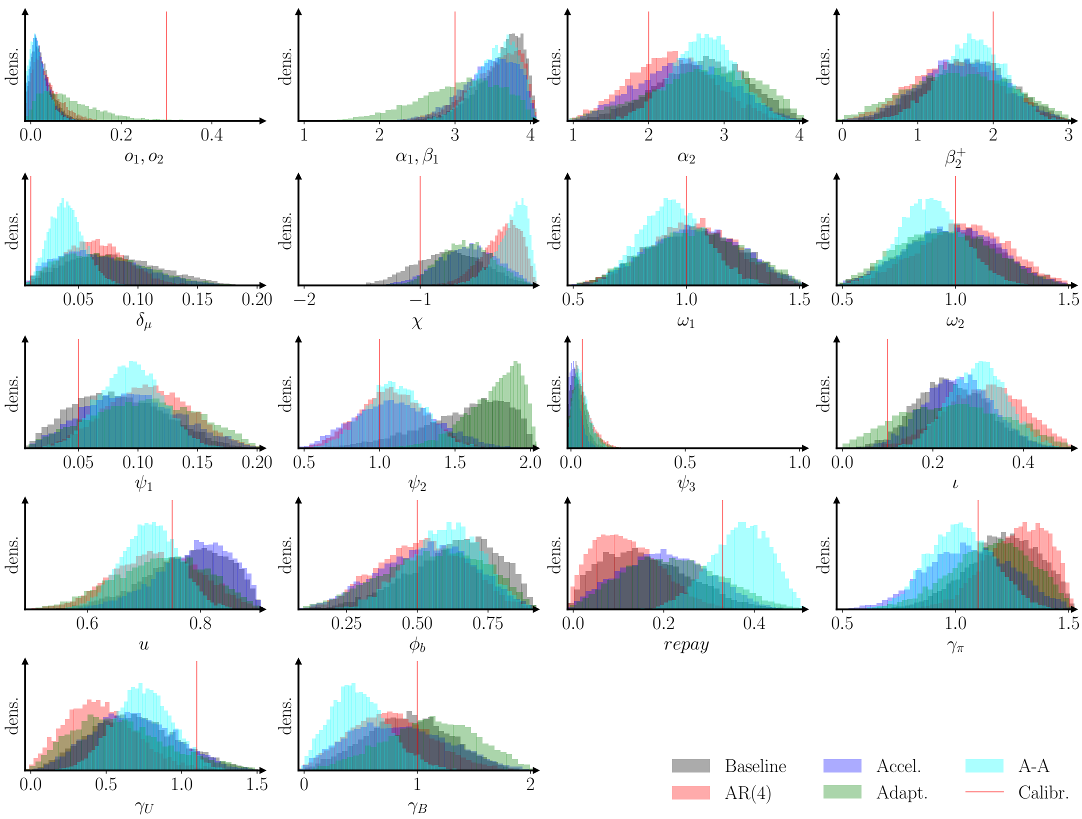


(b) Anchor-adjust expectations model

Figure 4: Posterior densities by empirical dataset



(a) Annual dataset



(b) Quarterly dataset

Figure 5: Posterior densities by expectation mechanism

different expectation specifications, and even when a fixed expectation specification is estimated on different time frequencies, as in figure 4, several calibrated parameters change. Crucially, this means that each expectation specification and data frequency estimation involves a distinct GP surrogate model, trained on distinct simulation runs, each produced by a given specification and calibration of K+S. While clearly those cases where estimates differ should be the focus of interest, the fact that in many cases parameter estimates do not differ across time frequencies and expectation specifications is more evidence that the GP surrogates have been trained reliably, as the independent posteriors they produce match closely over several distinct parameters.

4.3. Goodness of fit and expectation mechanisms

A natural question, given the differences in parameter estimates visible in figure 5, is what difference, if any, the expectation mechanism makes to the goodness-of-fit on the data. In order to provide some context for this, table 4 provides the mean and standard deviation of the 10 variables, for both empirical and simulated data.[8] For the two empirical columns, these descriptive statistics are obtained from the empirical datasets detailed in section 3.2. For the simulated data, the columns use the original calibration, the US-based calibration, setting the values of the parameters listed in table 2 to their annual and quarterly values, and the annual and quarterly mean estimates from table 3. In all cases the descriptive statistics are generated using the 1000 simulated runs of 300 observations used to train the MIC.

Examination of the first two columns of table 4 shows that the original calibration produces relatively high growth rates for output, consumption, investment, real wages and net worth, of around 3% per period. This is combined with low unemployment, at less than 1%, and an inflation rate of about 4%. A key feature of the model is that investment is more volatile than GDP, with consumption showing less volatility. While this investment accelerator is a central feature of the model's design, resulting from the Schumpeterian innovation process, its size is much larger than in the empirical data, where

[8] In the interest of space, plots of the annual and quarterly simulations are provided in the supplementary material for illustrative purposes.

	Calibrated simulations			Annual Estimation		Quarterly Estimation	
	Original	US Annual	US Quarterly	Empirical	Simulated	Empirical	Simulated
Δy	3.008	3.296	3.145	1.804	3.023	0.389	1.646
	(3.379)	(3.500)	(3.282)	(2.284)	(19.029)	(1.051)	(39.632)
Δc	3.010	3.299	3.147	1.888	3.022	0.448	1.647
	(1.768)	(1.918)	(1.773)	(1.796)	(10.206)	(1.153)	(15.739)
Δi	2.992	3.267	3.135	1.919	3.042	0.368	1.605
	(27.535)	(31.763)	(26.458)	(6.018)	(62.676)	(1.826)	(134.498)
ΔL	3.002	3.042	3.133	4.995	2.974	0.030	1.523
	(3.535)	(4.865)	(3.412)	(7.810)	(14.124)	(2.882)	(11.398)
Δnw	3.004	3.273	3.128	1.965	3.013	0.442	1.695
	(3.672)	(3.792)	(3.453)	(7.265)	(20.393)	(3.183)	(14.200)
Δw	3.010	3.304	3.150	1.479	3.018	0.263	1.652
	(1.471)	(1.517)	(1.472)	(1.693)	(1.890)	(1.070)	(2.499)
u	0.864	2.673	1.379	5.708	12.873	5.789	19.895
	(3.043)	(10.128)	(6.703)	(1.617)	(15.748)	(1.655)	(22.625)
r	9.817	9.032	9.665	4.365	3.680	3.582	4.094
	(3.181)	(3.537)	(3.161)	(3.458)	(4.026)	(2.866)	(6.758)
π	4.165	3.983	4.134	3.081	2.168	0.588	2.590
	(2.043)	(1.851)	(1.863)	(2.132)	(3.624)	(0.338)	(8.298)
lr	2.469	3.838	1.047	0.535	10.184	0.777	4.954
	(1.535)	(3.436)	(2.406)	(0.482)	(7.708)	(0.515)	(4.117)

Note: means provided for each variable, with standard deviations in parenthesis. The US calibration is obtained by imposing the parameter values in table 2 to the original calibration (annual and quarterly). The estimated simulations are obtained by imposing the parameter values in table 3 to the original calibration (annual and quarterly).

Table 4: Comparison of simulated and empirical descriptive statistics

the standard deviation of investment is only two to three times that of GDP. Imposing the US calibrated parameters of table 2 does not significantly modify the descriptive statistics, beyond lowering the loss ratio on loans to a more realistic value.

Given this, one can see that simulating the K+S models with the US-estimated parameters moves the simulated moments towards the empirical ones, for both empirical datasets. This again provides strong evidence that the BEGRS methodology is effective in delivering parameter estimates that genuinely shift the simulations in the direction of the empirical data. For the annual estimates, the main effect is an increase in the simulated unemployment rate and a reduction in the inflation rate and policy rates. For the quarterly estimates, this also involves a significant reduction the average growth of the macroeconomic aggregates (GDP, consumption, investment, wages, net worth) towards the lower values observed in the data. As highlighted in the previous section, the main mechanism for this is a slowing down of the rate of innovation, and thus productivity growth, in the model. Unfortunately, as the K+S economy slows down one also observes a dramatic increase in the volatility of the economy, visible in several locations. The first is the increase in the standard deviation of aggregates, particularly investment, which becomes unrealistically high for both sets of estimates, but is more pronounced for the quarterly estimates. The second location are the very high rates of unemployment and unemployment volatility observed in the K+S simulations based on US parameter estimates, again higher in the quarterly case.

These findings suggest that while the K+S parameter space does allow for the possibility of lower growth rates, this comes at the expense of relatively violent macro-economic crises. These lower growth rates are essentially produced by simulating an economy prone to deep crises and periods of depression, with little to no investment and high unvoluntary unemployment. As will be shown in the policy experiments of section 5, the behaviour of the K+S model changes when it is in this low-growth regime. The fact that this effect is present in both estimates but stronger in the quarterly setting suggests that this increase in volatility as growth slows is a gradual transition.

As explained in section 3.1, a more rigorous measure of the goodness of fit of the simulations with respect to the empirical data is obtained using the multivariate Markov Information Criterion (MIC) of Barde (2020). Table 5 shows the MIC scores obtained for each variable, as well as the aggregate MIC for the dataset.[9] The final column of the table shows the difference in MIC scores relative to the original calibration, with negative values indicating a better fit and positive values a worse one. This reveals that with the exception of the AR4 expectation model, all estimated specifications improve on the original calibration, in both empirical samples, as does the US calibration using table 2.

In the quarterly setting, the largest improvement is brought by the A-A expectation model, by a wide margin relative to other model, and this seems driven by a significantly improved performance on the loss ration lr and the real wage growth Δw. The picture is not as clear in the annual setting, where the A-A does not bring the largest improvement. This is consistent with the findings of section 4.2 above, which suggest that the A-A model is better identified than the others in the quarterly case, with a less clear picture in the annual case. A comparison of the descriptive statistics across all expectation specifications, however, reveals that the instability and volatility of the A-A mechanism

[9]Note that the aggregate MIC is not equal to the sum of the variable-level MIC scores, as it takes into account the mutual information that may exist between variables.

	Δy	Δc	Δi	ΔL	Δnw	Δw	u	r	π	lr	Aggr	Diff.
Annual data, $L_0 = 450$												
Original calibration	410.74	359.50	366.34	705.23	479.45	359.53	531.18	464.06	395.88	809.74	4797.04	0.00
US calibration	411.92	395.77	370.62	744.76	485.27	422.59	438.79	482.29	404.44	664.48	4624.12	-172.92
US est, baseline exp.	435.80	387.03	400.61	671.70	466.32	396.72	370.12	479.95	372.50	219.55	4121.95	-675.09
US est, AR4 exp.	349.26	311.10	487.72	1052.51	537.41	404.45	717.35	1005.81	427.12	215.65	5703.24	906.20
US est, Accel. exp.	468.06	419.60	437.92	645.81	440.27	421.19	392.14	488.94	370.79	218.81	4257.99	-539.06
US est, Adapt. exp.	405.27	362.74	374.52	651.74	456.02	420.63	335.88	479.64	345.91	237.58	4041.30	-755.74
US est, LAA. exp.	567.23	518.24	436.25	524.75	515.52	381.82	504.40	492.74	408.87	306.97	4409.67	-387.37
Quarterly data, $L_0 = 984$												
Original calibration	792.37	825.44	792.66	772.42	865.69	1031.67	978.95	1068.79	1006.34	1486.99	10087.60	0.00
US calibration	808.48	867.51	774.47	928.14	885.14	1088.51	699.58	1072.19	949.01	466.63	9081.37	-1006.23
US est, baseline exp.	939.16	754.17	757.42	903.92	929.69	821.74	826.21	1019.74	1060.94	733.57	8832.54	-1255.06
US est, AR4 exp.	1170.72	944.19	977.96	923.33	888.36	985.99	1264.47	1780.88	1019.47	2861.08	13060.13	2972.54
US est, Accel. exp.	1116.02	941.01	808.81	878.60	975.85	892.37	933.97	1014.59	996.29	800.90	9449.89	-637.71
US est, Adapt. exp.	1008.63	687.20	737.55	813.28	865.54	1048.42	1131.23	1391.14	955.82	778.02	9540.48	-547.12
US est, LAA. exp.	1075.16	856.52	936.99	788.94	940.48	686.98	993.59	1006.97	968.04	524.47	8581.55	-1506.05

Note: The first 10 columns provide the MIC score for the corresponding empirical variable, and the 'Aggr' column provides the aggregate MIC score over the entire dataset. Note that the aggregate score is not equal to the sum of the variable level scores, as it accounts for mutual information between variables. 'Diff' indicates the MIC score relative to the original calibration, with negative values indicating a preferred specification and positive values a worse specification.

Table 5: MIC scores, 6-bit discretisation, 1 lag

visible in table 4 is in fact lower than that generated by the other specifications[10] Combined with that the annual dataset contains significantly fewer observations than the quarterly dataset, potentially explaining that the lower MIC performance, this suggests that the A-A model is the preferred specification overall. This overall performance of the A-A specification is probably due to the fact that in the context described above, where firm-level investment becomes very volatile, expectations based on past demand alone, such as (1) - (4), are likely to themselves be very noisy. By contrast the anchor-and-adjust mechanism (5), which bases the expectation on aggregate GDP fluctuations, is probably able to filter out the noisy firm-level demand signal and provide more reliable expectations.

Finally, it is important to note that the overall goodness-of-fit of all models remains poor, as one might expect from the descriptive statistics in table 4. Because the MIC is based on a minimum description length approach (Grünewald, 2007), one can compare the score obtained to the original message length, which is 450 bits per variable (4500 on aggregate) for the annual dataset and 984 bits per variable (9840 on aggregate) for the quarterly dataset. Scores that are higher than those thresholds, either at the variable level or on aggregate indicate that the one-step-ahead predictive distribution induced by the simulated data is less informative than an uniform distribution. This is the case for the original calibration and the US calibration, for both annual and quarterly models. While most of the estimated specifications shown in table 5 do fall below those thresholds, indicating that the predictive power of these models outperforms a uniform distribution, the improvements are relatively small. This suggests that there is still room for improvement, however, this probably involves either a larger dataset, in order to better identify model parameters, or modifications to the model itself, which are discussed in section 6.

5. Policy experiments

The final step of the analysis is to replicate the fiscal and monetary policy experiments carried out in Dosi et al. (2015), in order to evaluate the sensitivity of these policy experiments to the parameter estimates obtained in the previous section. By default,

[10] A full table of descriptive statistics across all specifications is provided in the supplementary material.

K+S simulations do not constrain government deficits or debt, which are allowed to vary freely with the macro-economic context (labelled *No rule*). The first fiscal rule examined by Dosi et al. (2015) is the 3% rule of the *European Stability and Growth Pact* (*SGP*), whereby the government deficit is limited to 3% of GDP. When the rule is binding, government spending in the period is reduced accordingly. The second fiscal rule mirrors the *Fiscal Compact* (*FC*) and states that if the stock of government debt exceeds 60% of GDP, government spending should be reduced to ensure that 5% of the gap between the actual debt-to-GDP ratio and the 60% target is closed in each period.[11] Both the FC and SGP policies are also tested with an *escape clause* setting (SGP_{ec} and FC_{ec}), in which the fiscal rules are binding only when GDP growth is positive.

On the monetary policy side, three settings are examined. The first two are the single and dual-mandate Taylor rules, TR_{π} and $TR_{\pi,U}$. As previously mentioned, the dual-mandate Taylor rule forms the default setting for the estimation exercise, due to the use of the US empirical data. The single mandate rule is achieved by simply setting $\gamma_u = 0$ in simulations. A third monetary setting involves a bond spread rule, *spread*, where the interest rate on government bonds is correlated to the debt-to-GDP ratio as follows, with $\rho = 0.04$.

$$r_t^{bonds} = r_t \left(1 + \rho \frac{Debt_{t-1}}{GDP_{t-1}}\right) \tag{7}$$

This mechanism essentially adds a risk premium to government debt, where excessive borrowing relative to GDP is reflected in gradually increasing bond rates. An important clarification is that because the default monetary policy used in the BEGRS estimation of section 4.2 is the dual mandate Taylor rule, when the *spread* risk premium is applied in the policy experiments, it is added to this dual mandate Taylor rule. This differs from the original Dosi et al. (2015) experiments, where the default is the single-mandate Taylor rule and the *spread* risk premium is applied to that default.

[11] Note that in Dosi et al. (2015) this corresponds 5% per year, and a concern keeping this 5% target while shifting to a quarterly period increases the severity of this policy. The results of the analysis, however, show that this has little impact in practice.

	Original calibration			US calibration (annual)			US calibration (quarterly)		
	TR_{π}	$TR_{\pi,U}$	*spread*	TR_{π}	$TR_{\pi,U}$	*spread*	TR_{π}	$TR_{\pi,U}$	*spread*
Mean GDP growth									
No rule	1.000	1.023***	1.022***	1.000	1.011***	1.011***	1.000	1.060***	1.060***
	(0.000)	(10.097)	(9.598)	(0.000)	(7.123)	(7.123)	(0.000)	(29.470)	(29.479)
SGP	0.320***	0.997	0.998	1.012***	1.011***	1.009***	−0.974***	1.039***	1.031***
	(18.645)	(0.439)	(0.267)	(8.748)	(6.600)	(5.905)	(57.930)	(5.467)	(3.403)
FC	0.295***	0.978*	0.998	−0.751***	0.882***	0.930***	−1.059***	1.035***	1.038***
	(19.180)	(1.946)	(0.314)	(49.507)	(6.332)	(4.609)	(58.767)	(4.094)	(4.857)
SGP_{ec}	0.994**	1.023***	1.019***	1.019***	1.013***	1.009***	0.965***	1.064***	1.061***
	(2.524)	(11.026)	(8.051)	(13.552)	(8.945)	(4.809)	(17.402)	(35.178)	(27.638)
FC_{ec}	0.991***	1.023***	1.019***	−0.885***	0.933***	0.939***	0.959***	1.064***	1.061***
	(3.622)	(10.662)	(8.488)	(55.066)	(5.285)	(5.039)	(20.320)	(35.982)	(29.732)
Mean GDP growth volatility									
No rule	1.000	0.855***	0.856***	1.000	0.812***	0.812***	1.000	0.735***	0.735***
	(0.000)	(18.127)	(17.813)	(0.000)	(19.893)	(19.893)	(0.000)	(41.676)	(41.707)
SGP	18.745***	1.697***	1.703***	5.645***	1.178	1.689***	44.306***	1.138	1.331*
	(19.156)	(2.992)	(3.733)	(33.104)	(1.287)	(2.738)	(57.723)	(0.855)	(1.844)
FC	19.953***	2.099***	1.922***	73.749***	3.679***	2.077***	47.029***	1.095	1.315
	(20.188)	(4.188)	(3.430)	(63.540)	(5.528)	(3.630)	(65.561)	(0.899)	(1.579)
SGP_{ec}	1.460***	0.879***	0.935	5.329***	1.175	1.445***	4.494***	0.810***	1.006
	(15.868)	(12.035)	(1.358)	(28.769)	(1.570)	(2.592)	(24.292)	(5.060)	(0.069)
FC_{ec}	1.597***	0.890***	0.911***	68.595***	2.293***	2.174***	4.708***	0.816***	0.966
	(17.380)	(8.816)	(4.119)	(54.511)	(3.758)	(3.084)	(27.346)	(4.548)	(0.470)
Mean unemployment rate									
No rule	1.000	0.179***	0.193***	1.000	0.087***	0.087***	1.000	0.054***	0.055***
	(0.000)	(19.414)	(18.461)	(0.000)	(58.698)	(58.698)	(0.000)	(67.315)	(67.061)
SGP	5.893***	0.414***	0.464***	1.939***	0.084***	0.110***	3.385***	0.081***	0.092***
	(19.713)	(8.125)	(7.281)	(65.768)	(58.392)	(50.849)	(83.765)	(53.848)	(48.995)
FC	6.222***	0.518***	0.434***	3.247***	0.202***	0.136***	3.490***	0.084***	0.078***
	(20.270)	(5.699)	(7.748)	(176.209)	(32.130)	(41.768)	(92.444)	(52.163)	(53.906)
SGP_{ec}	1.594***	0.178***	0.192***	1.825***	0.084***	0.103***	1.770***	0.041***	0.066***
	(7.154)	(19.347)	(17.342)	(53.530)	(58.190)	(53.440)	(32.012)	(71.937)	(59.700)
FC_{ec}	1.798***	0.187***	0.195***	3.237***	0.178***	0.161***	1.886***	0.041***	0.066***
	(8.903)	(18.656)	(17.485)	(174.243)	(35.496)	(37.810)	(38.868)	(71.678)	(60.212)
Mean inflation rate									
No rule	1.000	1.196***	1.195***	1.000	2.116***	2.116***	1.000	2.118***	2.118***
	(0.000)	(20.613)	(20.410)	(0.000)	(171.911)	(171.911)	(0.000)	(157.078)	(157.154)
SGP	0.936***	1.191***	1.186***	0.998***	2.122***	2.110***	0.912***	2.112***	2.107***
	(4.500)	(19.543)	(18.860)	(3.494)	(182.634)	(157.551)	(9.865)	(139.221)	(134.793)
FC	0.937***	1.186***	1.187***	0.892***	2.080***	2.110***	0.915***	2.111***	2.112***
	(4.428)	(18.732)	(19.129)	(16.888)	(118.171)	(149.525)	(9.695)	(140.746)	(139.361)
SGP_{ec}	0.981	1.197***	1.199***	1.003***	2.122***	2.110***	0.993	2.128***	2.116***
	(1.431)	(20.692)	(20.883)	(6.182)	(183.690)	(157.562)	(0.972)	(172.182)	(147.144)
FC_{ec}	0.972**	1.196***	1.199***	0.894***	2.086***	2.100***	0.989	2.128***	2.116***
	(2.056)	(20.572)	(20.808)	(16.667)	(122.712)	(136.937)	(1.456)	(172.170)	(146.983)

Note: For each variable and policy combination, the entries provide the mean value over the replication relative to the mean value of the $No\ rule, TR_{\pi}$ policy combination. T-statistics for the difference in means are provided in parentheses, with *, ** and *** indicating significance at the 10%, 5% and 1% levels respectively.

Table 6: Policy experiments for calibrated models

5.1. Main findings

Tables 6, 7 and 10 show the results of the policy experiments obtained using 1000 Monte Carlo replications. In the interest of space, we present the impact on four variables (which corresponds to tables 1, 2, 3 and 5 in Dosi et al. (2015)). As is the case in Dosi et al. (2015), in each panel of the table the Monte Carlo average obtained for a given policy combination is divided be the average of the *No rule* / TR_{π} combination, which is the default K+S setting in their analysis. While this does not correspond to the default used here, which is the *No rule* / $TR_{\pi,U}$ combination, this choice ensures comparability with their original analysis. The t-statistics for differences in the mean are provided in parenthesis below each ratio, with stars indicating significance at the 10%, 5% and 1% respectively.

Table 6 shows the results of the policy experiments for the two calibrated models, the left-most panels represent the results obtained for the original calibration, to check for correct replication, the middle panels those obtained with the annual values of the US-calibrated parameters from table 2 and the right-most the quarterly values. As one would expect, the original calibration results in the left-most panels are generally in good agreement with the original tables in Dosi et al. (2015), allowing for the difference in Monte Carlo replications, 100 in the original work versus 1000 here. As highlighted above, the *spread* column has a slightly different interpretation, and should be compared to the $TR_{\pi,U}$ column here, not the TR_{π} column. Comparison of the *spread* and $TR_{\pi,U}$ columns in each panel shows very little quantitative difference, a finding that is consistent across all tables. This suggests that including a risk premium to bond rates in addition to the default monetary policy does not substantially modify the behaviour of the model, which is entirely consistent with the findings of Dosi et al. (2015).

The US calibrations, in the middle panels of table 6 for the annual calibration and the right-most panels for the quarterly calibration, share several qualitative similarities with the original calibration, but with amplified quantitative effects. In the case of a single-mandate Taylor rule TR_{π}, under binding fiscal constraints (*SGP* or *FC*) GDP growth becomes negative, rather than simply lower, relative GDP volatility raises 44-fold, with

	Annual estimates			Quarterly estimates		
	TR_π	$TR_{\pi,U}$	*spread*	TR_π	$TR_{\pi,U}$	*spread*
Mean GDP growth						
No rule	1.000	1.000	0.999	1.000	0.985*	0.988
	(0.000)	(0.080)	(0.281)	(0.000)	(1.859)	(1.588)
SGP	1.001	0.991**	1.003	1.011	1.010	0.998
	(0.407)	(2.420)	(1.247)	(1.453)	(1.225)	(0.297)
FC	0.983***	−0.326***	−0.414***	1.011	0.588***	0.387***
	(4.939)	(30.646)	(31.729)	(1.133)	(9.606)	(12.645)
SGP_{ec}	1.000	1.000	0.997	1.006	1.002	0.984*
	(0.162)	(0.022)	(0.782)	(0.733)	(0.305)	(1.960)
FC_{ec}	0.985***	−0.082***	−0.135***	1.014	0.517***	0.575***
	(4.256)	(31.521)	(32.484)	(1.446)	(11.835)	(9.999)
Mean GDP growth volatility						
No rule	1.000	0.871***	0.875***	1.000	1.202***	1.195***
	(0.000)	(20.223)	(19.253)	(0.000)	(9.373)	(8.768)
SGP	1.055***	1.046	0.921***	1.118***	1.376***	1.351***
	(8.732)	(0.739)	(5.338)	(5.159)	(14.802)	(13.232)
FC	1.469***	9.662***	10.104***	1.434***	5.862***	6.498***
	(38.236)	(28.607)	(29.768)	(15.099)	(39.005)	(39.954)
SGP_{ec}	1.094***	1.011	1.086	1.126***	1.411***	1.411***
	(12.012)	(0.224)	(1.056)	(5.568)	(15.913)	(14.387)
FC_{ec}	1.469***	7.916***	8.143***	1.436***	6.650***	6.474***
	(36.277)	(27.470)	(27.311)	(15.152)	(32.483)	(33.819)
Mean unemployment rate						
No rule	1.000	0.805***	0.805***	1.000	1.120***	1.110***
	(0.000)	(18.069)	(18.069)	(0.000)	(7.581)	(6.903)
SGP	0.989*	0.819***	0.795***	1.033**	1.121***	1.097***
	(1.719)	(12.190)	(19.551)	(2.094)	(7.370)	(6.062)
FC	1.720***	4.111***	4.236***	1.243***	2.808***	3.060***
	(51.498)	(43.293)	(45.800)	(13.635)	(36.590)	(39.338)
SGP_{ec}	1.030***	0.829***	0.834***	1.047***	1.157***	1.147***
	(4.043)	(12.096)	(11.093)	(2.967)	(9.476)	(8.842)
FC_{ec}	1.709***	4.055***	4.129***	1.242***	2.766***	2.773***
	(49.856)	(42.881)	(44.356)	(13.647)	(36.451)	(37.749)
Mean inflation rate						
No rule	1.000	1.173***	1.170***	1.000	0.966***	0.967***
	(0.000)	(21.124)	(21.054)	(0.000)	(5.972)	(5.600)
SGP	1.022***	1.180***	1.186***	0.998	0.981***	0.984***
	(4.943)	(21.459)	(23.459)	(0.375)	(3.336)	(2.771)
FC	0.944***	0.899***	0.882***	0.968***	0.745***	0.731***
	(11.533)	(8.565)	(10.031)	(5.468)	(28.498)	(30.356)
SGP_{ec}	1.007	1.174***	1.174***	0.990*	0.970***	0.965***
	(1.557)	(20.494)	(21.089)	(1.646)	(5.178)	(5.807)
FC_{ec}	0.947***	0.853***	0.854***	0.968***	0.740***	0.730***
	(10.837)	(11.517)	(12.048)	(5.467)	(29.782)	(29.720)

Note: For each variable and policy combination, the entries provide the mean value over the replication relative to the mean value of the *No rule*, TR_π policy combination. T-statistics for the difference in means are provided in parentheses, with *, ** and *** indicating significance at the 10%, 5% and 1% levels respectively.

Table 7: Policy experiments for A-A expectation model

high unemployment and lower inflation. Both SGP_{ec} and FC_{ec} provide a more benign environment, albeit still significantly less attractive that the *No rule* setting. The dual-mandate Taylor rule $TR_{\pi,U}$ reveals a very different picture, with improved GDP growth under all fiscal rules, improved or unchanged GDP volatility, and dramatically lower unemployment, at the cost of higher inflation. Again, this is simply a quantitatively magnified version of the original Dosi et al. (2015) findings.

The picture changes radically once one shifts to the policy experiments run with estimated parameters, as all display reversed behaviour for monetary policy. Table 7 shows the policy experiments for the A-A expectations model, which is the preferred model from an identification and goodness-of-fit perspective, using simulations obtained by running K+S on both annual and quarterly estimates from table 3. Examination of the single mandate Taylor rule TR_{π} reveals that the different fiscal policies no longer have any effect relative to the *No rule* setting, with the exception of slightly elevated GDP volatility and unemployment. It is now the dual-mandate Taylor rule $TR_{\pi,U}$ that provides worse macroeconomic outcomes: lower and more volatile growth under most fiscal rules, combined with higher unemployment a lower inflation. Another important result is that the extreme relative volatility under SGP and FC has completely disappeared.[12]

This key finding of a reversal in the impacts of single and dual-mandate Taylor rules is even stronger in the baseline estimates, shown in table 10 in appendix. Here we see a relatively small impact of all fiscal rules under TR_{π}, with slightly better growth, and lower unemployment with a higher volatility level and more inflation. Under $TR_{\pi,U}$ growth becomes negative, GDP volatility explodes and unemployment is much higher, all very amplified effects that are similar to what the US calibration exhibits under the TR_{π} monetary policy. Policy impacts for the accelerative and adaptive expectation models show a very similar picture, and so are provided in the supplementary material.

[12]It is important to note that this is probably due to the fact, mentioned above, that in a low-growth regime the K+S model exhibits much higher volatility overall, regardless of policy.

5.2. Discussion: the interaction of the expectations mechanisms and low-growth regime

As explained in sections 4.2 and 4.3, moving the simulated moments of the K+S model towards their empirical counterparts necessarily requires reducing the growth rates of macroeconomic variables relative to the original calibration, which is achieved primarily by reducing the rate of arrival and magnitude of productive innovations. Given the structure of the model, these lower levels of growth in productivity do result in a reduction in output growth, but also increase the volatility of the economy, and of investment behaviour in particular. This is already visible in the annual estimates, but becomes much more pronounced on the quarterly data, where the absolute value of growth rates are lower.

The key finding of the policy experiment is that this change in growth regime affects the interaction between expectation mechanisms and model behaviour. In this more volatile context, expectations that are too myopically tied to the demand faced by firms contribute to the instability. The performance of the A-A expectations observed within this context is probably linked to the fact that by linking expected demand to aggregate fluctuations they dampen the volatility of demand. In the relatively high and stable growth environment of the original policy experiments, the role of the expectation mechanism is not critical, explaining the aforementioned report in Dosi et al. (2015) that the different mechanisms essentially produce the same qualitative behaviour. This is not the case with the low-growth regime generated using the estimated parameters. The posterior distributions shown in figure 5, the goodness-of-fit results in section 4.3 and the policy experiments above show that the choice of expectation mechanisms can indeed matter within the K+S framework, and make a significant difference to the model behaviour.

To be clear, this does not invalidate the original findings in Dosi et al. (2015) that under the original calibration, with strong and steady innovation and productivity growth, the choice of expectation mechanism has little impact on the behaviour of the model. Instead, the aim is to establish that in an edge case where the economy is suffering from protracted crises, the behaviour of the model becomes much more sensitive to the expectations mechanism. This suggests at minimum that the K+S model exhibits different behaviour depending on the growth regime, which should be taken into account when selecting the expectations mechanism.

Importantly, given the relatively poor quality of the fit obtained, we make no claim that these policy experiments, or indeed the parameter estimates, correctly reflect some underlying truth. In fact, it is highly likely that both the parameter estimates and policy experiments would again change should future researchers introduce changes in the model mechanisms that reduce the volatility of the economy in low-growth environments. Instead, the key takeaway from the findings of the policy experiment is that the interaction of the model design and expectation mechanisms can depend on the parameter settings, and care should be taken to validate these mechanisms against data when making applied and quantitative policy recommendations.

6. Conclusion

This paper demonstrates that it is feasible to perform Bayesian estimation of a relatively large ABM model using a limited amount of simulation runs, obtain reasonable parameter estimates that improve the goodness-of-fit on empirical data, perform a selection exercise on a set of alternative specifications for expectations, and gain valuable insights into the behaviour of the model. As was highlighted throughout, the results obtained from this empirical exercise are by no means perfect, but they nevertheless provide a proof-of-concept for future research aiming to take the models to the data and better understand, as a result, which ABM mechanisms to include and how to approach model design. Given that many possible competing mechanisms could be considered for inclusion in any given economic ABM model, having a clear validation process that can provide parameter estimates and select amongst resulting mechanisms is critical for ensuring empirical applicability of the model.

In fact, the general takeaway from the exercise is less about the actual parameter estimates obtained than it is about providing an insight into how estimation workflows, such as the one presented here, can serve as disciplining device as part of an iterative design process, improving both the design and validation of ABMs. Initial estimation and selection exercises can highlight areas of improvement in a model at an early stage, allowing for better performance on the data in later estimation steps. This would ensure that validation of the ABM is directly embedded during the design phase, resulting in more empirically reliable models. For large ABMs, where the challenge of validation of the full set of model parameters and mechanisms remains, a potential strategy in this respect is to break up the validation of the model into stages. First one would perform unit testing of the individual mechanisms in separate model modules, for example using rich micro-level or experimental data. Once this is done, the estimation of deep macroeconomic parameters governing the markets connecting these modules can be carried out in a second stage, using an all-up estimation strategy similar to the one used here.

Several suggestions emerge from the overall empirical exercise regarding improving the performance of the K+S model, particularly for ensuring greater stability in the macroeconomic aggregates in low-growth regimes. Of particular concern from an empirical point of view is the unrealistically high ratio of investment volatility relative to GDP volatility, around 8 in the original calibration, which increases even further for low levels of growth. This mainly results from the very stylised input-output structure, with a single capital sector and a single downstream sector. While this is extremely effective at demonstrating how disruptive innovation in the upstream sector, and coordination of investment decisions in the downstream sector can generate such fluctuations, it leads to an exaggeration of their size. As an example, including a range of capital goods sectors connected in an input-output structure would allow for similar Schumpeterian innovation dynamics with each sector, while dampening the effect on fluctuations. A radical innovation occurring in one sector would still lead to downstream firms rapidly upgrading their machines, but this would nevertheless only represent a portion of the economy's overall capital stock.

Acknowledgements: The author is grateful to Andrea Roventini, Mauro Napoletano, Francesco Lamperti and Tania Treibich for providing access to the K+S C++ source code, and to the latter more specifically for recommendations regarding replications of the original policy experiments. The author also wishes to thank participants in the 2022, 2023 and 2025 ABM4Policy workshops, the 2022 workshop on data-driven economic agent-based models, the

WEHIA 2025 conference and 2025 ECONDAT conference for helpful comments and suggestions that helped shape the estimation strategy. The author also acknowledges the very helpful suggestions, particularly with respect to annual estimation provided by two anonymous referees. Finally, special thanks goes to the specialist and High Performance Computing systems provided by Information Services at the University of Kent for maintaining the computer cluster on which the Monte Carlo analysis was run. Any remaining errors are entirely down to the author.

References

Akaike, Hirotugu (1974) "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, Vol. AC-19, pp. 716–723.

Barde, Sylvain (2020) "Macroeconomic simulation comparison with a multivariate extension of the Markov information criterion," *Journal of Economic Dynamics and Control*, Vol. 111.

——— (2024) "Bayesian estimation of large-scale simulation models with Gaussian process regression surrogates," *Computational Statistics & Data Analysis*, Vol. 196, p. 107972.

Caiani, Alessandro, Antoine Godin, Eugenio Caverzasi, Mauro Gallegati, Stephen Kinsella, and Joseph E Stiglitz (2016) "Agent based-stock flow consistent macroeconomics: Towards a benchmark model," *Journal of Economic Dynamics and Control*, Vol. 69, pp. 375–408.

Christopoulou, Rebekka and Philip Vermeulen (2012) "Markups in the Euro area and the US over the period 1981–2004: a comparison of 50 sectors," *Empirical Economics*, Vol. 42, pp. 53–77.

Delli Gatti, Domenico and Jakob Grazzini (2020) "Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic Agent Based Models," *Journal of Economic Behavior & Organization*, Vol. 178, pp. 875–902.

Dosi, Giovanni, Giorgio Fagiolo, Mauro Napoletano, and Andrea Roventini (2013) "Income distribution, credit and fiscal policies in an agent-based Keynesian model," *Journal of Economic Dynamics and Control*, Vol. 37, pp. 1598–1625.

Dosi, Giovanni, Giorgio Fagiolo, Mauro Napoletano, Andrea Roventini, and Tania Treibich (2015) "Fiscal and monetary policies in complex evolving economies," *Journal of Economic Dynamics and Control*, Vol. 52, pp. 166–189.

Dosi, Giovanni, Giorgio Fagiolo, and Andrea Roventini (2006) "An evolutionary model of endogenous business cycles," *Computational Economics*, Vol. 27, pp. 3–34.

——— (2010) "Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles," *Journal of Economic Dynamics and Control*, Vol. 34, pp. 1748–1767.

Dosi, Giovanni, Mauro Napoletano, Andrea Roventini, Joseph E Stiglitz, and Tania Treibich (2020) "Rational heuristics? Expectations and behaviors in evolving economies with heterogeneous interacting agents," *Economic Inquiry*, Vol. 58, pp. 1487–1516.

Dosi, Giovanni, Mauro Napoletano, Andrea Roventini, and Tania Treibich (2017a) "Micro and macro policies in the Keynes+ Schumpeter evolutionary models," *Journal of Evolutionary Economics*, Vol. 27, pp. 63–90.

Dosi, Giovanni, Marcelo C Pereira, Andrea Roventini, and Maria Enrica Virgillito (2017b) "When more flexibility yields more fragility: the microfoundations of Keynesian aggregate unemployment," *Journal of Economic Dynamics and Control*, Vol. 81, pp. 162–186.

Erumban, Abdul Azeez (2008) "Lifetimes of machinery and equipment: evidence from Dutch manufacturing," *Review of Income and Wealth*, Vol. 54, pp. 237–268.

Fagiolo, Giorgio, Mattia Guerini, Francesco Lamperti, Alessio Moneta, and Andrea Roventini (2019) "Validation of agent-based models in economics and finance," in *Computer Simulation Validation*: Springer, pp. 763–787.

FDIC (1997) *Statistics on Banking : A Statistical History of the United States Banking Industry, Historical 1934-1996*: Federal Deposit Insurance Corporation, Division of Research and Statistics.

Ganong, Peter, Pascal Noel, and Joseph Vavra (2020) "US unemployment insurance replacement rates during the pandemic," *Journal of public economics*, Vol. 191, p. 104273.

Gilli, Manfred and Peter Winker (2003) "A global optimization heuristic for estimating agent based models," *Computational Statistics & Data Analysis*, Vol. 42, pp. 299–312.

Goddard, John, Hong Liu, Donal McKillop, and John OS Wilson (2014) "The size distribution of US banks and credit unions," *International Journal of the Economics of Business*, Vol. 21, pp. 139–156.

Grazzini, Jakob and Matteo Richiardi (2015) "Estimation of ergodic agent-based models by simulated minimum distance," *Journal of Economic Dynamics and Control*, Vol. 51, pp. 148–165.

Grazzini, Jakob, Matteo G Richiardi, and Mike Tsionas (2017) "Bayesian estimation of agent-based models," *Journal of Economic Dynamics and Control*, Vol. 77, pp. 26–47.

Grünewald, Peter D. (2007) *The Minimum Description Length Principle*: MIT Press.

Haldane, Andrew G and Arthur E Turrell (2018a) "Drawing on different disciplines: Macroeconomic agent-based models," *Journal of Evolutionary Economics*, pp. 1–28.

——— (2018b) "An interdisciplinary model for macroeconomics," *Oxford Review of Economic Policy*, Vol. 34, pp. 219–251.

Haubrich, Joseph G (2020) "A brief history of bank capital requirements in the United States," *Economic Commentary*.

Hoffman, Matthew D, Andrew Gelman et al. (2014) "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.," *J. Mach. Learn. Res.*, Vol. 15, pp. 1593–1623.

Kukacka, Jiri and Jozef Barunik (2017) "Estimation of financial agent-based models with simulated maximum likelihood," *Journal of Economic Dynamics and Control*, Vol. 85, pp. 21–45.

Lamperti, Francesco (2018) "Empirical validation of simulated models through the GSL-div: An illustrative application," *Journal of Economic Interaction and Coordination*, Vol. 13, pp. 143–171.

Lamperti, Francesco, Valentina Bosetti, Andrea Roventini, and Massimo Tavoni (2019) "The public costs of climate-induced financial instability," *Nature Climate Change*, Vol. 9, pp. 829–833.

Lamperti, Francesco, Valentina Bosetti, Andrea Roventini, Massimo Tavoni, and Tania Treibich (2021) "Three green financial policies to address climate risks," *Journal of Financial Stability*, Vol. 54, p. 100875.

Lamperti, Francesco, Giovanni Dosi, Mauro Napoletano, Andrea Roventini, and Alessandro Sapio (2018a) "Faraway, so close: Coupled climate and economic dynamics in an agent-based integrated assessment model," *Ecological Economics*, Vol. 150, pp. 315–339.

Lamperti, Francesco, Giovanni Dosi, Mauro Napoletano, Andrea Roventini, and ALESSANDRO Sapio (2020) "Climate change and green transitions in an agent-based integrated assessment model," *Technological Forecasting and Social Change*, Vol. 153, p. 119806.

Lamperti, Francesco, Andrea Roventini, and Amir Sani (2018b) "Agent-based model calibration using machine learning surrogates," *Journal of Economic Dynamics and Control*, Vol. 90, pp. 366–389.

Nakamura, Emi and Jón Steinsson (2018) "Identification in macroeconomics," *Journal of Economic Perspectives*, Vol. 32, pp. 59–86.

National Science Board (2022) *Science & Engineering Indicators 2022, Research and Development: U.S. Trends and International Comparisons.*

Poledna, Sebastian, Michael Gregor Miess, Cars Hommes, and Katrin Rabitsch (2023) "Economic forecasting with an agent-based model," *European Economic Review*, Vol. 151, p. 104306.

Smets, Frank and Rafael Wouters (2007) "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American Economic Review*, Vol. 97, pp. 586–606.

Talts, Sean, Michael Betancourt, Daniel Simpson, Aki Vehtari, and Andrew Gelman (2018) "Validating Bayesian inference algorithms with simulation-based calibration," *arXiv preprint arXiv:1804.06788.*

Wiese, Samuel, Jagoda Kaszowska-Mojsa, Joel Dyer, Jose Moran, Marco Pangallo, Francois Lafond, John Muellbauer, Anisoara Calinescu, and J Doyne Farmer (2024) "Forecasting macroeconomic dynamics using a calibrated data-driven agent-based model," *arXiv preprint arXiv:2409.18760.*

A. K+S fixed parameter calibrations

Parameter	Symbol	Value
Initialisation parameters		
Number of firms in capital-good sector	F_1	50
Number of firms in consumption-good sector	F_2	200
Number of banks	B	10
Initial equity of bank as multiple of firms total net worth	EqB_0	1
Machine output in consumption-good units per period	m_2	40
Worker output in capital-good units per period	m_1	0.1
Initial value of firm productivity	A_0	1
Initial number of workers in labor market	LS_0	250000
Average initial net worth of capital-good sector firms	$W_{1,0}$	1000
Average initial net worth of consumption-good sector firms	$W_{2,0}$	1000
Initial wage level	w_0	1
Initial firm capital stock	K_0	800
Initial mark-up of firms in the consumption-good sector	$\bar{\mu_0}$	0.2
Distributional supports		
Uniform distribution supports	$[\phi_1, \phi_2]$	$[0.1, 0.9]$
Support of the pareto distribution for number of bank clients	$[pareto_k, pareto_p]$	$[2, 200]$
Beta distribution support for the innovation draws	$[\underline{x_1}, \bar{x_1}]$	$[-0.15, 0.15]$
Bedhavioural parameters		
Interest rate on deposits	r^D	0
Bank spread mark-up on loans	μ_{deb}	0.3
Share of revenus spent on R&D by capital-good sector firms	ν	0.04
Share of R&D allocated to innovation	ξ	0.5
Firm search capabilities	ζ_1, ζ_2	0.3
New customer sample size parameter	ϖ	0.5
Payback period	b	120
Dividend rate	d	0
Credit limit in multiple of net assets or sales	ϕ_1	2
Minimum market share required to remain in consumption sector	$exit_2$	0.00001
Policy parameters		
Minimum bank capital adequacy rate	τ^b	0.08
Target interest rate for central bank	r_T	0.025
Natural unemployment rate	u^*	0.05
Unemployment benefit as a share of average wage	w_u	0.4

Table 8: Benchmark parameter values for the K+S model simulations

The following benchmark parameters from table 8 are left unchanged based on evidence from US data, as follows:

- The share of firm expenditure on R&D is left at 4% (0.04) as this is consistent with the US case, as shown in table RD10 of National Science Board (2022).
- The replacement rate of wages provided by unemployment benefits is left at 40%, based on evidence from Ganong et al. (2020) for the US in the pre-pandemic period.

- The target inflation and natural unemployment rates are left at 2% and 5% respectively. The former is based on the US Federal Reserve's mandate the latter on historical values of the US Noncyclical Rate of Unemployment (FRED series reference NROU).

- The minimum bank capital adequacy ratio is left at 8%. This stems from the Basel I requirement in 1988, covering most of the quarterly dataset, which starts in 1984. While the annual dataset starts in 1949, which predates this requirement, Haubrich (2020) points to the 1939 FDIC requirement of 10% capital to total assets, suggesting that the 8% value is broadly appropriate.

The following parameters, listed in table 2, are calibrated directly from available US data as follows:

- The mark-up of capital goods firms is increased significantly, based on evidence from Christopoulou and Vermeulen (2012).

- The population growth, which is set to 0 in the original Dosi et al. (2015) model is set using the quarterly imputed growth of the US civilian labour force over that period (FRED series reference CLF16OV).

- The tax rates are set using the 1984-2020 average of the tax revenue as a % of GDP indicator for the USA, taken from the OECD revenue statistics.

- The central bank mark-down on reserves is set to 0, as the US Federal reserve's 'Interest on Reserve Balances', very closely tracks the policy rate (effective funds rate). A potential concern is the fact that this policy was only introduced in October of 2008, prior to which the Federal Reserve did not pay interest on reserves. In order to maintain comparability of the simulations, the parameter is set to 0 across both samples.

- The inverse of bond maturity is set to 1/23, based on the "Historical Weighted Average Maturity of Marketable Debt Outstanding" of about 70 months (approximately

23 quarters) provided by the US Treasury presentations to the Treasury Borrowing Advisory Committee.[13]

- The Pareto parameter for the size distribution of banks is reduced to 0.7, based on the average value over 1995-2010 in table 2 of Goddard et al. (2014).

- The interest rate spread is set using the ICE BofA US Corporate Index Option-Adjusted Spread daily data provided by the FRED database for AAA, AA, A and BBB corporate bonds.[14] The parameter is set to the average of median spread calculated for each rating jump (i.e. AAA to AA, AA to A, etc.) between 01/01/1997 and 08/09/2022. The median is used within each rating jump to reduce the impact of outliers observed during the 2008 financial crisis.

- The technical lifetime of machines, initially set to 20 periods, i.e. 20 years, is simply multiplied by 4 to account for the shift to quarterly data. This remains in line with empirical estimates of machinery lifetime such as Erumban (2008). They report an expected average lifetime of 25.5 years for machinery and 8.6 years for computers. Given that the K+S model uses a single capital good, the 20 year lifetime parameter provides a reasonable middle ground.

A-A	Anchor and Adjust expectations mechanism
ABM	Agent-based model
BEGRS	Bayesian estimation with gaussian regression surrogates
DSK	Dystopian Schumpeter-Keynes model (variant of K+S, below)
ESS	Effective sample size
FRED	Federal Reserve Economic Data
FDIC	Federal Deposit Insurance Corporation
GP	Gaussian process
K+S	“Keynes meets Schumpete” model Dosi et al. (2006)
MIC	Markov information criterion
NUTS	No U-turn Sampler
SBC	Simulation-based calibration

Table 9: List of acronyms used

[13]https://home.treasury.gov/system/files/221/TreasuryPresentationToTBACQ42024.pdf

[14]Specifically, FRED series references BAMLC0A1CAAA, BAMLC0A2CAA, BAMLC0A3CA and BAMLC0A4CBBB

B. Supplementary policy results

	Annual estimates			Quarterly estimates		
	TR_π	$TR_{\pi,U}$	$spread$	TR_π	$TR_{\pi,U}$	$spread$
Mean GDP growth						
$No\ rule$	1.000	0.854***	0.851***	1.000	0.664***	0.668***
	(0.000)	(33.076)	(32.694)	(0.000)	(23.275)	(23.087)
SGP	1.000	0.947***	0.944***	1.204***	1.000	0.983
	(0.067)	(11.685)	(12.026)	(16.543)	(0.036)	(1.208)
FC	0.938***	−0.376***	−0.344***	1.065***	−0.243***	−0.370***
	(8.196)	(36.712)	(37.839)	(2.914)	(27.918)	(30.154)
SGP_{ec}	1.003	0.950***	0.943***	1.200***	0.938***	0.926***
	(1.280)	(11.826)	(12.920)	(16.182)	(4.568)	(5.344)
FC_{ec}	0.908***	−0.383***	−0.405***	1.083***	−0.340***	−0.288***
	(8.021)	(42.076)	(41.614)	(3.785)	(31.064)	(30.010)
Mean GDP growth volatility						
$No\ rule$	1.000	0.675***	0.682***	1.000	0.810***	0.811***
	(0.000)	(27.888)	(26.870)	(0.000)	(15.747)	(15.955)
SGP	1.609***	5.292***	5.482***	2.513***	4.241***	4.388***
	(27.425)	(22.655)	(21.495)	(27.546)	(31.362)	(29.793)
FC	2.460***	26.235***	24.560***	6.158***	30.793***	32.229***
	(10.368)	(29.657)	(29.482)	(22.791)	(56.751)	(61.319)
SGP_{ec}	1.548***	4.719***	4.865***	2.771***	3.426***	3.478***
	(30.359)	(19.918)	(20.236)	(30.957)	(36.048)	(34.563)
FC_{ec}	2.787***	45.813***	46.046***	6.427***	35.406***	35.341***
	(9.007)	(33.713)	(35.210)	(23.373)	(58.517)	(59.380)
Mean unemployment rate						
$No\ rule$	1.000	2.984***	2.986***	1.000	1.481***	1.485***
	(0.000)	(36.773)	(36.470)	(0.000)	(22.893)	(22.756)
SGP	0.994	2.204***	2.237***	0.966**	1.399***	1.424***
	(0.355)	(26.233)	(26.781)	(2.129)	(17.319)	(17.628)
FC	1.872***	4.665***	4.667***	1.569***	3.970***	4.012***
	(19.181)	(50.485)	(50.545)	(18.743)	(135.316)	(150.705)
SGP_{ec}	1.001	2.345***	2.390***	1.002	1.516***	1.534***
	(0.040)	(30.729)	(32.863)	(0.125)	(23.173)	(23.897)
FC_{ec}	1.930***	4.674***	4.809***	1.621***	3.956***	3.963***
	(19.800)	(50.997)	(54.191)	(19.882)	(139.532)	(139.596)
Mean inflation rate						
$No\ rule$	1.000	1.041***	1.042***	1.000	0.893***	0.896***
	(0.000)	(13.181)	(13.392)	(0.000)	(19.359)	(18.632)
SGP	1.014***	1.041***	1.041***	1.087***	0.990	0.985*
	(8.189)	(12.263)	(12.073)	(15.873)	(1.213)	(1.746)
FC	1.016***	0.890***	0.900***	1.015**	0.380***	0.375***
	(8.778)	(11.717)	(11.295)	(2.527)	(43.492)	(44.420)
SGP_{ec}	1.011***	1.027***	1.024***	1.075***	0.965***	0.963***
	(6.477)	(7.918)	(7.909)	(13.854)	(5.409)	(5.331)
FC_{ec}	1.013***	0.830***	0.842***	1.006	0.348***	0.378***
	(6.737)	(13.998)	(14.970)	(0.983)	(43.426)	(43.943)

Note: For each variable and policy combination, the entries provide the mean value over the replication relative to the mean value of the $No\ rule, TR_\pi$ policy combination. T-statistics for the difference in means are provided in parentheses, with *, ** and *** indicating significance at the 10%, 5% and 1% levels respectively.

Table 10: Policy experiments for baseline expectation model