An open OBR macro model for UK fiscal scoring*

Vahid Ahmadi†

July 2026

Abstract

I describe an open-source Python implementation of the Office for Budget Responsibility's macroeconometric model of the UK economy, built for near-term fiscal-reform scoring in the PolicyEngine Macro suite. The emulator transpiles the OBR's 372 published EViews equations into Python, solves them by Gauss–Seidel iteration with a demand-side closure swap, and anchors the baseline to the OBR's November 2025 *Economic and Fiscal Outlook* using the OBR's own add-factor device. In anchored mode the model reproduces the published EFO path with a GDP MAPE of 0.18 per cent and consumption of 0.29 per cent over 2025Q1–2027Q4, hard-gated in continuous integration. The free-running fit is deliberately reported without flattery: of 21 headline variables, 10 are passthroughs and only 4 of the 11 actually computed land within band, with company profits 54.6 per cent off—a gap traced to unpublished OBR calibration constants rather than live bugs. A static-costing bridge injects PolicyEngine microsimulation costings into the model as a household-disposable-income shock, delivering the OBR's own "static costing in, second-round effects out" workflow. A worked example—raising the basic rate of income tax by 1p—yields a static costing of £6.46 billion in 2026 rising to £7.38 billion by 2030, within the range of HMRC's published ready reckoner, with second-round GDP effects building from −0.02 per cent on impact to −0.06 per cent by end-2027. Three further experiments—a £5bn government-consumption expansion, a VAT-style one per cent consumer-price shock, and a 2pp corporation-tax rise through the user-cost-of-capital channel—exercise the direct macro levers and are benchmarked against the official ready-reckoner numbers where they exist.

Current-vintage note (21 July 2026). The deployed emulator and current validation tables are anchored to the March 2026 EFO. Over 2025Q1–2027Q4 the anchored GDP MAPE is 0.15 per cent and consumption MAPE is 0.25 per cent; the free-running values are 5.75 and 9.56 per cent. References below to November 2025 describe the preserved historical study and must not be read as the live baseline.

Keywords: macroeconometric models, fiscal policy, policy costings, OBR, microsimulation, open source

*Acknowledgements: I thank Max Ghenis for his great suggestions and helpful discussions. This paper is a PolicyEngine working paper describing the OBR macroeconometric emulator in the PolicyEngine Macro suite. It was prepared with AI assistance (Claude, Anthropic). All remaining errors are my own.

†Research Associate at PolicyEngine. Email: vahid@policyengine.org.

1 Introduction

When the UK government changes a tax or benefit, two questions follow immediately. The first is static: who pays what the morning after? The second is dynamic: how do consumption, investment, incomes, and prices adjust over the following years, and what does that adjustment do to the revenue estimate? The Office for Budget Responsibility (OBR) answers the second question with a large-scale macroeconometric model in the Treasury/Cowles-Commission tradition—a model whose equation listing and variable definitions are published, but which is operated in proprietary EViews software, on a databank the OBR does not fully publish, and with judgement imposed through unpublished residual adjustments (Office for Budget Responsibility, 2013, 2025b).

This paper describes an open-source Python emulator of that model, built as the near-term fiscal-scoring member of the *PolicyEngine Macro* suite.[1] The emulator takes the OBR's published model code of 15 October 2025—372 equations—transpiles it mechanically into Python, initialises it from the November 2025 *Economic and Fiscal Outlook* (EFO) detailed forecast tables (Office for Budget Responsibility, 2025a) and a 348-series snapshot of Office for National Statistics data, and solves it by Gauss–Seidel iteration. Policy analysis proceeds the way the OBR's own does: the baseline is anchored to the published forecast through add-factors, a shock is applied to an exogenous lever (or injected through a microsimulation bridge), and the reform effect is read as the deviation between the shocked and baseline solutions.

The contribution is threefold. First, *replication as infrastructure*: to my knowledge this is the first publicly runnable implementation of the OBR's macroeconomic model. The OBR publishes the equation listing and a briefing paper describing the model (Office for Budget Responsibility, 2013) but no official working paper accompanies the model itself, and the model cannot be run without EViews and the OBR's internal databank. The emulator makes the official forecaster's transmission mechanisms—the marginal propensities, error-correction speeds, and accounting identities that shape UK fiscal scoring—inspectable and executable by anyone with Python.

Second, *integration with microsimulation*: a static-costing bridge connects PolicyEngine's UK tax–benefit microsimulation (PolicyEngine, 2025) to the macro model, reproducing the OBR's own policy-costing workflow (Office for Budget Responsibility, 2014) in which a static costing from the revenue departments enters the macroeconomic forecast and second-round demand effects come out. A reform expressed as PolicyEngine parameters is costed at household resolution, converted to a shock on nominal household disposable income, and propagated through the model's consumption and expenditure blocks.

Third, *honest validation*: the project distinguishes sharply between what the model reproduces by construction and what it genuinely forecasts. The anchored baseline matches the EFO to within 0.2–0.3 per cent by design and this invariant is hard-gated in continuous integration; the free-running model, scored without the add-factors and de-seeded from the EFO path, is documented as weak—only 4 of 11 computed headline variables land within band—with each residual error traced either to a fixed bug or to a documented "floor" imposed by constants the OBR does not publish. Section 9 places the model on the suite's *verification gradient*, which grades each model class by how far its outputs can be audited against ground truth.

The rest of the paper proceeds as follows. Section 2 places the project in the UK macroecono-

[1] `https://policyengine-macro.vercel.app/obr/`; source at `https://github.com/PolicyEngine/obr-macroeconomic-model`. PolicyEngine Macro is a suite of open macroeconomic simulation models—overlapping-generations, macroeconometric, and structural-VAR classes—that score the same PolicyEngine reform objects and report comparable real-world aggregates (PolicyEngine, 2025).

metric modelling tradition and the replication literature. Section 3 sets out the model's equation structure block by block—the ECM template, the expenditure core, consumption, investment, the labour market, trade, government spending, incomes, prices, and the full receipts-to-borrowing fiscal chain—together with the add-factor architecture and the anchoring mechanism. Section 4 describes the implementation: transpiler, solver, closure swap, microsimulation bridge, and hosted interfaces. Section 5 presents calibration and validation against the November 2025 EFO. Section 6 reports a worked reform—1p on the basic rate of income tax—and compares it with HMRC's published ready reckoner (HM Revenue and Customs, 2025); Section 7 adds three further experiments through the direct macro levers (a government-consumption expansion, a VAT-style price shock, and a corporation-tax rise). Section 8 consolidates every comparison with official results in one place and extends it beyond the anchor vintage—to the OBR's March 2026 EFO, to ONS outturn data published since anchoring, to the fiscal-multiplier literature, and to the certified policy-costings record. Section 9 discusses limitations, and Section 10 concludes. Appendices provide a variable glossary, the complete free-running scorecard, and solver details.

2 Related literature

The emulator sits at the intersection of three literatures: the UK tradition of large-scale macroeconometric policy models, the post-crisis debate over what class of model official institutions should operate, and the replication and open-science movement in economics.

2.1 The UK macroeconometric modelling tradition

The model being emulated is a direct descendant of the estimated structural models pioneered by Klein (1950) at the Cowles Commission: a system of behavioural equations estimated on national-accounts data and closed by accounting identities, solved numerically to produce forecasts and policy simulations. The Treasury built the first version of its UK model in 1970, and the model passed into joint OBR/Treasury ownership when the OBR was created in 2010, with governance set out in a published memorandum of understanding (Office for Budget Responsibility and HM Treasury, 2011; Office for Budget Responsibility, 2013). Through the 1970s and 1980s the UK sustained an unusually rich ecosystem of competing models of this class—the Treasury, the National Institute of Economic and Social Research (NIESR), the London Business School, the Bank of England, and the Liverpool model among them—subjected to systematic comparative evaluation by the ESRC Macroeconomic Modelling Bureau at Warwick (Wallis et al., 1987). The Bureau's reviews established a discipline that this paper consciously imitates at small scale: deposit the model, document its properties, decompose its forecasts into model and judgement, and compare simulation responses across models on common shocks. That comparative infrastructure largely disappeared in the 1990s; one reading of the present project is an attempt to rebuild a small piece of it in open source, for the one UK model whose equation listing is still published.

The OBR/Treasury model's closest living relatives are NIESR's NiGEM, a quarterly estimated global model with country blocks linked by trade and financial flows, maintained commercially and used by central banks and finance ministries for scenario analysis (Hantzsche et al., 2018), and—across the Atlantic and in a more explicitly optimising idiom—the Federal Reserve's FRB/US (Brayton and Tinsley, 1996), whose public release of equations, data, and simulation code is the transparency standard this project aims at for the UK. The Bank of England, by contrast, moved its central forecasting platform to an estimated DSGE core (COMPASS) surrounded by a suite of auxiliary models (Burgess et al., 2013); the OBR/Treasury model is thus the last large non-DSGE structural econometric model in operational use by a UK macro-fiscal institution.

2.2 Which model class for policy institutions?

The choice embodied in the OBR model—error-correction econometrics over microfounded optimisation—was contested territory well before the financial crisis and became more so after it. Wren-Lewis (2018) argues that the "microfoundations hegemony" crowded out precisely the class of structural econometric models (SEMs) to which the OBR model belongs, and that policy analysis would have been better served had SEMs retained equal academic standing. Blanchard (2018), taxonomising models into foundational, DSGE, policy, toy, and forecasting classes, places institutional models of the OBR type in the "policy" class, where fitting the data and capturing actual transmission mechanisms trump theoretical purity. Hendry and Muellbauer (2018) make the econometric case directly: for forecasting and policy in the presence of structural breaks, congruent equilibrium-correction systems—exactly the Davidson et al. (1978)/Engle and Granger (1987) architecture of Section 3.1—dominate models whose cross-equation restrictions are derived from misspecified theory. This paper takes no side in that debate; its premise is narrower: whatever the merits of the model class, the model that actually produces the UK's official fiscal forecast should be publicly runnable. The multi-model design of the PolicyEngine Macro suite—macroeconometric, overlapping-generations, and structural-VAR members scoring the same reform—is itself an implementation of the "suite of models" philosophy that Burgess et al. (2013) describe at the Bank of England.

2.3 Replication, emulation, and open models

Finally, the project draws method and motivation from the replication literature. Systematic replication attempts find that even published academic results frequently cannot be reproduced from the authors' own materials (Chang and Li, 2015), and the best-known failure in the fiscal-policy domain—the spreadsheet error in Reinhart and Rogoff's debt-threshold result, found only when a graduate student obtained and re-ran the original file (Herndon et al., 2014)—is a standing argument that consequential numbers should be recomputable by outsiders. Official fiscal scoring is a stronger case still: OBR forecasts condition UK budget decisions directly, yet the model that produces them has never been publicly executable. The emulator is best understood as a *replication-as-infrastructure* exercise: not a one-off audit of a published result, but a maintained, continuously tested re-implementation whose agreement with the official output is machine-checked on every commit (Section 5), in the same spirit in which PolicyEngine re-implements statutory tax–benefit rules as open code (PolicyEngine, 2025). The honest-scorecard discipline of Section 5.2—publishing the free-running failures alongside the anchored successes, and refusing to tune unpublished constants to hit known answers—is the transparency analogue of pre-registration: it separates what the replication demonstrates from what it merely reproduces.

3 The model

The OBR's macroeconomic model descends from the Treasury model first built in 1970 and has been jointly maintained by the OBR and HM Treasury since 2010 (Office for Budget Responsibility, 2013; Office for Budget Responsibility and HM Treasury, 2011). It is a quarterly, large-scale macroeconometric model in the estimated-structural tradition of Klein (1950) and Fair (2004): a small number of estimated behavioural equations embedded in a dense web of national-accounts identities, with expectations largely implicit and adjustment dynamics carried by error-correction terms (Davidson et al., 1978; Engle and Granger, 1987). It sits in the lineage of the large UK policy models reviewed by Wallis et al. (1987), and is the UK institutional counterpart of semi-structural policy models such as the Federal Reserve's FRB/US (Brayton and Tinsley, 1996). The 15 October 2025 model listing used here contains 372 live equations over 636 variable codes. Of the 636 codes, 253 are exogenous inputs—the fiscal levers, market assumptions, world demand, and demographic paths on which the OBR conditions its forecast—and 383 are endogenous, solved by the model; of those endogenous variables, roughly 29 are determined by econometrically estimated behavioural equations, 216 by accounting identities, and 138 by calibrated or technical relationships (the count of live equations, 372, is slightly smaller because a few listed relationships ship commented out or otherwise inactive in the published file). All equations quoted in this section are transcribed verbatim (coefficients included, rounded to four decimals) from the published listing; Appendix A defines every mnemonic used.

3.1 The anatomy of an OBR equation: ECM long run and short run

Nearly every estimated behavioural equation in the listing has the same equilibrium-correction form, and reading the model is easier once that template is fixed. A representative equation for a variable y driven by long-run determinants x and short-run drivers z is

$$\Delta \ln y_t \;=\; \underbrace{\alpha + \textstyle\sum_i \beta_i \, \Delta \ln z_{i,t}}_{\text{short-run dynamics}} - \lambda \underbrace{\left[\ln y_{t-1} - \textstyle\sum_j \gamma_j \ln x_{j,t-1}\right]}_{\text{lagged equilibrium error}}, \qquad 0 < \lambda < 1. \tag{1}$$

The bracketed term is the residual of a long-run (cointegrating) relationship $\ln y^* = \sum_j \gamma_j \ln x_j$ up to a constant: when y sat above its equilibrium value last quarter, growth is pulled down by λ times the gap, so the level converges to y^* at a geometric rate—a gap half-life of roughly $\ln 2/\lambda$ quarters. The β_i terms govern the impact response before the level correction has had time to bite. Three consequences matter for policy scoring. First, *impact and long-run multipliers differ*: a permanent income shock moves consumption on impact by the short-run elasticity (0.103 in equation (3)) but eventually by the long-run one (0.439), so reform effects build over the horizon rather than arriving at once—exactly the profile visible in Figure 3. Second, the adjustment speed λ is the model's memory: at $\lambda = 0.125$ (consumption) roughly 12.5 per cent of any disequilibrium closes per quarter, so a two-year scoring window captures only part of the eventual response. Third, the intercept α and the implicit constant of the long-run relation are where unpublished calibration lives; as Section 5 shows, it is precisely these constants (not the published slopes) that impose floors on the free-running fit. The OBR's add-factors (Section 3.11) enter equation (1) as period-specific shifts to α.

3.2 The expenditure core

Real GDP at market prices ($GDPM$) is tied to its expenditure components by the identity

$$GDPM = CGG + CONS + IF + DINV + VAL + X - M + SDE, \tag{2}$$

where CGG is government consumption, $CONS$ private consumption, IF total fixed investment, $DINV$ inventory change, VAL valuables, X and M exports and imports, and SDE the statistical discrepancy. In the published listing this identity is arranged with inventories as the residual; the emulator's closure swap (Section 4) re-arranges it so that output responds to demand shocks.

3.3 Consumption

The centrepiece behavioural equation is an error-correction consumption function in the tradition of Davidson et al. (1978). As published (coefficients verbatim from the OBR listing):

$$\begin{aligned}\Delta \ln CONS = {} & 0.2646 + 0.1030\,\Delta \ln RHHDI - 0.0084\,\Delta LFSUR \\ & + 0.1269\,\Delta \ln\left(\frac{1000\,GPW}{PCE/100}\right) - 0.0004\,\Delta\left[R_{-1} - \pi_4^{PCE}\right] \\ & - 0.1251\left[\ln CONS_{-1} - 0.4393 \ln RHHDI_{-1}\right. \\ & \left. - 0.1059 \ln\left(\tfrac{1000\,GPW_{-1}}{PCE_{-1}/100}\right) - 0.2216 \ln\left(\tfrac{NFWPE_{-1}}{PCE_{-1}/100}\right)\right],\end{aligned} \tag{3}$$

where $RHHDI$ is real household disposable income, $LFSUR$ the LFS unemployment rate, GPW gross housing wealth, $NFWPE$ net financial wealth of the personal sector, PCE the consumption deflator, R Bank Rate, and π_4^{PCE} annual consumption-deflator inflation. Mapped onto template (1): the long-run solution ties consumption to income (elasticity 0.439), housing wealth (0.106), and financial wealth (0.222) with adjustment speed $\lambda = 0.125$ per quarter—a half-life for a consumption gap of about five and a half quarters; the short run responds to income growth, unemployment changes, housing-wealth growth, and the change in the ex-post real interest rate. A companion equation splits off durables ($CDUR$) with relative-price, income, and user-cost terms, and a house-price equation (PD) ties prices to housing wealth per house and the earnings–deflator wedge, closing a housing-collateral loop back into equation (3) through GPW.

3.4 Investment

Total investment is built from its components,

$$IBUS = IF - GGI - PCIH - PCLEB - IH - IPRL, \tag{4}$$

with business investment behaviour specified as an error-correction accelerator on market-sector output with a user-cost channel:

$$\begin{aligned}\Delta \ln IBUSX = {} & 0.1992\,\Delta \ln IBUSX_{-3} + 1.0057\,\Delta \ln MSGVA_{-1} - 0.0012\,CBIUD \\ & - 0.0418\left[\ln IBUSX_{-1} - \ln(1000\,KMSXH_{-2}) + KGAP_{-2} - 0.0884\right],\end{aligned} \tag{5}$$

where $MSGVA$ is market-sector gross value added, $CBIUD$ a CBI capacity-utilisation survey balance, $KMSXH$ the market-sector capital stock, and $KGAP$ the gap between actual and desired capital—the term through which the corporation-tax rate ($TCPRO$), by raising the user cost of capital, lowers desired capital and hence investment. Note the near-unit short-run accelerator coefficient on lagged market-sector output growth (1.0057) against the slow error correction ($\lambda = 0.042$, a gap half-life of about sixteen quarters): investment chases demand hard in the short run and repairs its level only slowly, which is why an unconstrained output–investment loop is explosive on a databank without a live supply block (Section 4). In the October 2025 listing this behavioural line ships commented out (the OBR runs investment off its databank with judgement); the emulator's corporation-tax experiments therefore operate the channel through an explicit investment closure with the cost-of-capital relationship restored and stabilised.

3.5 Labour market

The labour-market block combines one estimated equation—market-sector employment—with a set of identities that convert employment into unemployment and participation. Market-sector employees (EMS) follow a labour-demand ECM on market-sector output and the real producer wage:

$$\begin{aligned}\Delta \ln EMS \;=\; & -0.0113 \;+\; 0.4370\,\Delta \ln EMS_{-1} \;+\; 0.1932\,\Delta \ln EMS_{-2} \\ & +\; 0.1714\,\Delta \ln MSGVA_{-1} \;-\; 0.0062\left[\ln\!\left(\tfrac{EMS_{-1}}{MSGVA_{-1}}\right) + 0.4\ln\!\left(\tfrac{PSAVEI_{-1}}{PMSGVA_{-1}}\right)\right],\end{aligned} \tag{6}$$

(a 2010Q4 break dummy omitted for readability). The long-run solution makes the employment–output ratio depend on the producer real wage with elasticity 0.4: labour demand is downward-sloping, but the error-correction speed is extremely slow ($\lambda = 0.0062$, a half-life of over a hundred quarters), so employment is dominated in any scoring window by the lagged-adjustment and output-growth terms—the model's rendering of labour hoarding. Around this single behavioural equation sit the identities

$$ETLFS = 1000\,\frac{HWA}{AVH}, \qquad ULFS = \frac{POP16 \cdot PART16}{100} - ETLFS, \tag{7}$$

$$LFSUR = 100\,\frac{ULFS}{ETLFS + ULFS}, \qquad ER = 100\,\frac{ETLFS}{POP16}, \tag{8}$$

which build LFS employment from total hours worked (HWA) over average hours (AVH), unemployment from the participation-implied labour force less employment, and the unemployment and employment rates from those stocks. Workforce jobs, self-employment, and public-sector employment (ECG, ELA) scale by ratio links to the government-sector aggregates. The crucial configuration fact—central to the honest scorecard of Section 5.2—is that HWA, AVH, and $PART16$ are exogenous inputs in this configuration: total employment is a *trivial identity* over passthrough series, and only EMS and the unemployment split are genuinely computed. Unemployment then feeds back into the demand side twice: through the $\Delta LFSUR$ term in consumption (3) and through the wage equation below.

Private-sector average earnings ($PSAVEI$) are set by an estimated wage ECM:

$$\begin{aligned}\Delta \ln PSAVEI \;=\; & -0.0282 \;+\; 0.575\,\Delta \ln PMSGVA + 0.250\,\Delta \ln PMSGVA_{-1} \\ & +\; 0.105\,\Delta \ln PMSGVA_{-2} \;+\; 0.070\,\Delta \ln PMSGVA_{-3} \;-\; 0.0096\,\Delta LFSUR \\ & +\; 0.264\,\big(\Delta \ln MSGVA - \Delta \ln EMS\big) \\ & +\; 0.282\,\big(\Delta \ln CPI - \Delta \ln PMSGVA\big) \\ & -\; 0.0433\,\Big[\ln PSAVEI_{-1} - \ln\big(\tfrac{MSGVA_{-1}}{EMS_{-1}}\big) - \ln PMSGVA_{-1} \\ & \qquad +\; \ln\big(1 + \tfrac{EMPSC_{-1}}{WFP_{-1}}\big) + 0.0137\,LFSUR_{-1}\Big],\end{aligned} \tag{9}$$

a bargaining-style wage equation: the long run ties the real product wage to labour productivity ($MSGVA/EMS$) one-for-one, shifted down by employer social contributions (the tax wedge, $EMPSC/WFP$) and by unemployment (a wage curve with semi-elasticity 0.0137); the short run passes through producer-price inflation (with distributed-lag weights summing to one, so there is no long-run money illusion), productivity growth, the consumer–producer price wedge, and the change in unemployment. Earnings, employment, and public-sector pay then aggregate to the economy-wide wage bill

$$\begin{aligned}WFP \;=\; & ADJW \cdot PSAVEI \cdot (EMS - ESLFS) \\ & + \tfrac{52}{4000}\, CGWADJ \cdot ERCG \cdot ECG + \tfrac{52}{4000}\, LAWADJ \cdot ERLA \cdot ELA,\end{aligned} \tag{10}$$

which is the largest single input to household income (15) and to income-tax receipts in the fiscal block: equations (6)–(10) are the transmission path by which a demand shock eventually moves PAYE receipts.

3.6 Trade

The trade block is asymmetric in this configuration: aggregate export and import volumes (X, M) are exogenous conditioning inputs (the world-demand judgement), while the block's live equations decompose them, price them, and determine the import *content* of demand. Non-oil goods flows are residuals, $XNOG = X - XS - XOIL$ and $MNOG = M - MS - MOIL$, and current-price aggregates are built by deflator-weighted sums such as $XPS = (PXNOG/100)XNOG + (PXS/100)XS + (PXOIL/100)XOIL$. The import content of final expenditure is imposed through fixed input–output weights:

$$MGTFE = 0.176\,CONS + 0.064\,CGG + 0.175\,IF + 0.094\,DINV + 0.410\,XNOG + 0.049\,XS, \tag{11}$$

with a services analogue ($MSTFE$; weights 0.081, 0.030, 0.059, 0.012, 0.029, 0.093): a pound of goods exports carries far more import content (0.41) than a pound of government consumption (0.064), one reason fiscal multipliers on CGG exceed those on trade-intensive demand. Services import volumes are the block's estimated behavioural equation, an ECM on weighted expenditure and the relative import price:

$$\begin{aligned}\Delta \ln MS \;=\; & 0.819\,\Delta \ln MSTFE + \cdots - 0.477\,\Delta \ln PMSREL - 0.271\,\Delta \ln MS_{-1} \\ & -\; 0.171\,\big[\ln MS_{-1} - 1.079 \ln MSTFE_{-1} - 0.662 \ln SPECX_{-1} \\ & \qquad +\; 0.874 \ln PMSREL_{-1} + \cdots\big] - 0.032,\end{aligned} \tag{12}$$

whose long run has an expenditure elasticity slightly above one (trade deepening) and a relative-price elasticity of −0.87. Trade *prices* are estimated ECMs in domestic costs and world prices converted at the exchange rate—for export prices,

$$
\begin{aligned}
\Delta \ln PXNOG \; &= 0.636\,\Delta \ln PPIY_{-1} + 0.103\left(\Delta \ln WPG - \Delta \ln RXD\right) - 0.131\,\Delta \ln RX \\
&\quad - \; 0.161\left[\ln PXNOG_{-1} + 0.330 \ln RX_{-1} - 0.921 \ln PPIY_{-1}\right. \\
&\quad\quad - \; \left. 0.079 \ln(WPG_{-1}/RXD_{-1})\right] + \cdots ,
\end{aligned}
\tag{13}
$$

with an import-price twin ($PMNOG$; long-run weights 0.55 domestic, 0.45 world). The long-run weights say UK exporters are mostly domestic-cost setters (0.92 on producer prices $PPIY$) with modest world-price-taking (0.08 on world goods prices WPG/RXD), and both equations make the exchange rate (RX, RXD) a first-order transmission channel: a depreciation raises export and import prices within quarters, feeding the CPI–RPI block and, through $PMSREL$ in (12), expenditure switching. Oil trade prices are pure conversions of the exogenous Brent price, $\Delta \ln PXOIL = \Delta \ln PBRENT - \Delta \ln RXD$.

3.7 Government spending

Government consumption is the model's canonical fiscal lever, and its treatment illustrates the model's nominal-to-real architecture. The Spending-Review path enters in cash terms ($CGGPS$, nominal government consumption); real government consumption follows a bridging equation

$$
\Delta \ln CGG \;=\; 0.0007 \;+\; 0.3739\,\Delta \ln CGGPS \;+\; 0.1802\,\Delta \ln CGGPS_{-1} \;-\; 0.4198\,\Delta \ln CGG_{-1}, \tag{14}
$$

a distributed lag converting nominal spending plans into real activity, with the implied deflator for government consumption defined residually as $GGFCD = 100\,CGGPS/CGG$. Procurement is the cash residual $CGP = CGGPSPSF - (CGWS + LAWS) - LAPR - (RCGIM + RLAIM)$: central-government procurement is what remains of the spending envelope after central and local pay bills ($CGWS$, $LAWS$), local procurement, and capital-consumption imputations. The split matters for incidence: the pay component feeds the wage bill (10) and household income directly, while procurement enters demand without a first-round household-income counterpart. Government-sector social contributions scale with the pay bills by ratio links (e.g. $CGASC$ grows in line with $CGWS$; similarly $EESCCG$, $LASC$, $EESCLA$), and government investment enters the capital account as nominal $CGIPS$ (a direct lever in the emulator) deflated onto real GGI inside total investment. In the emulator's demand closure a shock to real CGG enters identity (2) directly—hence the impact multiplier of approximately one reported in Section 7—while its cash counterpart flows through the fiscal chain of Section 3.10 into borrowing.

3.8 Incomes

Household disposable income is a pure accounting identity over the sector accounts:

$$
\begin{aligned}
HHDI \; &= MI + FYEMP - EMPSC - EESC - TYWHH + NMTRHH + SBHH \\
&\quad + (PIRHH - PIPHH + FSMADJ) - HHSB + HHISC \\
&\quad + (EECOMPC - EECOMPD) + OSHH,
\end{aligned}
\tag{15}
$$

summing mixed income, employee compensation less social contributions, taxes on income ($TYWHH$), net transfers and benefits, net property income, and households' operating surplus

($OSHH$, mostly imputed rent). Real income deflates by the consumption deflator, $RHHDI = 100\,HHDI/PCE$, and feeds equation (3). This chain—$HHDI \rightarrow RHHDI \rightarrow CONS \rightarrow GDPM$—is the transmission path used by the microsimulation bridge. On the other side of the factor-income split, company profits are the residual of the operating-surplus account:

$$FYCPR \;=\; OS - OSHH - OSGG - OSPC - RENTCO + SA - FISIMPS. \tag{16}$$

The wage bill is equation (10); the taxes-on-income term closes the loop back to the fiscal block through

$$TYWHH = TYEM + TSEOP + CC + CGT + OCT - NPISHTC, \tag{17}$$

PAYE and self-assessment income tax, council tax, capital gains tax, and other current taxes: the same receipts lines that aggregate into government revenue in Section 3.10 are subtracted here from household income—precisely the double-entry the microsimulation bridge declines to re-traverse (Section 4.1).

3.9 Prices

The retail-price block builds the RPI from mortgage-interest and other items, $PR = I7\,[(1 - W_4)\,PRXMIP/I9 + W_4\,PRMIP/I4]$, with annual RPI inflation $RPI = 100\,(PR/PR_{-4} - 1)$; the consumption deflator moves with the CPI through $PCE/PCE_{-4} = CPI/CPI_{-4}$. The CPI itself is exogenous in this configuration (the OBR imposes its inflation judgement), so the deflator path is conditioned on the EFO inflation forecast—which also makes an exogenous CPI shock a legitimate lever for price-shock experiments (Section 7). Rent components ($PRENT$, $CPIRENT$) index to a weighted average of earnings growth, house-rental values, and house prices, tying the price block back to the labour-market and housing equations.

3.10 The fiscal block: receipts, spending, and the PSNB chain

The fiscal block is the deepest identity chain in the model: roughly a hundred equations that aggregate individual taxes and spending lines into the headline public-finance aggregates. It is worth setting out in full, because it is the path by which any scored reform becomes a borrowing number.

Receipts. Individual tax streams aggregate bottom-up. Income-tax receipts are

$$INCTAXG = TYEM + TSEOP + TCINV - INCTAC + CTC - NPISHTC, \tag{18}$$

PAYE plus self-assessment plus tax on investment income, less accruals adjustments and reliefs. Indirect taxes bundle as $CETAX = VREC + TXFUEL + TXTOB + TXALC + CUST + CCL + AL + TXCUS$ (VAT, fuel, tobacco, alcohol, customs, climate-change levy, aggregates levy, customs adjustments). These roll up into taxes on income and wealth,

$$\begin{aligned} PUBSTIW \;&= TYEM + TSEOP + PRT + TCINV + CT + CGT + FCACA \\ &\quad + BETPRF + BETLEVY + OFGEM - NPISHTC \\ &\quad - TYPCO + PROV - LAEPS, \end{aligned} \tag{19}$$

(including corporation tax $CT = NSCTP + NNSCTP$, split North Sea/onshore), and finally into public sector current receipts:

$$
\begin{aligned}
PSCR &= PUBSTIW + PUBSTPD + OCT + CC + INHT + EENIC + EMPNIC \\
&\quad + (RCGIM + RLAIM + OSPC) + PSINTR + (RNCG + HHTCG) \\
&\quad + LARENT + PCRENT + \cdots ,
\end{aligned} \tag{20}
$$

taxes on income and wealth, taxes on production, council tax, inheritance tax, employee and employer National Insurance, imputed capital consumption, public-sector interest and dividend receipts, and rent.

Spending. Public sector current expenditure ($PSCE$) is the mirror aggregate—pay, procurement, subsidies, net social benefits ($CGSB + LASBHH$), current grants, debt interest, and EU transactions—carried in the listing in inverted form (subsidies as the residual, reflecting the OBR's own solution ordering). Depreciation is $DEP = RCGIM + RLAIM + PCCON$, and gross public-sector investment accumulates as

$$
\begin{aligned}
PSGI &= CGIPS + LAIPS + IPCPS + IBPC + DINVCG \\
&\quad + (NPACG + NPALA) + \cdots + ASSETSA,
\end{aligned} \tag{21}
$$

central and local government investment, public-corporation investment, and asset transactions.

The borrowing chain. The headline aggregates then follow in strict sequence:

$$
\begin{aligned}
PSCB &= PSCR - PSCE - DEP && \text{(current balance)} && (22)\\
PSNI &= PSGI - DEP - ASSETSA && \text{(net investment)} && (23)\\
PSNB &= -PSCB + PSNI && \text{(net borrowing)} && (24)\\
PSNCR &= PSNB + PSLSFA + PSACADJ && \text{(net cash requirement)} && (25)\\
\Delta PSND &= PSNCR - ILGAC + \Delta FLEASGG && && \\
&\quad + \Delta FLEASPC + PSNDRES && \text{(net debt)} && (26)
\end{aligned}
$$

Current receipts less current spending less depreciation give the current balance; the deficit is the current deficit plus net investment; cash-basis and accruals adjustments ($PSLSFA$, $PSACADJ$) convert borrowing to the cash requirement; and the cash requirement, less index-linked gilt accruals and finance-lease adjustments, cumulates into the net-debt stock. Parallel decompositions attribute borrowing to sub-sectors ($CGNB$, $LANB$, $PCNB$ for central government, local authorities, and public corporations, with $GGNB = CGNB + LANB$), and the sub-sector chain feeds general-government gross debt ($GGGD$) for the international definitions. The one add-factor in this chain sits on $PSNBCY$ (the listing's `@ADD` on calendar-year borrowing), which is where the OBR's fiscal judgement enters. Every behavioural response elsewhere in the model reaches the public finances through this chain: a consumption shock moves VAT ($VREC$), the wage bill moves PAYE and NICs, profits move corporation tax, and the sum lands mechanically in equations (22)–(26). In the emulator these identities solve alongside the rest of the system, so scored reforms carry second-round fiscal aggregates on the same accounting basis as the OBR's published tables—subject to the calibration floors on the chain's unpublished input series documented in Section 5.

3.11 Add-factors and the anchoring mechanism

Every estimated equation in the listing carries an add-factor declaration (EViews `@ADD`), the residual-adjustment device through which the OBR imposes judgement on the pure model solution (Office for Budget Responsibility, 2013): for a behavioural equation $y_t = f(x_t) + a_t$, the add-factor a_t shifts the equation's intercept period by period. The emulator uses the same device for two distinct purposes.

Anchoring. For the baseline, add-factors are computed at every quarter as exactly the residuals that reconcile each behavioural equation to the published EFO path, so the anchored baseline reproduces the OBR's published aggregates *by construction.* Two published income aggregates ($HHDI$, $RHHDI$) are computed by level identities that receive no add-factor; because their unpublished inputs drift, they are instead held exogenously at their EFO values in the anchored solve—closing a leak that otherwise pulls consumption (through the $\Delta \ln RHHDI$ term) about 2 per cent off the EFO. This is anchoring to ground truth, not snapshotting a model error: the shock and baseline runs share the identical anchored structure, so reform deltas isolate the policy effect.

Forecasting. For genuine forecasts the emulator implements the OBR's own held-add-factor method: residuals are fitted over a recent base window (2024Q1–2025Q4), held constant, and the model is projected forward eight quarters. The forecast is the model's structural dynamics anchored by the frozen judgement terms—a real forecast rather than a replay of the input.

Exogenous variables. The 253 exogenous inputs comprise the fiscal levers above, monetary and financial-market assumptions, world demand and trade prices, demographics, and the technical adjustment factors. Where the EFO does not publish an input the model needs, the emulator draws on a vendored ONS snapshot: of 261 endogenous variables referenced by the listing but absent from the EFO tables, 199 had identifiable ONS series codes, yielding a 348-series databank that gives the circular fiscal and financial sub-blocks real balancing data.

4 Method and implementation

The emulator is a pure-Python package (`obr_macro`, Python 3.11–3.13, NumPy/pandas) that ships the OBR equation listing, the November 2025 EFO detailed tables, and the ONS input snapshot as package data, so `pip install` from GitHub is sufficient to run it. The pipeline has five stages.

1. Transpile. A deterministic transpiler rewrites each EViews line into an equivalent Python expression: `dlog(X)` and `d(X)` become log- and first-differences with balanced-parenthesis scanning, lags `X(-n)` become indexed look-backs, `@elem(VAR,"1970Q1")` becomes a base-period constant lookup, `@recode(@date=...)` becomes a period dummy (with EViews month-coded quarters normalised), `@trend` becomes a time index, and `^` becomes `**`. Each equation is classified (identity, error-correction, growth-rate, simultaneous) and compiled once; the dashboard's Equations tab displays every published EViews line beside its transpiled Python for line-by-line audit.

2. Initialise. History and forecast values load from the EFO tables and the ONS snapshot from 2016Q1; any variable an equation references but the data do not supply is created and seeded. Left-hand sides that arrive as transformations ($\Delta \ln y$, Δy, ratios) are solved analytically for the level of y.

3. Solve. Each quarter's 372 interdependent equations are solved simultaneously by Gauss–Seidel iteration (Fair, 2004): cycle through the equations in order, update each left-hand variable in place, and repeat until the maximum change falls below tolerance (with damping and stall detection). Quarters are solved sequentially forward over the horizon; scenario solves run 2025Q1–2027Q4 and baseline solves extend to 2031Q1.

4. Closure swap. The published listing fixes GDP and lets identities distribute it; for policy shocks the emulator swaps the closure, re-arranging identity (2) so that $GDPM$ becomes endogenous and inventories $DINV$ become the residual. Output then responds to demand. Under this demand closure the impact multiplier on a government-consumption shock is approximately one by construction (£1.25bn per quarter of CGG raises quarterly GDP by about £1.25bn on impact). For corporation-tax experiments an optional *investment closure* restores the user-cost channel: $TCPRO$ moves the cost of capital, desired capital, and business investment via equation (5), with the market-sector supply feedback stabilised by tracking-pass add-factors to prevent an explosive output–investment accelerator loop (an eigenvalue-above-one interaction that was diagnosed, bounded, and locked with sign- and magnitude-invariant tests).

5. Shock and compare. `run_reform(name, var, shock, periods)` solves the baseline and the shocked path with identical structure—add-factors on in both—and reports per-quarter deviations (ΔGDP in £bn, per cent of GDP, Δ consumption, Δ investment). Shocks may be scalar (£m per quarter, or decimal rate changes) or explicit quarterly paths.

4.1 The static-costing bridge from PolicyEngine microsimulation

The bridge reproduces the OBR's policy-costing workflow (Office for Budget Responsibility, 2014): static costing in, second-round effects out. A reform is expressed as PolicyEngine parameters, using the same {`parameter_path: value`} dictionary as the rest of the suite—for example {`"gov.hmrc.income_tax.rates.uk[0].rate": 0.21`}—and processed in three steps:

1. *Microsimulation costing.* PolicyEngine's UK model (PolicyEngine, 2025) runs the reform against the enhanced Family Resources Survey microdata for each year of the scoring window, returning the annual budgetary impact in £bn together with household-level detail (winners, losers, net-income change).

2. *Translation.* Annual costings convert to a quarterly additive shock on nominal household disposable income ($HHDI$, £m per quarter): sign-flipped (revenue raised means households keep less), scaled by 1000/4, and applied flat within each year—a deliberately crude and declared interpolation, since the microsimulation produces annual numbers.

3. *Macro propagation.* The call `run_reform(var="HHDI", shock=path)` propagates the shock through the OBR's own demand-side incidence channel, $HHDI \to RHHDI \to CONS \to GDPM$ (equations (15) and (3)), with $HHDI$ exogenised identically in baseline and shocked runs so the delta isolates the reform. The output is the second-round demand effect.

Two boundaries are enforced rather than fudged: the economy's feedback onto disposable income itself (income → tax → income) is not recycled, and corporation-tax reforms are refused by the bridge—corporation tax is not household-borne in the microsimulation—with a pointer to the direct $TCPRO$ lever. Supply-side channels (participation, saving, capital deepening) are deliberately left to the suite's overlapping-generations member; that division of labour is the point of a multi-model suite.

4.2 Hosted surfaces

The emulator is drivable from AI workflows and the command line through the suite's integration layer (`pe-macro`): a CLI and a Model Context Protocol (MCP) server exposing tools including `score_reform` (a PolicyEngine reform through a chosen model, with a common `ScoreResult` schema across model classes), `obr_shock` (raw variable shocks), forecasting, and household calculators. The hosted MCP server deploys automatically to Modal on every merge—merges to the model repository trigger the same redeploy—and a live dashboard (`https://obr-macroeconomic-model.vercel.app`) provides the scenario explorer, a searchable glossary of all 636 variable codes with ONS series, and the full equation listing beside its transpiled Python. Nightly CI validates the live endpoint against the correctness suite.

5 Calibration and validation

All results in this paper are conditioned on the November 2025 EFO vintage (Office for Budget Responsibility, 2025a) and the 15 October 2025 model code. Validation is organised around a principle borrowed from the reliability literature: separate what the model reproduces *by construction* from what it genuinely generates, and gate the former hard while reporting the latter honestly.

5.1 The anchored baseline: a hard gate

With add-factors on (Section 3.11), the anchored baseline must reproduce the OBR's published aggregates—that is what anchoring means, and any deviation is a broken transmission chain, not a forecast error. Table 1 reports the anchored fit over the scored horizon 2025Q1–2027Q4, and Figure 1 shows the two headline paths against the published EFO: the emulator tracks the published levels to within a fraction of a per cent throughout, with the largest gaps at the end of the horizon where the EFO's own quarterly interpolation is coarsest. The GDP mean absolute percentage error is 0.18 per cent and consumption 0.29 per cent; continuous integration hard-fails the build if either exceeds 1 per cent, if any published aggregate turns non-finite on the horizon, or if the expenditure identity (2) fails to close within 0.5 per cent of GDP in any quarter. (An earlier version of the anchored solve missed the EFO by 1.93 per cent on GDP because $HHDI/RHHDI$ ran raw and leaked into the consumption equation; anchoring the published income aggregates closed the gap to 0.18 per cent.)

5.2 The free-running model: an honest scorecard

The true measure of calibration is the raw model with add-factors off. Two flattery traps are removed before scoring. First, *de-seeding*: the solver's data frame is pre-filled with the EFO path, so a quarter whose Gauss–Seidel iteration exits on a stall would otherwise sit near the EFO seed; the scorecard re-solves each period after reseeding every live computed variable from the model's own previous-period value. Second, *passthrough labelling*: a 0.00 per cent error usually does not mean a good forecast—it means the variable was left at the OBR's published

Table 1: Anchored baseline versus the OBR November 2025 EFO, 2025Q1–2027Q4

Invariant	Metric	Value	CI gate
Real GDP ($GDPM$)	MAPE vs EFO	0.18%	hard fail if $\geq 1\%$
Consumption ($CONS$)	MAPE vs EFO	0.29%	hard fail if $\geq 1\%$
Key aggregates finite	NaN/inf on horizon	none	hard fail on any
Expenditure identity	max gap, % of GDP	$< 0.5\%$	hard fail if $\geq 0.5\%$
Spending multiplier sign	$\Delta GDPM$ for $+CGG$	positive	hard fail on sign flip
Corp.-tax investment sign	ΔIF for $\pm TCPRO$	correct, bounded	hard fail on sign/blow-up

Notes: The anchored fit is a by-construction invariant, not a forecast claim: add-factors absorb the model's tracking error so the baseline reproduces the published EFO path. Structural sign invariants (positive fiscal multiplier; a corporation-tax cut raises, and a rise lowers, investment, with the response bounded) are gated alongside it. Source: `tests/test_model_invariants.py` in the model repository.

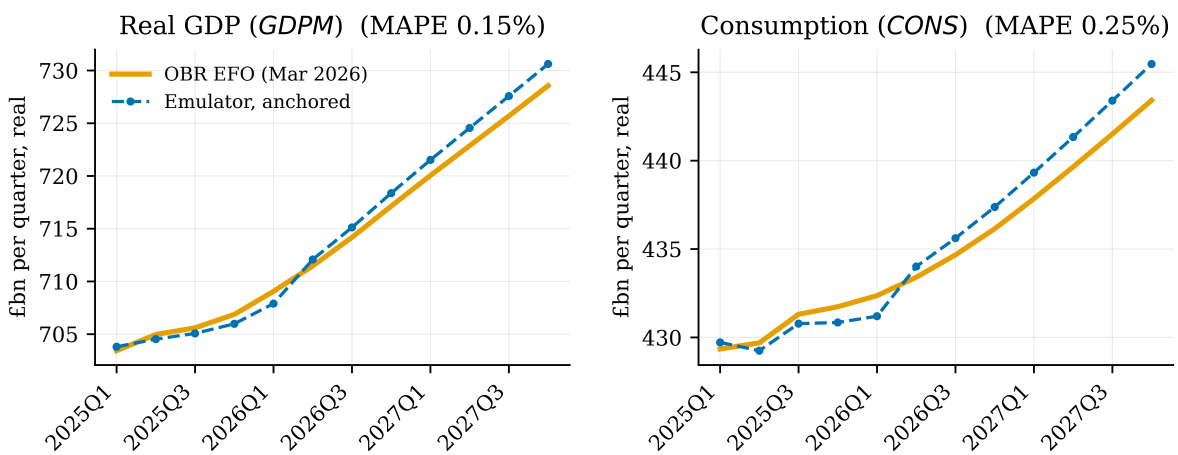

Figure 1: Anchored baseline versus the OBR November 2025 EFO, 2025Q1–2027Q4. Real GDP and consumption in £bn per quarter (chained volumes). The anchored fit is a by-construction invariant (Section 3.11), hard-gated in CI at 1 per cent MAPE. Generated by `figures/fig_anchored.py`.

value because its equation is dead or exogenous. Only variables the model actually computes measure calibration, and the employment identity ($ETLFS = HWA/AVH$, both passthrough inputs) is flagged as trivial rather than counted as a behavioural win.

The scorecard is now also emitted as structured JSON. Its schema retains all headline rows but excludes anchored and passthrough quantities from the accuracy denominator by construction, and states explicitly that the exercise is not a historical-vintage forecast backtest. This turns the paper's disclosure rule into a machine-checkable acceptance contract rather than a presentation choice.

Table 2 presents the scorecard, and Figure 2 shows what the divergence looks like in levels for three of the computed variables: the free-running model *contracts* while the EFO grows—the raw dynamics lose the OBR's judgement about the recovery path and drift down at roughly 1–2 per cent a year—and household income sits £30–120bn below the published path (about £75bn on average across the horizon), the direct image of the unpublished $OSHH$ constant discussed below. The full 21-variable panel, including every passthrough, is reproduced in Appendix B.

"Within band" collapses per-metric thresholds—rates within 1.0pp, net balances within 1.5 per cent of GDP (the OBR's own convention for scoring a difference of two large gross flows), levels within 10 per cent MAPE.

Table 2: Free-running (raw) model versus the OBR November 2025 EFO, 2025Q1–2027Q4

Variable	Metric	Error (3yr)	Status
Real GDP	MAPE	5.70%	computed — fair
Consumption	MAPE	9.45%	computed — fair
Trade balance	% of GDP	0.62%	computed — fair
Employment	MAPE	0.00%	computed — *trivial identity*
Unemployment rate	pp	1.01	computed — poor
RPI inflation	pp	2.03	computed — poor
Household income	MAPE	14.76%	computed — poor
Real household income	MAPE	14.48%	computed — poor
Business investment	MAPE	15.48%	computed — poor
Current account	% of GDP	2.76%	computed — poor
Company profits	MAPE	54.57%	computed — **off**
Investment, exports, imports, employees, avg. earnings, CPI, CPI inflation, GDP deflator, wages, compensation		0.00%	passthrough (held at OBR value)

Notes: De-seeded raw solve, add-factors off. Of 21 headline variables, 11 are computed and 10 are passthroughs; of the 11 computed, 4 land within band (36%)—real GDP, consumption, the trade balance, and the trivial employment identity. Bands: rates <1.0pp, net balances <1.5% of GDP, levels <10% MAPE. Source: `docs/calibration_scorecard.md`, generated by `obr_macro/calibration_score.py`.

Three residual errors deserve explanation because they are *floors*, not bugs. The company-profits error (54.6 per cent) traces to households' operating surplus: the published equation $OSHH = 12874 + 0.85\,IROO - DIPHH^{mf}$ evaluates to roughly £38bn against the OBR databank's roughly £77bn, and the gap lives in the additive base constant and the vintage of the unpublished rent index feeding $IROO$. Because $HHDI$ *adds* $OSHH$ (income runs ∼15 per cent low) while $FYCPR$ *subtracts* it (profits are inflated), one unpublished constant drives a large share of the error in three scored variables at once. Re-tuning the constant to hit the target would be fitting to the answer, so it is documented and gated against regression instead. Similarly, RPI inflation decays because the unpublished $I7/PR$ index history runs out around 2026Q2, and the current-account error is bounded by overseas rate-of-return normalisers the OBR does not publish. These figures are far better than early development versions (business investment 318 per cent, household income ∼932 per cent, company profits ∼375 per cent), which reflected genuine bugs in the incomes, investment, and financial-account blocks that were found and fixed; what remains is bounded by unpublished inputs.

5.3 Held-add-factor forecast

Between the two extremes sits the forecasting configuration (Section 3.11): add-factors fitted over 2024Q1–2025Q4, held constant, projected over 2026Q1–2027Q4. Scored against the EFO, real GDP comes in at 2.2 per cent MAPE, consumption 3.6 per cent, unemployment 0.75pp, and the current account 0.9 per cent of GDP—within band—while business investment (∼18 per cent), company profits (∼47 per cent), and household income (∼12 per cent) sit over band: 10 of 16 headline variables are computed and 6 of those land within band. Honesty caveats apply and

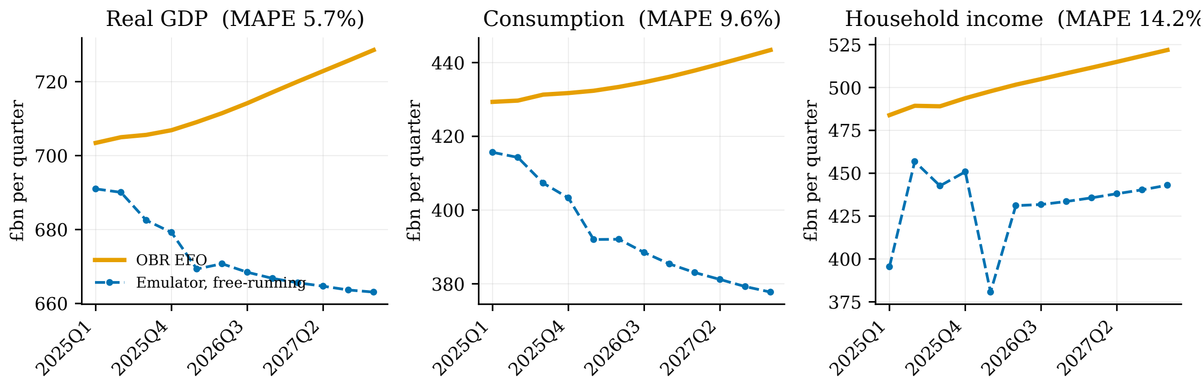

Figure 2: Free-running (raw, de-seeded, add-factors off) model versus the OBR November 2025 EFO, 2025Q1–2027Q4, £bn per quarter. The honest divergence: without the OBR's judgement terms the model drifts below the published path for GDP and consumption, and household income inherits the unpublished $OSHH$ constant's level gap. Generated by `figures/fig_free_running.py`.

are documented with the result: the forecast is initialised at the EFO values it is scored against (a neutral jump-off scores worse); the add-factor base window extends into quarters that are themselves OBR forecast rather than outturn, so the held add-factors partly encode "agree with the OBR"; and the ONS input snapshot is current-vintage data merged into an autumn-2025 baseline, so revision slack is absorbed into the add-factors.

6 Results: a worked reform

To illustrate the full pipeline, consider raising the basic rate of income tax by one percentage point (20p to 21p) from April 2026—the canonical illustrative change in HMRC's ready reckoner. Expressed as a PolicyEngine reform it is a single parameter, {`"gov.hmrc.income_tax.rates.uk[0].rate": 0.21`}, scored through the bridge over a five-year window, 2026–2030.

Static costing. PolicyEngine's microsimulation puts the direct Exchequer yield at £6.46 billion in 2026, rising with fiscal drag and earnings growth to £7.38 billion by 2030 (Table 3). Household net income falls by a corresponding amount, and losers far outnumber winners—as expected for a broad-based rate rise.

Second-round effects. Converted to a quarterly $HHDI$ shock (roughly $-$£1.6bn to $-$£1.8bn per quarter) and propagated through the consumption function (3) and the expenditure identity, the reform lowers GDP by 0.02–0.03 per cent relative to baseline over the first scored year: Figure 3 traces the quarterly path, from -0.020 per cent on impact in 2026Q1 through -0.040 per cent by 2026Q4 to -0.057 per cent by 2027Q4. The drag builds slowly because households smooth part of the income loss: the short-run marginal propensity to consume out of income in equation (3) is modest, and the equilibrium-correction term of Section 3.1 then cumulates the income loss into the consumption level at $\sim$12.5 per cent per quarter, so the demand effect remains small relative to the mechanical costing throughout the window. This is the sign and order of magnitude the OBR's own indirect-effect adjustments deliver for income-tax measures (Office for Budget Responsibility, 2014).

Table 3: Basic rate of income tax +1pp, scored through the static-costing bridge

	2026	2030
Static costing (PolicyEngine microsimulation), £bn/yr	+6.46	+7.38
Quarterly $HHDI$ shock, £bn/qtr	−1.62	−1.85
Second-round GDP effect (OBR emulator), % of GDP	−0.02	−0.06*
HMRC ready reckoner, 1p on basic rate, £bn/yr	6.9 (2026–27) to ≈8.2 (2028–29)	

Notes: Reform `{"gov.hmrc.income_tax.rates.uk[0].rate": 0.21}` from 2026, scored via `obr_score_reform` (Section 4.1): PolicyEngine annual costings on enhanced FRS microdata, converted to a flat-within-year quarterly shock on nominal household disposable income, propagated through the anchored OBR emulator under the demand closure. The GDP effect is the deviation of $GDPM$ from the anchored baseline; *the macro solve runs to 2027Q4, so the second column reports the end-of-window (2027Q4) effect, not a 2030 value. HMRC comparison: HM Revenue and Customs (2025).

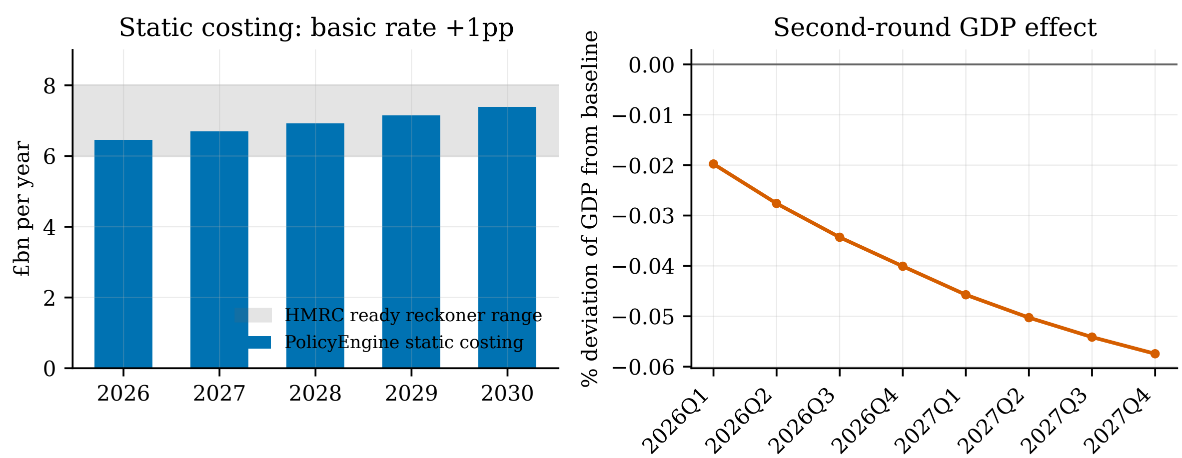

Figure 3: The 1p basic-rate reform. Left: PolicyEngine static costing by year (endpoints as scored through the bridge, interior years interpolated) against the range of HMRC's published ready reckoner—the June 2025 vintage puts 1p on the basic rate at £6.9bn in 2026–27 and £8.2bn a year in the two following years (HM Revenue and Customs, 2025). Right: quarterly second-round GDP effect of the corresponding $HHDI$ shock through the anchored emulator under the demand closure. Generated by `figures/fig_reform.py`.

Comparison with HMRC's ready reckoner. HMRC's *Direct effects of illustrative tax changes* bulletin—the official ready reckoner, produced on the OBR-certified costing basis—puts a 1p change in the basic rate at £6.9 billion in 2026–27, rising to about £8.2 billion a year in 2027–28 and 2028–29 (HM Revenue and Customs, 2025), and published vintages over recent years span roughly £6–8 billion per 1p. The PolicyEngine static costing of £6.46–7.38 billion sits within that range, toward its lower end; differences of this size are expected across costings that differ in data vintage (survey microdata versus HMRC administrative Survey of Personal Incomes), baseline policy assumptions (threshold-freeze paths), and behavioural adjustments (HMRC's figures embed taxable-income elasticities for higher incomes). Under the OBR's policy-costing conventions (Office for Budget Responsibility, 2014), the ready-reckoner figure corresponds to the post-behavioural *direct* effect, while the economy-wide feedback is applied at the forecast level—exactly the decomposition the bridge reproduces: the microsimulation supplies the direct effect, the macro model the indirect one.

Direct macro levers. Reforms outside the household tax–benefit space use the exogenous levers directly—government consumption CGG, nominal government investment $CGIPS$, the corporation-tax rate $TCPRO$, and (because the CPI is exogenous in this configuration) the consumer price level. Section 7 works three such experiments in full.

7 Further experiments: the direct macro levers

The bridge of Section 4.1 covers reforms whose incidence is household-borne. This section works three experiments that instead pull the model's exogenous levers directly, chosen to exercise three different transmission channels: a pure demand-side expenditure shock, a pure price shock, and a cost-of-capital shock. All are run with `run_reform` over 2025Q1–2027Q4 with the shock applied for twelve quarters; Table 4 reports annual averages of the quarterly deviations, and the underlying quarterly paths are written to `figures/experiments_quarterly.csv` by the script `figures/run_experiments.py` that generated them.

Table 4: Three policy experiments through the direct levers, annual averages of quarterly deviations from the anchored baseline

	ΔGDP, % of GDP	ΔGDP £bn/qtr	ΔConsumption £bn/qtr	ΔInvestment £bn/qtr
A. Government consumption +£ 5bn/yr (CGG +£1.25bn/qtr)				
2025	+0.18	+1.250	0.000	0.000
2026	+0.19	+1.250	−0.000	0.000
2027	+0.19	+1.249	−0.001	0.000
B. VAT-style price shock (CPI level +1 per cent, +1.4 index points)				
2025	−0.15	−1.014	−1.014	0.000
2026	−0.41	−2.737	−2.737	0.000
2027	−0.46	−3.070	−3.070	0.000
C. Corporation tax +2pp ($TCPRO$ +0.02, investment closure)				
2025	−0.005	−0.034	−0.000	−0.034
2026	−0.04	−0.295	−0.004	−0.290
2027	−0.13	−0.901	−0.014	−0.880

Notes: Each experiment solves a structurally identical baseline and shocked pair (add-factors shared) and reports the delta, 2025Q1–2027Q4, shock applied for 12 quarters. Experiment C uses the investment closure with the reconstructed $\Delta \ln IBUSX$ equation (5) and the stabilised market-sector feedback (Section 4). Source: `figures/run_experiments.py`; quarterly detail in `figures/experiments_quarterly.csv`.

A. Government consumption +£5bn a year. The canonical demand shock: real CGG rises £1.25bn per quarter. Under the demand closure the impact multiplier is one by construction: GDP rises by exactly the shock in 2025Q1, and the interesting content is what happens *after* impact: the multiplier neither builds (no accelerator operates, because investment is a passthrough outside the investment closure) nor decays materially (imports are exogenous in this configuration, so the leakage that would normally erode a spending multiplier through equation (11) is inoperative), and consumption crowds out only trivially ($-$£1bn *cumulative* by 2027Q4) through the small negative feedback of the price and income channels. The honest reading is that the emulator's CGG multiplier of ≈ 1.0 throughout is the direct mechanical effect plus approximately zero second-round dynamics, which brackets it between the OBR's own impact-multiplier assumptions for spending measures (below one once import leakage and offsets are applied (Office for Budget Responsibility, 2014)) and older Keynesian estimates above one: the emulator omits the import leakage (downward force) and the induced-income consumption loop (upward force) alike, and reports the shock arithmetic in between.

B. A VAT-style price shock. Because the CPI is an exogenous conditioning path (Section 3), a one per cent rise in the consumer price level (+1.4 index points held for twelve quarters) is a legitimate lever, and it is the closest supported analogue to the price effect of an indirect-tax rise: the OBR's costing convention for VAT changes assumes passthrough into consumer prices, which then erode real incomes. The transmission here is the deflator chain $CPI \rightarrow PCE \rightarrow RHHDI$ into the consumption ECM (3): with nominal income initially unchanged, real income falls one per cent, and consumption follows—-0.15 per cent of GDP in 2025, deepening to -0.46 per cent ($-$£3.1bn per quarter) by 2027 as the error-correction term drags the consumption *level* toward its lower long-run solution. The entire GDP effect is consumption ($\Delta CONS = \Delta GDP$ to the pound in Table 4), which is the configuration speaking: investment, trade, and the labour-demand response that would spread the shock are passthrough or exogenous here. For scale, HMRC's ready reckoner puts the *direct revenue* of a 1pp VAT standard-rate rise at £8.8–9.6bn a year (HM Revenue and Customs, 2025); a full VAT scoring would put that costing through the bridge as an $HHDI$-equivalent real-income loss, and the experiment here isolates the complementary price-channel arithmetic on consumer demand.

C. Corporation tax +2pp, and why the bridge refuses it. By design, the bridge declines corporation-tax reforms: in the microsimulation, corporation tax has no household-resolution incidence—attributing it to households would smuggle in an incidence assumption (shareholders? workers? consumers?) that PolicyEngine's UK model deliberately does not make—so translating a corporation-tax costing into an $HHDI$ shock would be a category error, and the correct route is the model's own lever. Raising $TCPRO$ by two percentage points with the investment closure active works the user-cost channel $TCPRO \rightarrow TAF \rightarrow COC \rightarrow KSTAR \rightarrow KGAP \rightarrow IBUSX$: a higher tax rate raises the post-tax cost of capital, lowers desired capital, opens a capital-stock gap, and depresses business investment through the error-correction term of equation (5). The response is small and slow—$-$£0.03bn per quarter in year one, building to $-$£0.9bn per quarter (-0.13 per cent of GDP) by 2027—exactly the profile the slow investment ECM ($\lambda = 0.042$) implies, and consistent in sign and magnitude class with the OBR's treatment of corporation-tax measures as chiefly a receipts effect with modest medium-term investment drag. For comparison, the *direct* Exchequer yield of +2pp on the main rate is about £7.2bn in 2026–27 rising to £8.0bn in 2028–29 on HMRC's ready reckoner (HM Revenue and Customs, 2025); the emulator's contribution is the second-round path, not the static number. The response is hard-gated in CI for sign and boundedness (Table 1), and Section 9 flags the stabilised closure it depends on as a stop-gap.

Across the three experiments a single lesson recurs: where a channel is live, the emulator reproduces the OBR architecture's sign, shape, and magnitude class; where a channel is passthrough, the emulator does not invent a response—it reports zeros and says why. That property—experiments degrade legibly rather than plausibly—is the design goal the honest scorecard of Section 5.2 exists to protect.

8 Comparison with official results

The preceding sections benchmark each result against its official counterpart where one exists. This section consolidates those comparisons into a single side-by-side accounting—our numbers, the official source, and the deviation—so that the reader can audit the claim of the paper in one place, and then steps outside the November 2025 information set: against the OBR's subsequent March 2026 EFO (vintage drift), against ONS outturn data published after the anchor vintage (forecast versus outturn), against the published fiscal-multiplier literature, and against the official policy-costings record. Table 5 collects four panels: the anchored levels against the published EFO, the worked income-tax reform against HMRC's ready reckoner, the two direct-lever experiments against the corresponding reckoner lines, and the tracking errors of the three model configurations. Figure 4 shows the two like-for-like comparisons graphically.

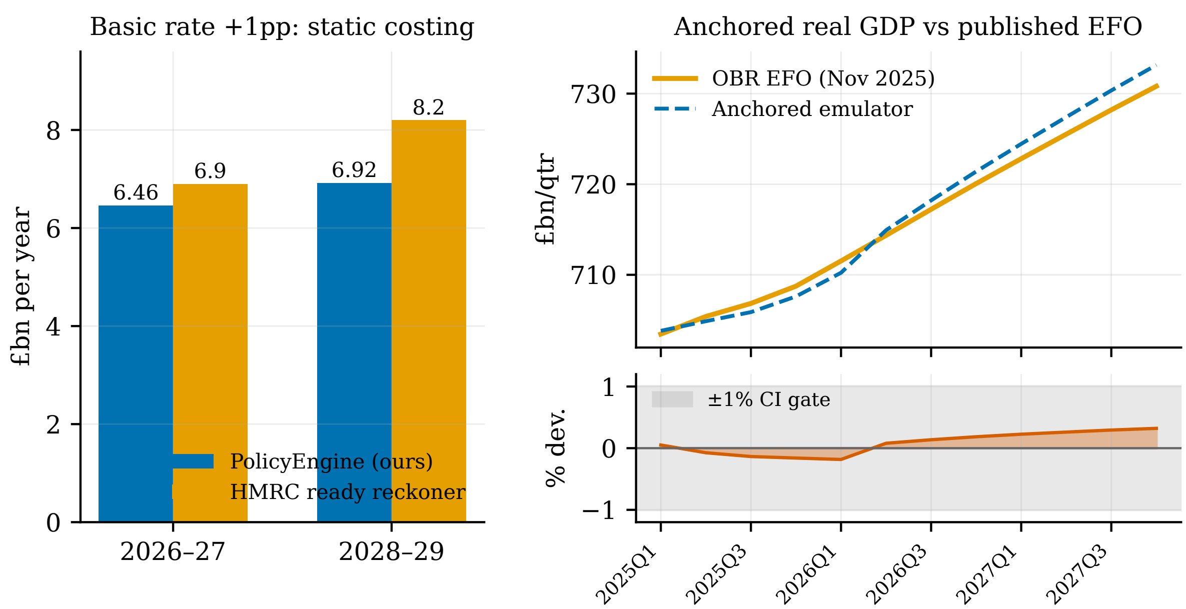

Figure 4: Ours versus official. Left: the PolicyEngine static costing of the basic-rate +1pp reform against HMRC's ready reckoner (June 2025 vintage); the 2028–29 bar for the emulator is interpolated between the scored endpoints. Right: the anchored real-GDP path overlaid on the published EFO (top) and the quarterly percentage deviation against the ±1 per cent CI gate (bottom). Generated by `figures/fig_comparison.py` from `figures/fig_anchored_data.csv`.

Table 5: Our results versus the official sources

		Ours	Official	Deviation
A. Anchored levels vs OBR EFO, November 2025 (Office for Budget Responsibility, 2025a), £bn/qtr				
Real GDP	2025Q1	703.8	703.5	+0.05%
	2025Q4	707.6	708.7	−0.16%
	2026Q4	721.4	720.1	+0.19%
	2027Q4	733.2	730.8	+0.32%
Consumption	2025Q1	429.6	429.2	+0.08%
	2025Q4	429.9	431.1	−0.26%
	2026Q4	438.9	437.6	+0.30%
	2027Q4	446.7	444.4	+0.53%
B. Basic rate +1pp costing vs HMRC ready reckoner (HM Revenue and Customs, 2025), £bn/yr				
2026–27		6.46	6.9	−6.4%
2028–29†		6.92	8.2	−15.6%
2030 (end of window)		7.38	≈8.2	−10.0%
C. Direct-lever experiments vs the reckoner's direct yields (HM Revenue and Customs, 2025)				
VAT +1% price shock	GDP effect 2027, £bn/qtr	−3.1	—	—
VAT +1pp direct yield	£bn/yr	—	8.8–9.6	n/a
Corp. tax +2pp, 2nd round	GDP effect 2027, %	−0.13	—	—
Corp. tax +2pp direct yield	£bn/yr	—	7.2–8.0	n/a
D. Tracking error vs the published EFO path, MAPE 2025Q1–2027Q4				
Anchored	GDP / consumption	0.18% / 0.29%	CI gate < 1%	
Held add-factors	GDP / consumption	2.2% / 3.6%	within band	
Free-running (raw)	GDP / consumption	5.70% / 9.45%	over band	

Notes: Panel A: selected quarters from the anchored solve of Section 5 (full path in `figures/fig_anchored_data.csv`); the anchored fit is a by-construction invariant, so the deviations measure residual transmission error, not forecast skill. Panel B: HMRC *Direct effects of illustrative tax changes*, June 2025 vintage; †our 2028 figure is linearly interpolated between the scored endpoints (£6.46bn in 2026, £7.38bn in 2030). Panel C is deliberately *not* scored with a deviation: the reckoner lines are direct Exchequer yields while the experiments deliver the complementary second-round macro paths (Section 7), so the objects differ and a percentage gap would be a category error. Panel D: Sections 5 and 5.2; the free-running MAPEs are the honest de-seeded scores with passthroughs excluded.

8.1 The consolidated ledger

Three readings follow. First, the like-for-like rows agree to well within the tolerances the sources themselves imply: the anchored levels sit within a third of a per cent of the published EFO everywhere on the horizon (largest at 2027Q4, where the EFO's own quarterly interpolation is coarsest), and the income-tax costing lands 6–16 per cent below HMRC's reckoner—inside the £6–8bn range spanned by recent published vintages, with the shortfall concentrated in the later years where HMRC's figures embed administrative-data fiscal drag that survey micro-data capture less fully (Section 6). Second, where no like-for-like object exists—the VAT and corporation-tax rows—the table reports the official direct yield and our second-round path side by side without manufacturing a percentage: the emulator's contribution there is the piece the reckoner does not publish, not a rival estimate of the piece it does. Third, panel D quantifies how much of the model's fidelity is inherited rather than generated: the same equations that track the EFO to 0.18 per cent when anchored miss it by 5.7 per cent free-running, which is precisely the gap the OBR's own add-factor judgement closes in the official process, and the reason reform deltas are always scored against the anchored baseline.

8.2 Vintage drift: the March 2026 EFO

Everything above compares against the November 2025 EFO because that is the vintage the emulator is anchored to. The OBR's next forecast, the March 2026 EFO (Office for Budget Responsibility, 2026), provides a natural out-of-vintage check: how far has the official view moved away from the baseline this paper conditions on? The answer is material and one-directional. The March EFO revises 2026 real GDP growth down to 1.1 per cent, 0.3 percentage points below November, citing weaker-than-expected GDP outturns in late 2025, a loosening labour market, and subdued business surveys; growth then averages 1.6 per cent a year over 2027–2030, and 2026 CPI inflation is revised down 0.2 percentage points. Table 6 sets the emulator's anchored calendar-year growth rates beside both vintages. The anchored 2026 paths—1.5 per cent GDP growth and 1.4 per cent consumption growth on the model side, against the November EFO's 1.4 and 1.2—now sit 0.3–0.4 percentage points above the official March view. This gap is *vintage drift*, not model error: it is almost exactly the OBR's own November-to-March revision, and an emulator re-anchored to the March tables would inherit the new path by construction (re-anchoring at each vintage is the maintenance loop described in Section 10). The practical lesson for users is about levels, not deltas: reform effects are differences between structurally identical runs and are insensitive to modest baseline drift, but anyone quoting the anchored *baseline* should treat it as a November 2025 snapshot that the official forecaster has since marked down.

Table 6: Vintage drift: anchored baseline versus the November 2025 and March 2026 EFOs

Real growth, % per year	Anchored emulator	EFO Nov 2025	EFO Mar 2026
GDP, 2026	1.5	1.4	1.1
GDP, 2027	1.8	1.5	1.6[a]
Consumption, 2026	1.4	1.2	—
Consumption, 2027	1.9	1.5	1.6[b]

Notes: Calendar-year growth of annual averages of the quarterly paths in `figures/fig_anchored_data.csv`; the emulator column differs from the November EFO column only by the residual anchoring error of Table 1, compounded across quarters. Sources: Office for Budget Responsibility (2025a, 2026). [a]The March 2026 EFO reports 2027–2030 GDP growth as an average of 1.6 per cent a year rather than a 2027 point estimate. [b]March 2026 consumption is likewise published as a five-year average (1.6 per cent a year); no 2026 point figure appears in the summary text.

8.3 Forecast versus outturn: ONS data published since anchoring

Comparing one forecast vintage with another tests agreement, not accuracy. Since the paper's anchor vintage, the ONS has published outturn estimates for the quarters the anchored path forecasts, allowing a genuine forecast-versus-outturn check. On the latest vintage available at the time of writing (Office for National Statistics, 2026), quarter-on-quarter real GDP growth was 0.1 per cent in 2025Q2, 0.2 in Q3, 0.2 in Q4, and 0.6 in 2026Q1, with 2025 annual growth of 1.4 per cent. The left panel of Figure 5 sets the anchored emulator's quarterly growth against those outturns: the model (0.15, 0.14, 0.25, 0.37) tracks the three 2025 quarters to within 0.06 percentage points but—like the November EFO it is anchored to (0.28, 0.20, 0.27, 0.39)—misses the strong 2026Q1 outturn by roughly a quarter of a percentage point, the same surprise that led the OBR to describe late-2025 outturns as weaker and to re-profile growth in March.

Two honesty caveats govern the reading. First, the anchored path is the November EFO plus residual anchoring error, so this panel is primarily a test of the OBR's November vintage that the emulator inherits—the emulator's own contribution to the errors is the 0.02–0.13 point gap between the blue bars and the orange markers. Second, ONS quarterly estimates are themselves revised (2025Q4 was first published at 0.1 per cent and 2025 annual growth has moved between 1.3 and 1.4 across vintages), so the "outturn" is a moving target of the kind Section 9 flags under vintage sensitivity.

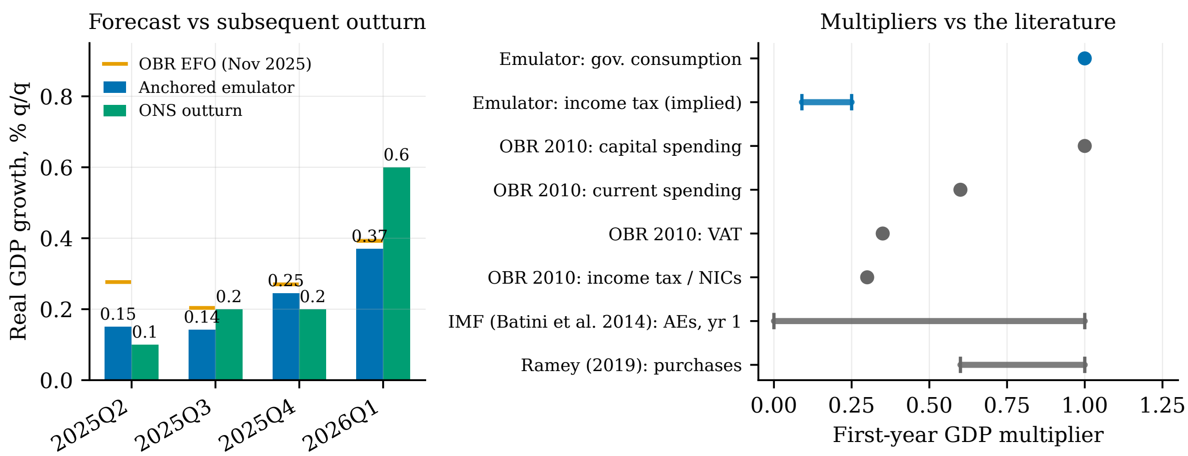

Figure 5: Stepping outside the anchor vintage. Left: quarter-on-quarter real GDP growth, the November-2025-anchored emulator forecast (bars, with the published EFO path as markers) against ONS outturns published after anchoring (Office for National Statistics, 2026). Right: the emulator's government-consumption multiplier and the implied income-tax multiplier of the 1p reform (impact to 2027Q4 range) against published first-year multiplier estimates: the interim OBR's June 2010 assumptions by instrument (Office for Budget Responsibility, 2010), the IMF's bunching range for advanced economies in normal times (Batini et al., 2014), and Ramey (2019). Generated by `figures/fig_outturn.py`; data in `figures/fig_outturn_data.csv`.

8.4 Multipliers against the published literature

Section 7 reports the emulator's multiplier on government consumption as approximately 1.0 throughout—the mechanical demand-closure effect with both the import leakage (downward force) and the induced-income loop (upward force) inoperative. The right panel of Figure 5 places that number, and the income-tax response, in the published range. The OBR's own multiplier assumptions, set out when the interim OBR costed the June 2010 consolidation and still its most explicit published statement, run from 0.3 for income-tax and NICs changes and 0.35 for VAT through 0.6 for current departmental and welfare spending to 1.0 for capital spending, tapering to zero over about five years as monetary policy and real wages adjust (Office for Budget Responsibility, 2010). The IMF's guidance note recommends first-year multipliers of 0 to 1 for advanced economies in normal times, with spending measures toward the top of the range and revenue measures toward the bottom (Batini et al., 2014), and Ramey (2019), surveying the post-crisis empirical renaissance, puts the plausible range for aggregate government-purchases multipliers at 0.6 to 1.0. Against these, the emulator's CGG multiplier of ≈1.0 sits at the top of the OBR's spending range—above the 0.6 the OBR would assume for a current-spending measure, for the reason Section 7 documents: import leakage is switched off in this configuration, so the force that would pull the multiplier below one is absent. The income-tax side is more comfortable: the 1p reform's second-round GDP effect, expressed as a ratio to the $HHDI$ shock (both in per cent of GDP), implies a tax multiplier of 0.09 on impact rising to 0.25 by 2027Q4—approaching the OBR's 0.3 income-tax assumption from below as the consumption error-correction term cumulates, and inside the revenue end of the IMF range. The honest summary is that the emulator's multipliers are bracketed by, and mostly consistent with, the published assumptions, with the one out-of-range feature (no import leakage on the spending multiplier) both explained and flagged as future work.

8.5 The policy costings record

A final check is against the official policy-costings record—the measure-by-measure database the OBR maintains alongside each EFO and the Treasury's certified costings documents. No recent fiscal event contains a basic-rate change to score against directly: the November 2025 Budget raised revenue through threshold freezes and rate changes on property, dividend, and savings income rather than the main rates (HM Treasury, 2025), so HMRC's ready reckoner (Table 5, panel B) remains the like-for-like benchmark for the worked reform. The nearest certified personal-tax measure is instructive on scale: the Budget's extension of the income-tax threshold freeze to 2030–31 is costed at £8.0 billion in 2029–30 (HM Treasury, 2025; Office for Budget Responsibility, 2025a)—the same order as the reckoner's £8.2 billion for 1p on the basic rate in 2028–29, and consistent with the well-known equivalence between freeze extensions and small rate rises under fiscal drag. The emulator's costing chain (PolicyEngine microdata to £6.5–7.4 billion for 1p) sits 10–15 per cent below both official numbers in the later years, for the microdata-versus-administrative-data reasons documented in Section 6; scoring the freeze extension itself through the bridge is a natural next validation exercise, since both the static costing and the OBR's certified number are published.

9 Limitations

The PolicyEngine Macro suite grades its members on a *verification gradient*: how far each model's outputs can be audited against ground truth should temper the weight placed on its answers. The OBR emulator is a *replication with a published anchor*—auditable against the EFO tables—which is a stronger epistemic position than a calibrated counterfactual, but the audit itself reveals clear limits. In the spirit of reporting them rather than rounding them away:

1. **The anchored fit is by construction, not skill.** The 0.18/0.29 per cent reproduction of the EFO (Table 1) demonstrates that the transmission chains are wired correctly; it says nothing about forecasting ability. All forecasting claims rest on the much weaker free-running and held-add-factor scorecards.

2. **The free-running model is weak, and says so.** Only 4 of 11 computed headline variables land within band; 10 of 21 are passthroughs held at the OBR value (Table 2). The suite's documentation describes the free-running fit as "deliberately report-only, and weak," and this paper endorses that description.

3. **Unpublished inputs impose hard floors.** The OBR does not publish its full databank: the $OSHH$ base constant (behind the 54.6 per cent company-profits error and the $\sim$15 per cent incomes errors), the $I7/PR$ index history (behind decaying RPI inflation), and the overseas rate-of-return normalisers (behind the current-account error) are unrecoverable from public sources. These errors are gated against regression, not chased to zero, because closing them would mean fitting to the answer.

4. **Dead and exogenised channels.** Exports, imports, CPI, the labour-demand block, and wages are exogenous in this configuration—held at the OBR path. Reforms whose first-round incidence runs through those channels (e.g. trade or supply-side labour measures) are outside the emulator's current scope, and employment "accuracy" is a trivial identity over two passthrough inputs.

5. **Two household-income equations do not fire.** The transpiler's left-hand-side parser recognises the $\mathrm{dlog}(X)$, $d(X)$, $X/X(-n)$ and level forms, but not a bare $\log(X)$ left-hand side. The October 2025 model file contains exactly two such equations—$\log(HHTFA)$ and $\log(NDIVHH)$—which are consequently indexed under a key that is not a databank column, so they never execute and their variables stay frozen at their historical paths. This is not cosmetic: $HHTFA$ enters $NMTRHH$ and $NDIVHH$ enters $PIRHH$, and both of those enter $HHDI$—the very quantity the microsimulation bridge shocks. In particular $\log(NDIVHH)$ is a function of lagged company profits $FYCPR$ and $CORP$, so household dividend income should respond to a corporate-profits change and here does not, which mutes the household-income channel of the corporation-tax experiment of Section 7 specifically. The defect was found while investigating solver warning output; the warning is now raised once per solve and names the affected variables rather than being suppressed. The parser has since gained a log left-hand-side kind ($\log(X) = rhs \Rightarrow X = \exp(rhs)$; merged upstream as obr-macroeconomic-model PR #16, after this paper's results were computed), but parser support does not by itself revive the equations, for a second and independent reason: their exogenous right-hand-side inputs, $MAJGDP$ and $CORP$, are absent from the databank entirely. No equation computes them and neither appears in the exogenous seed data, so the right-hand side would evaluate non-finite and the solver would skip the update. Parser support is therefore necessary but not sufficient: with the fix merged, the two lines remain passthroughs held at the OBR value, and the scorecard continues to report them as such. The channel's inertness is directly demonstrable—a five-percentage-point corporation-tax rise moves $FYCPR$ by $-£1{,}780$m while $\Delta NDIVHH$ is exactly zero. Reviving the channel requires sourcing a $CORP$ series, which is a calibration decision rather than a code change: the fitted constant -8.6056 implies a scale the published OBR material does not document, and choosing one to make the level come out right would be fitting to the answer. Every figure in this paper is therefore computed with these two channels inert.

6. **Bridge assumptions are crude and declared.** The static costing lands entirely on nominal household disposable income, flat within each year; the economy's feedback onto disposable income is not recycled; incidence is demand-side only; and corporation-tax reforms are refused rather than mistranslated. Supply-side responses belong to the suite's OLG member.

7. **Vintage sensitivity and judgement contamination.** The ONS input snapshot is

current-vintage data merged into an autumn-2025 baseline, so revision slack silently enters the add-factors; and the held-add-factor base window extends into quarters that are the OBR's own forecast, so "held judgement" partly means "agree with the OBR." The forecast scorecard also inherits a jump-off flatter from EFO seeding.

8. **Stabilised, not structural, investment closure.** The corporation-tax investment closure freezes market-sector output feedback at a tracking-pass reference to suppress an explosive accelerator loop; the mandated sign and boundedness invariants hold, but the closure is a stop-gap pending a populated market-sector supply block, and it is computationally expensive on the hosted service.

9. **No official counterpart to check dynamics against.** The OBR publishes the model listing and a description (Office for Budget Responsibility, 2013) but no official working paper, ready-reckoner of model multipliers, or reference simulation set for the current vintage; the emulator's shock responses can be checked for sign, magnitude class, and convention-consistency, but not line-by-line against an official impulse response.

10 Conclusion

This paper has described an open Python emulator of the OBR's macroeconometric model: 372 published EViews equations transpiled and solved in the open, anchored to the November 2025 EFO by the OBR's own add-factor device, and connected to household-resolution microsimulation through a static-costing bridge that reproduces the official "static costing in, second-round effects out" workflow. The anchored baseline reproduces the published forecast to 0.18 per cent on GDP and is hard-gated in CI; the free-running model is documented as weak, with every residual error traced to a fixed bug or a floor imposed by unpublished OBR inputs. A worked 1p basic-rate reform yields a costing within the range of HMRC's published ready reckoner, with small negative second-round GDP effects of the sign and size the official conventions imply.

The broader aim is methodological. Official fiscal scoring in the UK runs through a model that is described but not runnable by the public. Emulating it openly—with its judgement device made explicit, its passthroughs labelled, and its calibration failures published alongside its successes—moves the transparency frontier: researchers can now inspect the marginal propensities and error-correction speeds that shape UK costings, run counterfactual shocks through the official transmission mechanisms, and see precisely where the public information runs out. Within the PolicyEngine Macro suite, the emulator supplies the near-term, official-convention answer alongside the OLG model's long-run structural answer and the microsimulation's statutory one; scoring the same reform through all three, on a declared verification gradient, is the point.

Future work includes populating the market-sector supply block (retiring the stabilised investment closure), reviving the trade and labour-demand channels currently held exogenous, recycling income feedback through the bridge, and re-anchoring to each new EFO vintage as it is published.

References

Batini, N., Eyraud, L., Forni, L., and Weber, A. (2014). Fiscal multipliers: Size, determinants, and use in macroeconomic projections. Technical Notes and Manuals 14/04, International Monetary Fund.

Blanchard, O. (2018). On the future of macroeconomic models. *Oxford Review of Economic Policy*, 34(1–2):43–54.

Brayton, F. and Tinsley, P. A. (1996). A guide to FRB/US: A macroeconomic model of the United States. *Finance and Economics Discussion Series*, 1996-42. Board of Governors of the Federal Reserve System.

Burgess, S., Fernández-Corugedo, E., Groth, C., Harrison, R., Monti, F., Theodoridis, K., and Waldron, M. (2013). The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models. *Bank of England Working Paper*, No. 471. Bank of England, London.

Chang, A. C. and Li, P. (2015). Is economics research replicable? sixty published papers from thirteen journals say "usually not". *Finance and Economics Discussion Series*, 2015-083. Board of Governors of the Federal Reserve System.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *The Economic Journal*, 88(352):661–692.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Fair, R. C. (2004). *Estimating How the Macroeconomy Works*. Harvard University Press, Cambridge, MA.

Hantzsche, A., Lopresto, M., and Young, G. (2018). Using NiGEM in uncertain times: Introduction and overview of NiGEM. *National Institute Economic Review*, 244(1):R1–R14.

Hendry, D. F. and Muellbauer, J. N. J. (2018). The future of macroeconomics: Macro theory and models at the Bank of England. *Oxford Review of Economic Policy*, 34(1–2):287–328.

Herndon, T., Ash, M., and Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, 38(2):257–279.

HM Revenue and Customs (2025). Direct effects of illustrative tax changes. Official statistics bulletin, GOV.UK. Ready reckoner of tax changes for 2026–27 to 2028–29, April 2026 implementation.

HM Treasury (2025). Autumn budget 2025: Policy costings. Technical report, HM Treasury. Costings certified by the Office for Budget Responsibility.

Klein, L. R. (1950). *Economic Fluctuations in the United States, 1921–1941*. Cowles Commission Monograph No. 11. John Wiley & Sons, New York.

Office for Budget Responsibility (2010). Fiscal multipliers. Forecasts in depth box, https://obr.uk/box/fiscal-multipliers/; multiplier assumptions first set out in the interim OBR's June 2010 Budget forecast.

Office for Budget Responsibility (2013). The macroeconomic model. Briefing Paper No. 5, Office for Budget Responsibility.

Office for Budget Responsibility (2014). Policy costings and our forecast. Briefing Paper No. 6, Office for Budget Responsibility.

Office for Budget Responsibility (2025a). Economic and fiscal outlook – november 2025. Technical report, Office for Budget Responsibility. CP 1403, published 26 November 2025.

Office for Budget Responsibility (2025b). OBR macroeconomic model. Forecasts in depth, https://obr.uk/forecasts-in-depth/obr-macroeconomic-model/. Model code and variable listing of 15 October 2025.

Office for Budget Responsibility (2026). Economic and fiscal outlook – march 2026. Technical report, Office for Budget Responsibility.

Office for Budget Responsibility and HM Treasury (2011). Memorandum of understanding: The macroeconomic model. Technical report, Office for Budget Responsibility.

Office for National Statistics (2026). GDP first quarterly estimate, UK: January to march 2026. Statistical bulletin, published 14 May 2026.

PolicyEngine (2025). PolicyEngine: Open-source tax and benefit microsimulation. https://policyengine.org. Software and documentation; UK model at https://github.com/PolicyEngine/policyengine-uk.

Ramey, V. A. (2019). Ten years after the financial crisis: What have we learned from the renaissance in fiscal research? *Journal of Economic Perspectives*, 33(2):89–114.

Wallis, K. F., Andrews, M. J., Fisher, P. G., Longbottom, J. A., and Whitley, J. D. (1987). *Models of the UK Economy: A Fourth Review by the ESRC Macroeconomic Modelling Bureau*. Oxford University Press, Oxford.

Wren-Lewis, S. (2018). Ending the microfoundations hegemony. *Oxford Review of Economic Policy*, 34(1–2):55–69.

A Variable glossary

Table 7 defines every OBR mnemonic used in this paper, with the definition from the OBR's published variable listing of 15 October 2025 and the ONS series identifier where the listing records one. The table is generated mechanically from the published workbook by `figures/make_glossary.py`; the full 636-code glossary, with live ONS series links, is browsable on the model dashboard's Glossary tab. Seven further codes appearing in the paper are emulator- or scorecard-internal rather than OBR mnemonics: `AWEI` (average weekly earnings, scorecard label), `CBIUD` (CBI investment-uncertainty balance, an input to the reconstructed investment equation, defaulted to zero), `COMP` (compensation of employees, scorecard label), `CPIGR` (CPI inflation rate), `EEES` (employees, scorecard label), `RPIGR` (RPI inflation rate, EFO scorecard series), and `PSNB` (used in the text for the seasonally adjusted net-borrowing concept carried in the listing as `PSNBNSA`/`PSNBCY`).

Table 7: Variable glossary: every mnemonic used in this paper, with the OBR's published definition and ONS identifier where one exists. The full 636-code glossary is browsable on the model dashboard.

Mnemonic	Definition (OBR variable listing)	ONS code
ADJW	Time varying coefficient for wages & salaries	No Codes
AVH	Average weekly hours, all workers	YBUV
CB	Current balance	HBOP
CC	Council tax accruals	NMIS
CDUR	HH final consumption expenditure: durable goods (CVM)	UTID
CETAX	Customs & Excise taxes	ACAC
CGG	General Govt final consumption CVM	NMRY
CGGPS	General Govt final consumption	NMRP
CGIPS	CG gross fixed capital formation	NMES
CGNB	Central Government Net Borrowing	-NMFJ
CGP	CG procurement expenditure	QWPT
CGSB	CG net social benefits to households	GZSJ
CGT	Capital Gains tax (paid by HH)	QYJX
CGTSUB	CG total subsidies: products & production	NMCD
CGWADJ	Time varying coefficient for central government wages & salaries	No Codes
CGWS	CG compensation of employees	QWPS
COC	Tax-adjusted real cost of capital	No Codes
CONS	HH (&NPISH) final consumption expenditure	ABJR+HAYO
CPI	Consumer Prices Index	D7BT
CT	Corporation tax (gross of tax credits)	CPRN
CTC	Child tax credit	-MDYL
DEP	Public Sector Depreciation	JW2S
DINV	Change in inventories (CVM)	CAFU
DIPHH	Total interest payments of HH (&NPISH)	ROYU
EARN	Whole economy average earnings (NA measure)	(DTWM-DTWP)/(MGRZ-MGRQ)
ECG	Central Government employment	G6NQ
EECOMPC	Employees compensation from abroad	IJAH
EECOMPD	Employees compensation due abroad	IJAI
EENIC	Employees' (& self-employed) payments of NICs	AIIH-CEAN
EESC	Employees' social contributions	L8PS+L8Q8+L8LU
ELA	Local Authority employment	G6NT
EMPNIC	Employers' payments of NICs	CEAN
EMPSC	Employers' social contributions	DTWP
EMS	Market sector employment (LFS)	MGRZ-G6NQ-G6NT-MGRT-MGRW
EPS	Private sector employment (WFJ)	No Codes
ER	16+ employment rate	MGSR
ERCG	CG average earnings index	(NMAI*1000/C9K9)*(4/52)

Mnemonic	Definition	ONS code
ERLA	LA average earnings index	(NMJF*1000/ C9KA)*(4/ 52)
ES	Employers & self-employed (WFJ)	DYZN
ESLFS	Employers & self-employed (LFS)	MGRQ
ET	Employed labour force (WFJ)	No Codes
ETLFS	Total LFS employment inc. self-employed	MGRZ
FISIMPS	Total nominal FISIM	IE9R
FSMADJ	FISIM adjsutment in HH disposable income	No Codes
FYCPR	Gross trading profits of all private companies	CAED+CAGD+ RITQ
FYEMP	Compensation of employees	DTWM
GAD	ONS population projection: total	No Codes
GDPM	Gross Domestic Product at market prices, CVM	ABMI
GGFCD	General Govt final consumption deflator	100*NMRP/ NMRY
GGI	General Government GFCF (CVM)	DLWF
GGNB	General Govt Net Borrowing (NSA)	-NNBK
GPW	Gross physical wealth of HH&NPISH	No Codes
HHDI	Household disposable income	RPHQ
HHISC	Household imputed social contributions	L8RF
HHSB	Household social benefits	QWMZ
HWA	Total hours worked	YBUS
I4	January base year indices for RPI components MIPS	DOBQ
I7	January base year indices for consumer prices RPI	CHAW
I9	January base year indices for consumer prices RPIX	CHMK
IBUS	Business investment	NPEL
IBUSX	Business investment ex. BNFL transfer to CG	GAN8
IF	Total gross fixed capital formation (CVM)	NPQT
IH	Private sector investment in dwellings	L636
INCTAC	Income tax accruals adjustment	CYNX+RUTC+ DKHE+DBKE+ KIY5
INCTAXG	Income tax gross of tax credits	LIPG
INHT	Inheritance tax	ACCH+LSON
IPRL	Private sector transfer costs on non-produced non-fin. assets	L637
IROO	Working variable for OSHH	No Codes
KGAP	Gap between capital stock and optimal level of capital	No Codes
KMSXH	Market sector capital excluding housing	No Codes
KSTAR	Optimal capital	No Codes
LAIPS	LA gross fixed capital formation	NMOA
LANB	Local Authority Net Borrowing	-NMOE
LAPR	LA procurement expenditure	QWRZ-NMKK
LASBHH	LA net social benefits to HH	GZSK
LATSUB	LA total subsidies: products & production	ADAK
LAWADJ	Time varying coefficient for LA wages & salaries	No Codes
LAWS	LA compensation of employees	QWRY
LFSUR	LFS unemployment rate	MGSX
M	Total imports (CVM)	IKBL
MC	Household consumption weighted by its import content	No Codes
MCGG	Government consumption weighted by its import content	No Codes

Mnemonic	Definition	ONS code
MDINV	Inventories (excl. alignment adjustment) weighted by its import content	No Codes
MGTFE	index of final demand weighted by import intensity (goods)	0.176*(ABJR+HAYO)+0.064*NMRY+0.175*NPQT+0.094*CAFU+0.410*(BQKQ - BOXX) + 0.049*IKBE
MI	Mixed income	RNKX
MIF	GFCF weighted by its import content	No Codes
MNOG	Imports of non-oil goods CVM	BQKO-BPIX
MOIL	Imports of oil	BPIX
MPS	Total imports (CP)	IKBI
MS	Imports of services (CVM)	IKBF
MSGVA	Market sector GVA, Â£CVM	No Codes
MSTFE	index of final demand weighted by import intensity (services)	0.081*(ABJR+HAYO)+0.030*NMRY+0.059*NPQT+0.012*CAFU+0.029*(BQKQ - BOXX) + 0.093*IKBE
MTFE	Import-weighted domestic demand	No Codes
MXG	Goods exports weighted by its import content	No Codes
MXS	Services exports weighted by its import content	No Codes
NFWPE	HH net financial assets (NSA)	NZEA
NMTRHH	Net misc. transfer receipts of HH (&NPISH)	RPHO-RPID
NPISHTC	NPISH tax credits	-CFGW
OCT	Other current taxes: rec'd by CG	NMCV-CQOQ-KIH3
OS	Whole economy Gross Operating Surplus	ABNG
OSGG	General Govt Gross Operating Surplus	NTAR
OSHH	Household & NPISH Gross Operating Surplus	CAEN
OSPC	Gross Operating Surplus of Public Corporations	CAEQ
PART16	16+ activity rate	MGWG
PCE	Consumers' expenditure deflator	100*(ABJQ+HAYE)/(ABJR+HAYO)
PCIH	Public Corporation investment in dwellings	L634
PCLEB	PC investment in existing buildings & transfer costs	L635
PCNB	Public Corporations Net Borrowing (NSA)	-CPCM
PD	Property transactions	No Codes
PGDP	Gross Domestic Product deflator	YBGB
PIPHH	Property income paid by HH (&NPISH)	ROYT
PIRHH	Property income rec'd by HH (&NPISH)	ROYL
PMGREL	Deflator for goods import-weighted TFE	No Codes

Mnemonic	Definition	ONS code
PMNOG	AVI of imports of non-oil goods	100*(BOKH-ENXO)/(BQKO-BPIX)
PMOIL	Price index for imports of oil	(ENXO/BPIX)*100
PMS	AVI of imports of services	100*(IKBC/IKBF)
PMSGVA	Market Sector GVA deflator	No Codes
PMSREL	Deflator for services import-weighted TFE	No Codes
POP16	Population of 16+ (LFS)	MGSL
PPIY	Producer output Price Index ex. taxes	GB7S
PR	Retail Prices Index (RPI) level	CHAW
PRENT	Housing: Rent RPI	DOBP
PRMIP	Housing: Mortgage Interest Payments RPI	DOBQ
PRODH	Non-oil productivity per hour	No Codes
PRXMIP	RPI excluding Mortgage Interest Payments	CHMK
PSAVEI	Private sector average earnings index (inc. bonus)	KAC4
PSCB	Public Sector Current Budget	JW2T
PSCE	Public Sector Current Expenditure	JW2Q
PSCR	Public Sector Current Receipts	JW2O
PSGI	Public Sector Gross Investment	No Codes
PSNBCY	Public Sector Net Borrowing (CYSA)	-RQBN-RPZD
PSNBNSA	Public Sector Net Borrowing (NSA)	-J5II
PSNCR	Public Sector Net Cash Requirement	JW38
PSND	Public Sector Net Debt	HF6W
PSNI	Public Sector Net Investment	-JW2Z
PUBSTIW	Public Sector taxes on Income & Wealth	ANSO
PUBSTPD	Public Sector taxes on Production (& products)	NMYE
PXNOG	AVI of exports of non-oil goods	100*(BOKG-ELBL)/(BQKQ-BOXX)
PXOIL	Price index for exports of oil	(ELBL/BOXX)*100
PXS	AVI of exports of services	100*(IKBB/IKBE)
R	Bank rate	No Codes
RCGIM	CG non-trading capital consumption	NSRN
RENTCO	Private sector companies rental income	DTWR+DTWS
RHHDI	Real household disposable income	NRJR
RLAIM	LA non-trading capital consumption	NSRO
RPI	Retail Prices Index (RPI) inflation	No Codes
RPRICE	Relative export prices	No Codes
RX	Sterling effective exchange rate	BK67
RXD	Sterling-dollar cross rate: $/Â£	AUSS
SA	Stock appreciation	DLRA+EQCB
SBHH	Household social benefits	RPHL
SDE	Statistical Discrepancy: GDP(E)	GIXS
SPECX	Trend specialisation in world trade & ind. production	No Codes
TAF	Tax-adjustment factor for private sector	No Codes
TB	Balance of trade in goods & services	IKBJ
TCINV	Other company taxes on investment	GRXE

Mnemonic	Definition	ONS code
TCPRO	Corporation tax rate	No Codes
TME	Total Managed Expenditure	KX5Q
TSEOP	Taxes on self-employment incomes	ZAFG
TYEM	Taxes on income from employment	DBBO
TYWHH	Household current taxes on income & wealth	RPHS+RPHT
ULFS	LFS unemployment (ILO)	MGSC
VAL	Net acquisitions of valuables (CVM)	NPJR
VREC	Net VAT receipts	EYOO
WFJ	Employed labour force (WFJ)	DYDC
WFP	Wages & salaries inc. benefits in kind	DTWM-DTWP
WPG	World Price of Goods	No Codes
WRGTP	Work related govt training programmes	LOJU
X	Total exports, CVM	IKBK
XNOG	Exports of non-oil goods, CVM	BQKQ-BOXX
XOIL	Exports of oil	BOXX
XPS	Total exports (CP)	IKBH
XS	Exports of services, CVM	IKBE

B The complete free-running scorecard

Table 8 reproduces the complete 21-variable free-running scorecard summarised in Table 2, organised by model block, with each variable's metric and passthrough status made explicit. The de-seeding procedure and band definitions are those of Section 5.2: rates within 1.0pp, net balances within 1.5 per cent of GDP, levels within 10 per cent MAPE.

C Solver and convergence details

Iteration scheme. Each quarter is solved by Gauss–Seidel iteration over the full equation list in listing order: every left-hand variable is updated in place from the current values of all others, and a sweep of all 372 equations constitutes one iteration. Convergence is declared when the maximum relative change across all variables in a sweep falls below 10^{-6}, where the relative change of an update from y^{old} to y^{new} is $|y^{new} - y^{old}| / \max(|y^{old}|, |y^{new}|, 10^{-6})$—symmetric in old and new magnitudes, with an absolute floor so near-zero variables (rates, balances) still register movement without inflating the criterion. The iteration cap is 60 sweeps per quarter.

Stall detection. A minority of equations oscillate or overflow from unit inconsistencies in the public databank and would burn the full iteration budget without converging. The solver therefore tracks the best (smallest) maximum-change achieved and exits with a recorded `stall` status if eight consecutive sweeps fail to improve on it; the exit status of every period (`tol`, `stall`, or `max_iter`) is retained for diagnosis, and equations that raise during evaluation are counted per variable rather than silently dropped, so dead equations are visible in the solve report. The de-seeded scorecard of Section 5.2 exists precisely because a stall exit near the EFO seed would otherwise flatter the raw score.

Sequential solve and typical cost. Quarters solve sequentially forward (lags reference solved history; the few lead references use the seeded path). A twelve-quarter anchored solve of the

Table 8: Complete free-running scorecard by block, raw de-seeded solve vs EFO, 2025Q1–2027Q4

Block	Variable	Code	Metric	Error	Status
Demand	Real GDP	$GDPM$	MAPE	5.70%	computed — fair
	Consumption	$CONS$	MAPE	9.45%	computed — fair
	Total investment	IF	MAPE	0.00%	passthrough
	Business investment	$IBUS$	MAPE	15.48%	computed — poor
	Exports	X	MAPE	0.00%	passthrough
	Imports	M	MAPE	0.00%	passthrough
Labour market	Employment	$ETLFS$	MAPE	0.00%	trivial identity
	Employees	$EEES$	MAPE	0.00%	passthrough
	Unemployment rate	$LFSUR$	pp	1.01	computed — poor
	Average earnings	$AWEI$	MAPE	0.00%	passthrough
Prices & wages	CPI index	CPI	MAPE	0.00%	passthrough
	RPI inflation	RPI	pp	2.03	computed — poor
	CPI inflation	$CPIGR$	pp	0.00	passthrough
	GDP deflator	$PGDP$	MAPE	0.00%	passthrough
	Wages & salaries	WFP	MAPE	0.00%	passthrough
Incomes	Household income	$HHDI$	MAPE	14.76%	computed — poor
	Compensation	$COMP$	MAPE	0.00%	passthrough
	Real household income	$RHHDI$	MAPE	14.48%	computed — poor
	Company profits	$FYCPR$	MAPE	54.57%	computed — off
External	Current account	CB	% of GDP	2.76	computed — poor
	Trade balance	TB	% of GDP	0.62	computed — fair

Notes: "Passthrough" means the variable's equation is dead or exogenous in this configuration and the series is held at the OBR's published value, so a 0.00 error carries no information about calibration. The employment identity $ETLFS = 1000\,HWA/AVH$ is computed but trivial, both inputs being passthrough. RPI inflation is scored as a rate against the EFO's $RPIGR$ series (scoring the modelled inflation *rate* against the RPI *index* is a category error the scorecard previously made, reading a meaningless 99.6 per cent). Net balances are scored as a per cent of GDP, the OBR's own convention for differences of large gross flows. Source: `docs/calibration_scorecard.md` and `obr_macro/calibration_score.py` in the model repository.

full system takes roughly 30–40 seconds on a laptop (single-threaded CPython; equations are compiled once and evaluated against a per-iteration variable dictionary); a reform run is two such solves on structurally identical clones.

Add-factor application. Anchoring residuals (Section 3.11) are added to behavioural equations' solutions in baseline mode and disabled in shock mode, so that shocked runs deviate from an identically structured baseline. A second, separate channel—*structural* add-factors—is applied in baseline and shocked runs alike and therefore cancels in any base-versus-shock delta; the investment closure uses it to hold the reconstructed $\Delta \ln IBUSX$ baseline on the OBR's published path without contaminating the corporation-tax response.

The investment closure. Corporation-tax experiments activate the reconstructed business-investment ECM (5) (published commented out, with one unbalanced parenthesis restored) in place of the $IBUSX$ residual identity, add the IF and $IBUS$ aggregation identities, and stabilise the result in two moves: market-sector output $MSGVA$ is frozen at a reference path from a tracking solve in which $IBUSX$ is pinned to its published values (breaking an accelerator loop whose closed-loop eigenvalue exceeds one on this databank—the raw loop runs $IBUSX$ from ∼£94,000m to ∼£7,000,000m in twelve quarters), and the equation's level is re-centred on the OBR path by structural add-factors. Because $KSTAR$ still responds to the cost of capital, the $TCPRO$ channel survives intact. Investment-closure solves cap Gauss–Seidel at 25 sweeps: the only variable still moving beyond that point is a rest-of-world financial-accounts series off the corporation-tax chain, and the investment response itself is settled to within ∼1 per cent by sweep 25.

Reproducibility. All figures and experiment tables in this paper are generated by the scripts in the paper's `figures/` directory (`fig_anchored.py`, `fig_free_running.py`, `fig_reform.py`, `run_experiments.py`, `make_glossary.py`), which run against the installed `obr_macro` package and write both the PDF figures and CSV audit files alongside themselves.