

Identification, Estimation and Misspecification and Their Effects on Forecasting and Policy

Wage–Price Pass-Through in a Dutch Agent-Based Model

Steven Hoekstra ^{*1,2}

¹*CeNDEF, University of Amsterdam, Amsterdam, Netherlands*

²*Tinbergen Institute, Amsterdam, Netherlands*

16 October 2026

Working paper

Abstract

Can estimation distinguish a change in cost pass-through from compensation for a wrong wage-setting rule? We study this question in a Dutch agent-based economy built on CANVAS, with a common markup pricing rule and two wage rules: price indexation and error correction. A known-truth experiment with 100 independent panels per cell crosses the generating and fitted wage rules with the presence or absence of a pass-through break, and measures coefficient bias, the error in a paired 5% wage-shock response, and whether fit, forecast and parameter-constancy warnings flag large response errors. When the indexation model is fitted to error-correction data, the estimated break in the cost coefficient is nearly unbiased, but its level is biased by +0.35 to +0.39 and the twelve-quarter consumer-price response is overstated by about +1.09 percentage points, against at most +0.06 in the correct-rule control. Fit and forecast warnings flag every such panel at an achieved control alarm rate of 0.10; adding parameter constancy does not improve on them at that rate. In Dutch data both wage models are atypical of the observed panel before and after estimation; before estimation, the episode's consumer-price increase exceeds 98% and 92% of the two models' prior paths, and prices depart from the prior well before wages do. The estimated cost-surprise coefficient is near zero under both, and the implied twelve-quarter consumer-price responses to the wage reset are close to zero, against about 5.5–5.7 percentage points with all coefficients fixed at one. Because the same estimator recovers the coefficient in model-generated data and the omitted wage rule biases it upward, we read the near-zero Dutch estimate as a symptom of a misspecified price block rather than as a measurement of pass-through; the Dutch responses are not suitable for policy analysis. The error-correction estimator fails simulation-based calibration, so its results are exploratory throughout.

Keywords: wage-price spiral; pass-through; Phillips curve; agent-based model; misspecification; Bayesian estimation.

JEL codes: C11, C15, C52, C53, E31, E37.

*Corresponding author: s.g.hoekstra@uva.nl

Research status. Results from the six-coefficient indexation estimator rest on an estimator that passes its simulation-based calibration screen; results for the error-correction model M_1 use the error-correction estimator, which fails that screen, and are exploratory. The Dutch estimates indicate that the price block is misspecified for Dutch data (Section 7.6); the Dutch wage-shock responses are reported as diagnostics, not as policy estimates.

1 Introduction

How much of a common wage increase reaches consumer prices? An estimated macroeconomic model can answer this question precisely and still answer it incorrectly. Its price equation must explain observed inflation using the wage, demand and expectation signals supplied by the rest of the model. If wage setting is wrong, a fitted cost coefficient may compensate for that error rather than measure the transmission of a common wage shock. A change in the fitted coefficient may then reflect a change in the omitted mechanism, and a stable coefficient may conceal a persistent transmission error.

The distinction matters in inflation episodes in which wages catch up with earlier price increases. Accounts of pandemic inflation distinguish product-market disturbances from subsequent wage adjustment (Bernanke and Blanchard, 2023). Historical wage–price accelerations do not uniformly develop into sustained spirals (Alvarez et al., 2024). Models used to assess those risks need a defensible account of both the initial cost increase and the wage response to the resulting inflation.

This paper asks whether estimation can distinguish a genuine change in cost pass-through from compensation for an omitted wage-setting mechanism. A second, connected question is whether model fit, out-of-sample forecasts and parameter constancy warn when a fitted model gives an economically misleading wage-shock response. Forecasting and identification are therefore parts of one transmission-reliability question.

We use a Dutch economy built on the stock-flow-consistent agent-based CANVAS model (Hommes et al., 2025; Poledna et al., 2023). Firms in 62 industries interact through an empirical production network, employ workers and trade with households, government and the rest of the world. Wages affect unit costs and household demand; prices feed into subsequent wage adjustment. The baseline retains the CANVAS price–quantity allocation rule but uses an explicitly specified markup pricing rule. This operational baseline is not presented as the published CANVAS pricing equation.

The main controlled alternative changes only the aggregate wage equation. Baseline wages respond to positive output-price inflation through an indexation coefficient. The alternative uses error correction: wage growth responds to past consumer-price inflation, wage-growth persistence, unemployment and a real-wage gap (Blanchard and Katz, 1999). Pricing, quantity setting, production, matching and expectations are held common in this comparison. Broader alternatives with productivity growth and search frictions are diagnostic extensions, not a replacement for this one-block experiment.

The transmission object is additional cumulative consumer-price inflation twelve quarters after a once-only 5% rise in every firm’s base wage. It is our policy example: a stylised economy-wide wage settlement, such as a coordinated collective-agreement round. Shocked and unshocked simulations start from the same state and use paired random streams. In a synthetic economy the generating response is known, so response errors and their coverage can be measured. In the Dutch application that truth is unobserved. Differences between defensible specifications therefore measure sensitivity, not which model gives the correct causal effect of a real-world intervention.

Our framework separates three failures that are easily conflated. An inaccurate neural posterior is an inference failure. An observed path that is atypical under a fitted predictive distribution

is a predictive discrepancy. Model-family misspecification is the stronger statement that no admissible parameter vector reproduces the relevant conditional distribution. Simulation-based calibration addresses the first issue, predictive checks the second, and a maximized conditional Monte Carlo test targets the third. None alone identifies the true wage mechanism.

The known-truth experiment gives the main answer. Each of its eight cells has 100 independent panels, so recovery bias, interval coverage and response errors are frequencies rather than examples. When the indexation model is fitted to error-correction data, the estimated break in the cost-surprise coefficient remains close to the truth: its bias is -0.03 against a true change of 0.25 , compared with $+0.02$ in the correct-rule control. The level of the coefficient is not: it is biased by $+0.35$ to $+0.39$, and its 90% intervals cover the truth in 0.15 – 0.33 of panels. Its twelve-quarter response to the common wage reset overstates the generating response by $+1.04$ to $+1.10$ percentage points, with a noise-corrected squared error of 1.168 – 1.352 , against at most $+0.06$ and 0.084 in the control. The omission therefore biases transmission and the level of the coefficient, but not the estimated break.

Standard warnings are only partly informative about these errors. Calibrated on correct-rule panels and evaluated on independent ones, a warning based on joint fit and fresh forecast losses flags every wrong-rule panel of the indexation model at an achieved control alarm rate of 0.10 ; adding a posterior probability of a rise in ϕ_{CP} leaves detection unchanged at that rate, and for the exploratory error-correction fits raises detection of a genuine break from 0.04 to 0.58 . Because the wrong-rule panels are flagged almost without exception, the unflagged panels carry only the much smaller correct-rule errors.

The Dutch application shows the same tension in observed data. Before estimation, both wage models are already atypical of the observed panel under their priors: at the episode origin 2021Q3, observed twelve-quarter consumer-price growth exceeds 98% of the indexation model's prior paths, and prices leave the prior's central band well before wages do. The estimated indexation model passes its calibration check and puts the cost-surprise coefficient near 0.01 – 0.02 ; the error-correction model gives 0.084 and 0.059 in the two windows. Both imply mean twelve-quarter CPI responses to the common wage reset close to zero, against about 5.5 – 5.7 percentage points with all coefficients fixed at one. After estimation their joint predictive paths remain atypical of the observed panel, and their wage and episode-price intervals have low empirical coverage.

We therefore read the Dutch estimates as a symptom of misspecification, not as a measurement of pass-through. Three facts support that reading. First, the same estimator recovers the cost-surprise coefficient in data generated by the model, with a bias of -0.01 and 90% coverage of 0.94 at a true value of 0.25 , so a near-zero Dutch estimate is not an artefact of the estimator. Second, the wage-rule omission studied in the experiment biases the coefficient upward, not toward zero, so it cannot explain the Dutch value either. Third, prior and posterior predictive checks fail for both wage rules, and the clearest failure is an episode price increase faster than the model generates under almost any prior coefficient values. The most plausible source, which we have not isolated, is the price block: it does not reproduce the episode's price dynamics with any coefficient value we estimated; a near-zero cost coefficient is then where the fit settles, not a description of how Dutch firms pass on wage costs. The Dutch wage-shock responses reported here should not be used for policy analysis. The paper's contribution is the diagnostic framework and its operating characteristics in known-truth data; the Dutch application shows those diagnostics flagging a model that is not yet fit for that purpose.

The error-correction estimator fails simulation-based calibration on its wage block. We report

its results throughout, marked M_1 , because the cross-model comparison is the question the paper asks; its intervals are not read as calibrated structural uncertainty. The headline synthetic conclusions rest on the cells fitted with the validated indexation estimator.

Section 2 places the question in the literature. Section 3 describes the model and Section 4 the data and observation design. Section 5 defines the estimands, the inference and the predictive checks. Section 6 reports the known-truth experiment and Section 7 the Dutch evidence. Section 8 concludes.

2 Related literature

Accounts of pandemic inflation assign different roles to the margins studied here. The sellers' inflation view emphasises firms' ability to widen margins during a common cost shock (Weber and Wasner, 2023), whereas Bernanke and Blanchard (2023) attribute the initial US surge mainly to product-market shocks and a later, more modest role to labour-market pressure. Their separation of price pressure and wage adjustment motivates the mechanism studied here. The error-correction alternative lets wages catch up with past inflation and respond to labour-market slack. Historical evidence shows that most wage–price accelerations subside without a sustained spiral (Alvarez et al., 2024). Error-correction wage equations, in which real wages return to a level consistent with productivity and the labour share, reconcile the wage curve with the Phillips curve (Blanchard and Katz, 1999), and coordinated bargaining institutions such as the Dutch system shape how quickly that correction occurs (Visser, 2016).

Production networks change how these forces reach aggregate inflation. Intermediate-input shares flatten network Phillips curves and generate heterogeneous sectoral responses to common shocks (Rubbo, 2023); euro-area gas shocks pass through differently to producer prices, wages and core inflation (Adolfson et al., 2024); and firm-level price adjustment is itself heterogeneous (Nakamura and Steinsson, 2008). Survey evidence shows that cost pass-through rose as broad cost pressures intensified in 2021–2022 (Asghar et al., 2023), and that lived experience of inflation shapes expectations (Salle et al., 2023). Pass-through is therefore not a constant of nature, and a model that cannot detect a change in it is of limited use in an episode such as 2021–2023.

Agent-based macroeconomic models represent propagation through distinct firms, balance sheets, input suppliers and bilateral relationships in goods, labour and credit markets (Axtell and Farmer, 2025). The Keynes–Schumpeter family generates endogenous cycles, fat-tailed firm growth and persistent unemployment (Dosi et al., 2010, 2015), and motivates the investment-embodied productivity channel of the broader diagnostic extension. EURACE combines spatial, credit and labour frictions (Dawid and Delli Gatti, 2018; Dawid et al., 2012), and Dawid et al. (2024) show that the form of firms' pricing rules determines how cost shocks reach inflation. Stock-flow-consistent agent-based models reproduce many empirical regularities (Caiani et al., 2016). By calibrating directly on national accounts, Poledna et al. (2023) show that an agent-based model can forecast as well as equilibrium and statistical models out of sample; the approach has been extended across OECD economies (Wiese et al., 2024), adopted in central banks (Borsos et al., 2025) and applied around the COVID break (Hommes and Poledna, 2023, 2026). CANVAS adds richer price and quantity rules and household heterogeneity and is used for policy analysis at the Bank of Canada (Hommes et al., 2025, 2026). In this literature the coefficients governing pass-through, wage adjustment and expectations are usually calibrated, often at one, even though they determine aggregate dynamics.

Latent agent states and decentralised matching make the likelihood of a large agent-based model intractable. Approximate Bayesian methods show that economic agent-based models can instead be estimated from simulated output (Grazzini et al., 2017; Delli Gatti and Grazzini, 2020; Platt, 2020). Simulation-based inference learns from parameter–data pairs without evaluating the likelihood (Cranmer et al., 2020); neural methods replace the tolerance and distance choices of approximate Bayesian computation with a learned conditional density or density ratio (Lueckmann et al., 2021; Dyer et al., 2024; Hermans et al., 2020). Such posteriors can be overconfident (Hermans et al., 2022; Delaunoy et al., 2022), so their calibration has to be checked before they are read (Talts et al., 2018).

A posterior can be well calibrated and still belong to the wrong model, and here the econometric literature on forecasting under misspecification is directly relevant. When the data-generating process differs from the forecasting model, estimation delivers pseudo-true values that make the misspecified model fit as well as it can, and forecast failure need not point to the coefficient that is wrong (Clements and Hendry, 1998). The Lucas critique concerns a different question, the invariance of coefficients to changes in policy (Lucas, 1976); no policy rule changes in our experiments. The known-truth design measures how a specified wage-rule omission propagates into estimated cost transmission and the implied wage-shock response. It separately evaluates whether predictive warnings and parameter constancy flag large response errors.

3 Model

CANVAS is a large-scale stock-flow-consistent agent-based model extending Poledna et al. (2023), recalibrated here for the Netherlands (Hommes et al., 2025). This section describes the blocks that matter for wage–price transmission: production and costs, price and quantity setting, wage setting and the labour market. The controlled models M_0 and M_1 differ only in wage setting. Diagnostic variants with search frictions, productivity growth and labour-force growth are described in Appendix B. Appendix A gives the remaining specification.

3.1 Overview

The economy comprises six institutional sectors: households, firms, a commercial banking sector, a central bank, government and the rest of the world. Firms operate in 62 industries classified according to NACE Rev. 2 and are linked through an empirical input–output network. Across the dated origin states used here, the economy contains several hundred representative-firm agents and several thousand worker agents in a simulation scaled to 0.001 of the national population (Appendix C).

Goods, labour and credit are exchanged through decentralised search-and-matching protocols in which agents form bilateral relationships, so identical firms facing identical aggregate shocks may set different prices. Accounting is stock-flow consistent and aligned with the European System of Accounts: every monetary flow has a balance-sheet counterpart. The production network transmits cost shocks across sectors through intermediate-goods trade, generating heterogeneity in firm-level inflation even under a common shock.

The model evolves in quarterly steps. Within each period, agents resolve insolvencies, update expectations, set prices and labour demand, trade in the credit, labour and goods markets, update aggregate price indices and the wage index, service debt, and pay taxes and dividends.

The government collects revenues and pays transfers according to Dutch fiscal rules, and the central bank sets the policy rate by a Taylor rule calibrated to the ECB's reaction function. The sequence matters for the mechanism under study: firms post prices before current goods-market matching, whereas the wage index responds to inflation realised in that round and enters firms' costs in the next quarter.

3.2 Production and costs

Each firm i in sector s operates a Leontief production function,

$$Y_i = \min(Q_i^s, N_i \alpha_i, K_i \kappa_i, M_i \beta_i), \quad (1)$$

where Q_i^s is planned production, N_i employment, α_i effective labour productivity, K_i the capital stock, κ_i capital productivity, M_i intermediate inputs and β_i materials productivity. Effective productivity is the firm's productivity anchor $\bar{\alpha}_i$ adjusted for capacity utilisation, and the firm wage w_i is its wage anchor $\bar{w}_i$ adjusted by the same factor (Appendix A).

Unit cost decomposes into labour, materials and capital components,

$$AC_i = \underbrace{(1 + \tau^{SIF}) \frac{\bar{w}_i}{\bar{\alpha}_i} P^W}_{AC_i^L} + \underbrace{\frac{1}{\beta_i} \sum_g a_{sg} \bar{P}_g}_{AC_i^M} + \underbrace{\frac{\delta_i}{\kappa_i} P^{CF}}_{AC_i^K}, \quad (2)$$

where τ^{SIF} is the employer social-insurance rate, P^W the aggregate wage index, a_{sg} the input-output coefficient from sector g to sector s , $\bar{P}_g$ the price index of sector g , δ_i the depreciation rate and P^{CF} the price of capital goods. Unit labour cost is built on the anchors rather than on utilisation-adjusted values, so it does not move with utilisation within the quarter. Equation (2) is the direct channel from wages to costs: the wage block sets P^W , which scales unit labour cost in every firm, and the pricing rule marks up over the result. Wages also affect prices indirectly through household demand.

3.3 Price and quantity setting

CANVAS routes a firm's demand signal either to its price or to its planned quantity. At the start of period t , firm i observes lagged demand $Q_i^d(t-1)$, goods offered $Q_i^o(t-1) = Y_i(t-1) + S_i(t-1)$ including inventories S_i , its own price $P_i(t-1)$ and the average price $\bar{P}_g(t-1)$ of its sector. Figure 1 combines excess demand or supply with a high or low relative price. A cheap firm facing excess demand raises its price, while an expensive firm facing excess demand raises planned production; under excess supply, a relatively expensive firm lowers its price and a relatively cheap firm lowers production.

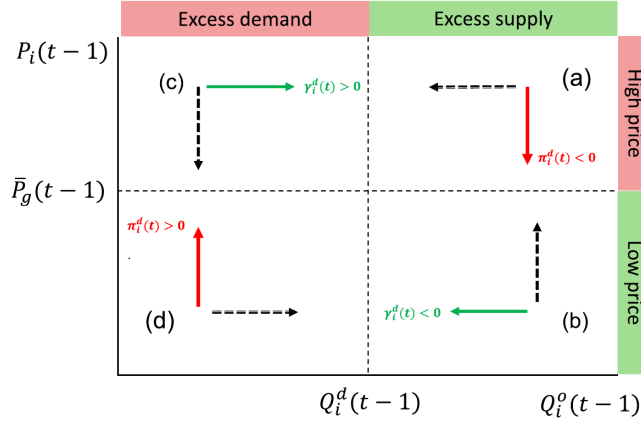


Figure 1: CANVAS four-quadrant price-quantity rule. Cases (a) and (d) route the demand imbalance to prices; cases (b) and (c) route it to planned quantities. Reproduced from Hommes et al. (2025, Figure 1) under CC BY 4.0.

The quantity signal is

$$\gamma_i^d(t) = \begin{cases} \frac{Q_i^d(t-1)}{Q_i^o(t-1)} - 1, & \begin{aligned} &Q_i^o(t-1) \leq Q_i^d(t-1) \text{ and } P_i(t-1) \geq \bar{P}_g(t-1), \\ &\text{or } Q_i^o(t-1) > Q_i^d(t-1) \text{ and } P_i(t-1) < \bar{P}_g(t-1), \end{aligned} \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

and the demand-pull price signal $\pi_i^{DP}(t)$ equals the same ratio in the two complementary cases and zero otherwise. Planned production follows

$$Q_i^s(t) = \max\left\{0, Q_i^o(t-1)(1 + \phi_{G_Q}\gamma^e(t))(1 + \phi_{DP_Q}\gamma_i^d(t)) - S_i(t-1)\right\}, \quad (4)$$

where $\gamma^e(t)$ is expected aggregate real growth, ϕ_{G_Q} scales that signal and ϕ_{DP_Q} the firm-specific demand signal. Subtracting opening inventories converts desired supply into current production. Because firms that adjust output must hire, the quantity coefficients reach inflation through labour demand and unemployment.

Prices are posted on decision-time unit costs, not on production costs realized after current goods-market matching. The calculation uses the current wage base and the latest posted input-price indices. Let cost growth relative to expected inflation be

$$s_{i,t} = \frac{AC_{i,t}}{AC_{i,t-1}(1 + \phi_{AE}\pi_t^e)} - 1. \quad (5)$$

The common markup rule is

$$1 + \mu_{i,t} = (1 + \mu_{i,t-1}) \frac{1 + \phi_{DP}\pi_{i,t}^{DP}}{1 + (1 - \phi_{CP})s_{i,t}}, \quad P_{i,t} = (1 + \mu_{i,t})AC_{i,t}. \quad (6)$$

The opening gross markup is reconciled as $P_{i,t-1}/AC_{i,t-1}$ after any firm-state reset. The denominator absorbs the part of the cost surprise not passed through. At $\phi_{CP} = 1$, it equals one and current costs pass through without multiplying prices by another expectation factor. At $\phi_{CP} = 0$, cost growth is absorbed so that the price factor is $(1 + \phi_{DP}\pi_{i,t}^{DP})(1 + \phi_{AE}\pi_t^e)$.

To first order, with $g_{i,t} = AC_{i,t}/AC_{i,t-1} - 1$,

$$\pi_{i,t} \approx \phi_{DP}\pi_{i,t}^{DP} + \phi_{CP}g_{i,t} + (1 - \phi_{CP})\phi_{AE}\pi_t^e. \quad (7)$$

Thus ϕ_{CP} governs cost-surprise transmission through the absorption denominator. It is not the same numeric coefficient as a multiplicative cost-pressure factor in the published rule. At full cost pass-through, expected inflation has no additional direct loading; it can still affect wages, quantities and subsequent costs through other rules.

Published CANVAS measures cost pressure using input-price indices relative to sector prices (Appendix A). A common change in the wage base does not directly change those reference-price ratios; its direct signal can therefore be small when the corresponding wedges are small. Homogeneity of those ratios does not prove that every common wage shock has no aggregate effect. Both controlled models in this paper use (6).

Expected inflation and expected real growth are formed by adaptive learning: each is an AR(1) in its own history, re-estimated every quarter on all available data (Appendix A). Both models use the same expectation rule. Their realized expectation paths may nevertheless differ because they learn from their own simulated histories.

3.4 Wage setting

The common wage index P^W scales every firm's anchor-based labour cost in (2). Measured compensation also varies with utilisation and employment composition. The controlled models differ in how the common index is updated.

In M_0 the wage index is indexed to output-price inflation,

$$P_{t+1}^W = (1 + \phi_{UNION} \max(\pi_t, 0)) P_t^W, \quad (8)$$

where $\phi_{UNION} \geq 0$ is the degree of indexation. With $\phi_{UNION} = 1$ wages fully track positive inflation, with $\phi_{UNION} < 1$ real wages absorb part of each price increase, and with $\phi_{UNION} > 1$ wages over-index. The maximum operator is a ratchet: indexation raises wages and never lowers them. Nothing else disciplines the wage level, so the real wage under (8) is a residual of inflation and ϕ_{UNION} rather than an outcome of the labour market. Unemployment and the labour share play no role, so this aggregate indexation rule has no direct unemployment term. An observed wage–unemployment relationship can still arise through utilisation and composition.

In M_1 wage growth follows an error-correction equation in the form of Blanchard and Katz (1999),

$$\Delta \log P_{t+1}^W = (1 - \beta_{IDX}) \Delta \log P_t^W + \beta_{IDX} \pi_t^{HH} - \delta_U (u_t - u^*) - \lambda_{RW} [\log(P_t^W / P_{t-1}^{VA}) - k], \quad (9)$$

where π^{HH} is consumer-price inflation in log form, u the unemployment rate, u^* the equilibrium rate and P^{VA} the value-added deflator. The coefficient β_{IDX} splits short-run wage growth between catch-up with past inflation and persistence of past wage growth; because the two weights sum to one, the equation is homogeneous in nominal terms. The slope δ_U is the wage Phillips curve: slack slows wage growth. The speed λ_{RW} pulls the wage relative to the output price back to the fixed real-wage target k . Unlike (8), the rule is two-sided and keys on a ratio of levels, so a period of high wages relative to prices restrains later wage growth. The equilibrium rate is pinned at $u^* = 0.052$ and the constant k is fixed (Appendix C).

3.5 Labour market

The controlled M_1 retains baseline matching and constant productivity anchors. In both models firms whose desired employment falls short of their workforce dismiss randomly selected workers, and firms with a shortfall post vacancies. In M_0 , vacancies and unemployed workers are matched at random until one side is exhausted, so every vacancy is filled whenever enough workers are unemployed. There is no search friction, and the labour market clears within the quarter.

3.6 What the difference means for transmission

In M_0 , the wage-base reset initially shifts unit labour costs; subsequent price indexation can feed inflation back into wages. In M_1 , the real-wage gap and unemployment can restrain later wage growth. Demand and employment responses also matter, so a sign or magnitude is not imposed on the aggregate response. The controlled experiment asks how these mechanisms change fitted transmission. Neither wage rule is presumed to be Dutch truth.

3.7 Estimated coefficients

M_0 estimates six coefficients,

$$\theta_0 = (\phi_{DP}, \phi_{CP}, \phi_{AE}, \phi_{DP_Q}, \phi_{G_Q}, \phi_{UNION}), \quad (10)$$

which the fixed empirical baseline sets at one. M_1 shares the five price and quantity coefficients and replaces indexation by three wage coefficients,

$$\theta_1 = (\phi_{DP}, \phi_{CP}, \phi_{AE}, \phi_{DP_Q}, \phi_{G_Q}, \beta_{IDX}, \lambda_{RW}, \delta_U). \quad (11)$$

Table 1 summarises the two models. Constants outside θ , among them u^* , the innovation step bound and the expectation rules, are fixed and listed in Appendix C.

Table 1: **The two models.** Blocks of the indexation model M_0 and the error-correction model M_1 ; only wage setting differs.

Block	M_0 (CANVAS)	M_1
Price setting	markup recursion (6)	same as M_0
Quantity setting	demand-gap rule (4)	same as M_0
Expectations	adaptive AR(1)	same as M_0
Wage setting	indexation (8)	error correction (9)
Labour market	random matching, no friction	same as M_0
Productivity	constant	same as M_0
Estimated coefficients	6	8

Notes: Fixed constants, including u^* and the expectation rule parameters, are listed in Appendix C.

4 Data and observation design

The Dutch economy is initialized from quarterly national accounts and the FIGARO input-output tables. Sector balance sheets, production technologies and wage bases are anchored to

the observed origin state rather than estimated as additional behavioral coefficients. Table 2 lists the sources. The application uses aggregate time series; existing sectoral initialization is not independent microeconomic validation of the behavioral rules.

Table 2: **Data sources.** External series used to initialise the model at each forecast origin, to inform estimation and to evaluate forecasts, with their frequency and the sample held in the calibration data.

Source	Variables	Frequency and sample
<i>Netherlands</i>		
Eurostat quarterly national accounts (namq_10_gdp)	Real GDP and gross value added; GDP, value-added and household-consumption deflators; household and government consumption, gross fixed capital formation, exports and imports; compensation of employees, wages and salaries, and operating surplus	Quarterly, seasonally and calendar adjusted; 1996Q1–2024Q4
Eurostat annual national accounts (nama_10_pe)	Total employment (persons), interpolated log-linearly to quarters	Annual; 1996–2024
Eurostat Labour Force Survey (une_rt_q)	Unemployment rate, ages 15–74	Quarterly, seasonally adjusted; 1996Q1–2024Q4
Eurostat FIGARO input-output tables (naio_10_fcp_ii)	62 × 62 inter-industry flows, final demand and value-added components by industry (NACE Rev. 2)	Annual; 2010–2023
Eurostat sector accounts, government finance, business and census statistics	Balance sheets of firms, households, banks and government; household income, transfers and taxes; fixed assets, employees and number of firms by industry	Quarterly and annual; 1997–2023
<i>Euro area</i>		
Eurostat quarterly national accounts (namq_10_gdp, EA19)	Real GDP and GDP deflator	Quarterly; 1996Q1–2024Q4
Eurostat money-market rates (irt_st_q)	Three-month Euribor	Quarterly; 1996Q1–2024Q4

Notes: All series come from one fully revised vintage. Samples are those held in the calibration data; initialisation at an origin uses observations up to that origin, and the windows described in this section select the estimation and evaluation targets. The euro-area series drive the rest-of-world block and the Taylor-rule parameters.

Seven quarterly series inform estimation: real gross value added, the household-consumption deflator, the value-added deflator, the GDP deflator, wages and salaries, total employment and unemployment. Wages and salaries (ESA D11, excluding employers’ social contributions) per person employed measure the observed wage rate, and real value added per person employed measures productivity. Employment counts all persons employed, including the self-employed, whereas the model’s workforce is calibrated to employees; a trend in self-employment therefore opens a wedge between observed and simulated wage-rate growth, which the forecast and predictive checks absorb as error. The household-consumption deflator is our CPI proxy, not a claim that it equals the harmonized consumer-price index.

Preprocessing preserves the declared origin and lead identities. Level paths are expressed relative to their origin; growth rates use log transformations and unemployment is in percentage points. Isolated accounting spike-and-reversal corrections belong to the observed-data preparation and are applied consistently across specifications. No burn-in or post-origin initialization adjustment is used to improve fit. Date lookup is checked against the observed lead-zero values.

The early estimation window uses six semi-annual origins from 2010Q1 through 2012Q3, with twelve-quarter targets ending before 2016. The pre-2020 window uses fourteen origins through 2016Q3, with targets ending before 2020. The corresponding evaluation periods are 2016–2019 (QOOS, quiet out of sample) and 2020–2024 (WOOS, wild out of sample). Forecast comparisons use only the eligible origin–lead cells of the declared target period; a twelve-quarter forecast need not contribute all its leads to a window’s score.

All observations come from a revised data vintage. These are retrospective forecasts, not a real-time exercise. Consecutive semi-annual origins share ten target quarters. Overlap is not treated as additional independent Dutch histories in hypothesis testing.

For neural estimation, one input is the assembled panel of origin-specific simulated paths and summaries. This is a conditional-restart observation design, not the likelihood of one continuously simulated Dutch history. Simulation-based calibration evaluates inference under that specified simulator design. It does not establish that the restarted joint observation law reproduces the dependence in the observed Dutch history. Conditional specification tests therefore state their origin-wise null separately and account for overlap when combining origins.

Five core variables enter forecast evaluation and joint predictive criticism: real value added, the CPI proxy, the wage rate, the value-added deflator and unemployment. Each forecast path retains the dependence across variables and all twelve quarters. Joint discrepancies, interval widths, coverage and forecast scores provide complementary evidence.

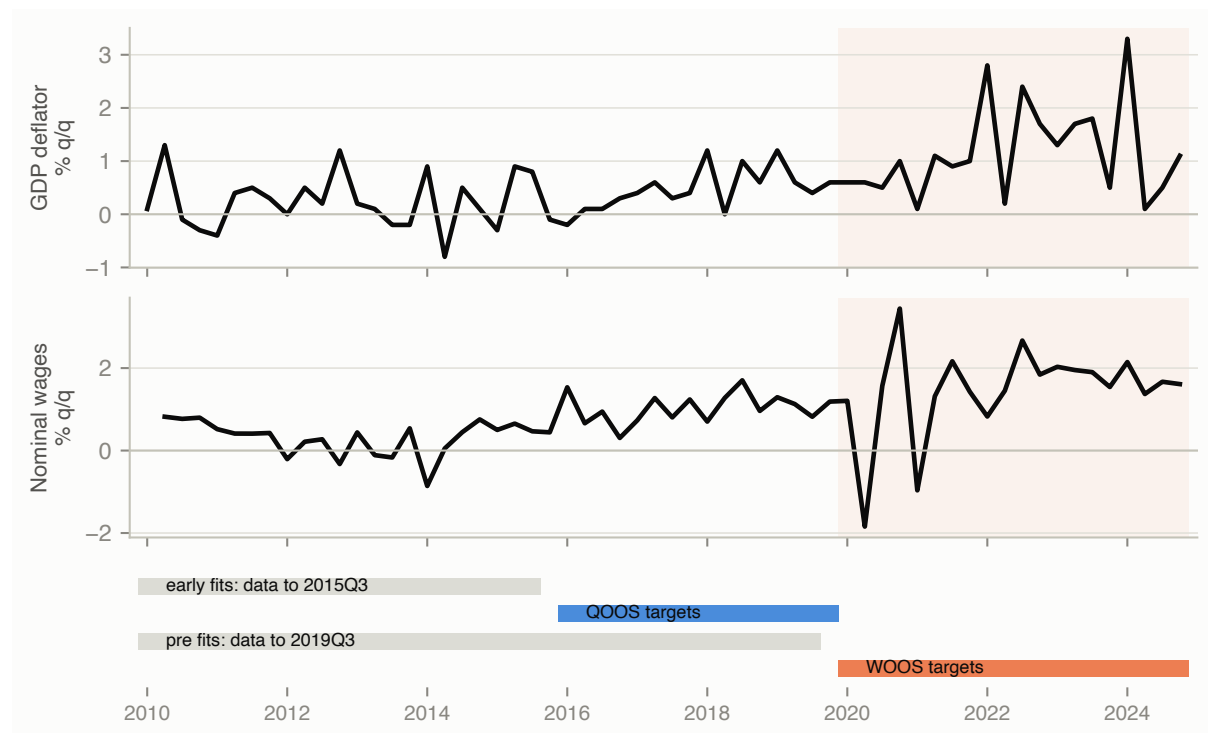


Figure 2: Quarterly GDP-deflator growth and nominal wage growth, 2010–2024. The shaded region marks the 2020–2024 episode; the bars show the early and pre estimation windows and the QOOS and WOOS evaluation windows. Source: Eurostat national accounts.

5 Estimands, inference and predictive criticism

5.1 The model family and the inferential target

Let Y_r denote a joint aggregate path in environment r , with true conditional law G_r . Model $m \in \{M_0, M_1\}$ generates $P_m(\cdot \mid \theta, \mathcal{I}_o)$, where $\mathcal{I}_o$ includes the declared origin information and initialization. Suppressing this conditioning when unambiguous, its observational family is

$$\mathcal{P}_m = \{P_m(\cdot \mid \theta) : \theta \in \Theta_m\}. \quad (12)$$

Observational misspecification means $G_r \notin \mathcal{P}_m$. A forecast failure at one calibration does not establish that family-wide statement. A parameter break can make a frozen pre-break forecast wrong even when the generating rule remains inside the same family.

The inference input is $X = S(Y)$, where S includes the retained paths, preprocessing and any frozen summary map. The exact model posterior and its neural approximation are different objects:

$$\Pi_m(d\theta \mid X_r) \propto p_{m,S}(X_r \mid \theta) \Pi_m(d\theta), \quad \hat{\Pi}_m(\cdot \mid X_r) \approx \Pi_m(\cdot \mid X_r). \quad (13)$$

Calibration of $\hat{\Pi}_m$ under simulations from P_m does not imply that G_r belongs to the family. Conversely, a failed calibration check prevents a neural interval from being used as evidence about structural uncertainty, even when the simulator contains the generating rule.

Where the divergence is finite and a minimizer exists, define the pseudo-true set

$$\Theta_{m,r}^* = \arg \min_{\theta \in \Theta_m} D_{\text{KL}}(G_{r,S} \parallel P_{m,S}(\cdot \mid \theta)). \quad (14)$$

Classical misspecified-likelihood theory and Bayesian misspecification theory motivate this distinction (White, 1982; Kleijn and van der Vaart, 2006). The projection is defined at the retained information resolution and need not equal a behavioral parameter. Misspecified simulation-based likelihood approximations also require care about their target and uncertainty (Frazier et al., 2024). Equation (14) is not a proved concentration result for our finite-sample neural estimator. We do not call a concentrated Dutch posterior pseudo-true merely because it is concentrated, or infer compatibility of the full path from compatibility of a few summaries.

5.2 A response can be wrong despite observational fit

For the once-only wage reset a , define the response at horizon h in log percentage points:

$$\tau_{m,h}(\theta) = 100 \mathbb{E} \left[\log \frac{\text{CPI}_{m,h}^a(\theta)}{\text{CPI}_{m,h}^0(\theta)} \right], \quad \hat{\tau}_{m,r,h} = \int \tau_{m,h}(\theta) \hat{\Pi}_m(d\theta \mid X_r). \quad (15)$$

The expectation is over paired simulation paths. Within-model response identification requires observationally equivalent parameters to give the same response:

$$P_{m,S}(\cdot \mid \theta_1) = P_{m,S}(\cdot \mid \theta_2) \implies \tau_{m,h}(\theta_1) = \tau_{m,h}(\theta_2). \quad (16)$$

Even this property does not equate responses across different economic rules that fit the available observations.

Let $\tau_{m,r,h}^\Pi$ integrate the response over the exact model posterior. For a selected $\theta_{m,r}^* \in \Theta_{m,r}^*$, adding and subtracting gives

$$\begin{aligned}\hat{\tau}_{m,r,h} - \tau_{r,h}^{\text{true}} &= (\hat{\tau}_{m,r,h} - \tau_{m,r,h}^\Pi) + (\tau_{m,r,h}^\Pi - \tau_{m,h}(\theta_{m,r}^*)) \\ &\quad + (\tau_{m,h}(\theta_{m,r}^*) - \tau_{r,h}^{\text{true}}).\end{aligned}\tag{17}$$

The terms distinguish neural approximation error, finite-data learning and prior effects, and transmission error at the observational projection. This algebraic identity does not separately identify those components from Dutch data. The known-truth experiment measures total response error and uses controlled comparisons to investigate its sources.

Likewise, a change in the shared cost coefficient can be written

$$\Delta\hat{\phi}_{CP} = \Delta\phi_{CP}^{\text{true}} + \Delta b_m + \Delta e_m,\tag{18}$$

where b_m is projection bias and e_m estimation error relative to that projection. Parameter movement may reflect a genuine break, changing compensation for the wrong rule, or estimation error. Stability can coexist with bias. Whether constancy improves warnings beyond fit and forecasting is measured in the replicated synthetic experiment (Section 6.5).

The framework supplies definitions and accounting identities, not a new asymptotic theorem. No global-identification, neural-consistency or misspecification-robust coverage result is claimed. Synthetic experiments reveal response errors because their generating laws are known. Dutch evidence can reveal incompatibilities and specification sensitivity, but not counterfactual truth.

5.3 Interpretive conventions

Four conventions apply throughout and are not repeated at every result. Dutch differences between specifications measure sensitivity, not which response is causally correct. Fitted predictive tail ranks are descriptive criticism, not frequentist p -values or model-family rejections. Monte Carlo standard errors are conditional on the stated origins, seeds and fitted posterior pools, not uncertainty across independent Dutch histories. Results for M_1 come from an estimator that fails simulation-based calibration and are exploratory.

5.4 Simulation-based Bayesian estimation

The simulator supplies parameter–data pairs although its likelihood cannot be evaluated. Neural posterior estimation learns a conditional density from these pairs (Cranmer et al., 2020; Dyer et al., 2024):

$$q_\psi(\theta \mid x) \approx p_{m,S}(\theta \mid x), \quad \theta \sim \Pi_m, \quad x \sim P_{m,S}(\cdot \mid \theta).\tag{19}$$

At the population training distribution, minimizing conditional negative log density minimizes the average divergence from the model posterior:

$$\mathbb{E}[-\log q_\psi(\theta \mid X)] = \mathbb{E}_X[D_{\text{KL}}(p_{m,S}(\theta \mid X) \parallel q_\psi(\theta \mid X))] + C.\tag{20}$$

This identity does not guarantee accurate inference at an atypical Dutch observation, finite-sample recovery or valid uncertainty under misspecification.

The training tables use 5,000 prior-box parameter points, ten stochastic realizations per point and 24 dated origin templates, each with a twelve-quarter horizon. Realizations enter separately, not as an averaged economy. A recurrent representation of the seven paths is combined with nineteen pooled wage–price statistics. These statistics are information inputs, not nineteen estimated parameters. Baseline M_0 has six coefficients and the controlled M_1 has eight.

All stochastic realizations of one parameter point stay in the same training, validation or held-out fold. The density is a box-supported spline flow (Durkan et al., 2019); posterior draws must satisfy the declared parameter bounds. Where used, an equal-weight five-member mixture shares the split and varies network initialization. An ensemble is not assumed to repair posterior calibration: it is assessed under the same acceptance rule.

5.5 Priors and the inference gate

Priors are independent uniforms on the intervals in Table 3. The five price and quantity response coefficients lie in $[0, 1]$. Indexation in M_0 admits values up to 1.25; the alternative admits partial wage catch-up, real-wage adjustment and a response to unemployment. These are declared model supports, not estimates of Dutch bargaining institutions.

Table 3: **Prior supports.** Each response coefficient has an independent uniform prior on the interval shown; a dash means the coefficient is not part of the model.

Coefficient	Role	M_0	M_1
ϕ_{DP}	Price response to firm excess demand	$[0, 1]$	$[0, 1]$
ϕ_{CP}	Pass-through of cost surprises	$[0, 1]$	$[0, 1]$
ϕ_{AE}	Pass-through of expected inflation	$[0, 1]$	$[0, 1]$
ϕ_{DPQ}	Quantity response to firm excess demand	$[0, 1]$	$[0, 1]$
ϕ_{GQ}	Quantity response to expected growth	$[0, 1]$	$[0, 1]$
ϕ_{UNION}	Wage indexation	$[0, 1.25]$	–
β_{IDX}	Wage catch-up weight	–	$[0, 1]$
λ_{RW}	Real-wage error correction	–	$[0, 0.5]$
δ_U	Wage response to unemployment	–	$[0, 1.5]$

Notes: Supports are declared model ranges, not estimates of Dutch bargaining institutions. The fixed benchmark sets every M_0 coefficient to one.

Before adoption, support, held-out parameter recovery and simulation-based calibration (SBC) are checked (Talts et al., 2018). SBC ranks the known generating coefficient among posterior draws under the simulated observation design. Uniformity and finite-draw rank dispersion are tested with prespecified multiplicity correction across coefficients, tests and estimation windows. Point recovery, posterior concentration or a visually plausible density cannot replace these checks.

The baseline five-member estimator passes its recorded gate in the Dutch designs, and its synthetic rank screen records no rejection (Appendix D). The error-correction estimator and both broader variants fail on wage-parameter calibration. The gate is not relaxed. The error-correction estimator is nevertheless used for the cross-model comparison; its M_1 results are labelled exploratory throughout.

5.6 Three different predictive checks

A prior predictive check asks whether the stated priors, initialization and rules generate plausible joint paths. A fitted posterior predictive check asks whether the observed paths look typical under the estimated predictive distribution,

$$\hat{P}_m(dy | X_r, \mathcal{I}_o) = \int P_m(dy | \theta, \mathcal{I}_o) \hat{\Pi}_m(d\theta | X_r). \quad (21)$$

A held-out check evaluates targets absent from estimation, using frozen early or pre-2020 fits and fresh complete forecast paths. These checks are not interchangeable.

A simple panel diagnostic averages paths across origins before scoring the resulting five-variable, twelve-horizon vector; we report it only as a secondary diagnostic. The main checks preserve that complete vector at each origin and rank the maximum origin-specific Mahalanobis discrepancy against the same maximum computed for complete simulated reference panels. Normalization uses batch A only, with fixed covariance regularization; reference batch B has disjoint parameter-draw groups. Two simulation siblings sharing a parameter draw remain dependent reference paths. The maximum is exploratory, not a preregistered statistic, and does not treat origins as independent.

These fitted ranks are descriptive joint posterior predictive criticism. They are not automatically uniform frequentist p -values: information has been reused, posterior approximation can add error, and reference siblings are not independent parameter draws. An independent reference batch does not remove these issues. Small tail areas indicate atypical fitted predictions, not the probability that the economic model is wrong or exclusion of every admissible parameter.

Held-out checks avoid fitting to their target quarters, but do not automatically have fixed-parameter size either. Section 6 reports their false-alarm and detection rates in the synthetic experiment; outside that design they carry no calibrated testing claim. Overlapping origins remain dependent.

5.7 Model-family specification testing

Predictive criticism at a fitted posterior is weaker than a test that no admissible parameter vector reproduces the conditional joint forecast law. We implement a maximized Monte Carlo test of that null (Dufour, 2006), searching the full parameter box rather than the posterior (Appendix G.2). A searched maximum is only a lower bound on the model p -value, and the certified upper bound remains one, so no family-wide rejection is reported.

5.8 Forecasts and the wage-shock experiment

The fixed empirical arm sets the six baseline response coefficients to one. The estimated arm uses the same rules, origin states and forecast information. Forecasts are free-running after the origin; conditional scenarios supplied with realized energy paths are outside the current reported results. Evaluation compares RMSE and proper predictive scores, together with coverage and interval width, over the core five variables and eligible origin-lead cells. Overlap and the short episode sample limit population-ranking claims (Diebold, 2015).

For the response experiment, every firm’s base wage is multiplied by 1.05 once at the first forecast step, before prices and quantities are decided. There is no repeated quarterly increase, automatic reversal or burn-in. Pair each shocked path with its own unshocked path using the same initial state and per-step random stream. The primary object is $\tau_{m,12}$ in (15); full wage, output, price and employment paths are secondary outcomes. Level IRFs use shocked/baseline ratios with one as the no-effect value. Inflation contributions are cumulated differences, never ratios of inflation rates.

Dutch response summaries condition on the saved finite posterior pool and use 128 paired seed streams per origin. Wages in the response comparison are compensation per employee; the common wage-index multiplier alone omits the shocked firm wage base. The synthetic response estimators are defined in Section 6.4.

6 Known-truth simulation study

6.1 An eight-cell design

The synthetic economy provides the truth that the Dutch application cannot. We cross two generating wage rules, the presence or absence of a cost-coefficient break, and two fitted wage rules. All cells share markup pricing, quantity setting, production, labor matching and aggregate expectations. The controlled comparison has six estimated coefficients in M_0 and eight in M_1 ; the fixed-all-ones empirical benchmark is not a synthetic arm.

Table 4: **Design of the known-truth experiment.** Each row is one cell of 100 independent panels: the wage rule that generates the data, whether ϕ_{CP} breaks from 0.25 to 0.50 between the initial and post phases, and the wage rule that is fitted.

Generating rule	Break in ϕ_{CP}	Fitted rule	Interpretation
M_0	No	M_0	Correct-rule recovery control
M_0	No	M_1	Alternative fitted to baseline truth
M_1	No	M_0	Omitted error-correction wages
M_1	No	M_1	Correct-rule recovery control
M_0	Yes	M_0	Correct rule, genuine change
M_0	Yes	M_1	Rule difference and genuine change
M_1	Yes	M_0	Wage omission and genuine change
M_1	Yes	M_1	Correct rule, genuine change

Pre- and post-boundary panels use the same fourteen origin templates, with independently realized twelve-quarter paths. The break is between phase sets, never inside one simulated path. The no-break arm has the same placebo boundary. Synthetic phases are not Dutch early and pre-2020 estimation periods, and the experiment does not represent one continuously evolving economy crossing a break. No burn-in is used.

Each cell has 100 independent outer panels. A panel consists of the fourteen origin templates in each of three phases: a shared initial phase and independent no-break and break post phases. The no-break and break post phases share common random numbers, so the break comparison is paired. Every panel is conditioned on by the frozen estimator of the fitted model, giving 1,200 posterior conditionings and no retraining. Each conditioned posterior generates 128 parameter draws with two simulations per draw and origin for predictive evaluation. Fresh held-out targets come from a separate seed namespace and never enter estimation.

Each fitted model is assessed for coefficient recovery, the estimated change in ϕ_{CP} , joint prior and posterior predictive adequacy, held-out forecasting and the paired wage-shock response. Correct-rule break cells distinguish a stale coefficient from a wrong economic rule. Cross-rule cells need not be globally distinguishable at every parameter value; model-family rejection is an outcome to establish, not a label guaranteed by changing one equation.

6.2 What each cell can establish

The cells fitted with M_0 use the indexation estimator, which passes its empirical calibration gate; its synthetic rank screen records no rejection (Table 18). They carry the main result: M_0 fitted to M_0 data is the correct-rule control, and M_0 fitted to M_1 data is the omitted-error-correction experiment. The cells fitted with M_1 ask whether fitting the right wage rule helps; because that estimator fails calibration, their coverage figures are not calibrated statements and their bias and response errors are exploratory.

A prior predictive check precedes estimation. Under the indexation prior, every error-correction-generated panel has a worst-origin rank at or below 0.05 (share 1.00), with mean worst-origin rank 0.021–0.024 across phases. Under the error-correction prior, no indexation-generated panel does (share 0.00; mean rank 0.61–0.65). The check is asymmetric: the indexation prior cannot produce error-correction data, while the error-correction prior covers indexation data. Table 5 reports every cell as a range over the three phases, which barely differ; the ranks are descriptive, not tests.

Table 5: **Prior predictive check in the known-truth experiment.** Mean worst-origin upper-tail rank of the generated panels under the fitted model’s prior predictive distribution, and the share of panels with rank at or below 0.05. Each entry is the range over the initial, post no-break and post break phases.

Fitted model and generating data	Mean worst-origin rank	Share with rank ≤ 0.05
<i>A. Indexation model M_0 fitted (validated estimator)</i>		
M_0 on M_0 data (correct rule)	0.705–0.744	0.00
M_0 on M_1 data	0.021–0.024	1.00
<i>B. Error-correction model M_1 fitted (exploratory estimator)</i>		
M_1 on M_1 data (correct rule)	0.687–0.721	0.00
M_1 on M_0 data	0.610–0.651	0.00

Notes: The three phases are pooled into a range because they barely differ; a single value means all three phases give the same figure. Each phase averages 100 independent panels; all panels are usable. A rank near zero means the generated data are atypical of the fitted model’s prior. Ranks are descriptive, not tests. Monte Carlo standard errors are below 0.02 throughout.

6.3 Coefficient bias and the cost-coefficient break

All 1,200 panel–phase conditionings completed, with no whole-panel failure. Table 6 reports recovery of ϕ_{CP} ; Appendix F reports every coefficient.

In the correct-rule cells the projection bias b_m is zero, so the error in $\Delta\hat{\phi}_{CP}$ is estimation error alone. It is small: against a true change of 0.25, the bias of the estimated break is +0.02 under M_0 and –0.02 under M_1 , and the estimated placebo change rounds to zero under both. Against this benchmark, a cross-rule deviation in $\Delta\hat{\phi}_{CP}$ measures Δb_m . Fitting M_0 to error-correction data biases the level of ϕ_{CP} by +0.38 before the break and +0.35 after it, and the 90% intervals

Table 6: **Recovery of the cost-surprise coefficient ϕ_{CP} .** Bias of the posterior mean and coverage of the central 90% interval in each phase, and bias of the estimated change $\Delta\hat{\phi}_{CP}$ from the initial phase. The true ϕ_{CP} is 0.25, rising to 0.50 in the break phase.

	Initial phase		Post, no break		Post, break		Bias of $\Delta\hat{\phi}_{CP}$	
	Bias	Cov.	Bias	Cov.	Bias	Cov.	No break	Break
Fitted model and generating data								
<i>A. Indexation model M_0 fitted (validated estimator)</i>								
M_0 on M_0 data (correct rule)	−0.010 (0.008)	0.94 (0.02)	−0.005 (0.008)	0.89 (0.03)	+0.006 (0.007)	0.93 (0.03)	+0.005	+0.016
M_0 on M_1 data	+0.382 (0.009)	0.26 (0.04)	+0.387 (0.010)	0.33 (0.05)	+0.350 (0.006)	0.15 (0.04)	+0.005	−0.032
<i>B. Error-correction model M_1 fitted (exploratory estimator)</i>								
M_1 on M_1 data (correct rule)	+0.015 (0.008)	0.97 (0.02)	+0.014 (0.008)	0.94 (0.02)	−0.007 (0.007)	0.97 (0.02)	−0.001	−0.021
M_1 on M_0 data	+0.078 (0.007)	0.96 (0.02)	+0.066 (0.009)	0.98 (0.01)	−0.084 (0.009)	0.99 (0.01)	−0.012	−0.162

Notes: Averages over 100 independent panels, with Monte Carlo standard errors in parentheses. Nominal coverage is 0.90. The true change is 0 without a break and 0.25 with one; its bias is the difference of phase-mean biases on the same panels. RMSE and every other coefficient are in Table 24. M_1 -fitted rows use an estimator that fails calibration.

cover the truth in only 0.15–0.33 of panels. The bias is almost constant across phases, so the estimated break remains close to the truth (bias −0.03). Level and change therefore answer different questions: the omitted wage rule distorts what ϕ_{CP} measures but not how it moves. The distortion is not confined to ϕ_{CP} : demand pull is biased by +0.28 (coverage 0.41) and the growth-expectation coefficient by +0.11 (coverage 0.36). In the exploratory reverse cell, M_1 fitted to indexation data, the projection bias changes sign with the break, from +0.08 to −0.08, and the bias of the estimated break is −0.16, about two thirds of the true change.

Figure 3 shows what the level bias looks like in a single panel. Under the correct wage rule the joint posterior of ϕ_{CP} and ϕ_{DP} concentrates around the truth, and the posterior means of the 100 panels scatter tightly around it. Fitted to error-correction data, the posterior is wider and irregular but places almost no mass near the truth: higher cost pass-through and much higher demand pull jointly substitute for the missing wage dynamics. The posterior itself gives no sign that it is centred in the wrong place; that signal has to come from the predictive checks of Section 6.5.

6.4 Transmission errors

For each panel and phase the fitted model’s response $\hat{\tau}_{m,r,12}$ is compared with the generating model’s response on the same origins. Each comparison uses eight paired paths per origin, split into two independent halves A and B of four paths. Each half draws a fresh set of sixteen posterior parameters from the equal-weight five-member mixture, independent of the draws used for forecasts and warnings, and pairs generating and fitted paths through common random numbers. With half errors e^A and e^B , the product $e^A e^B$ is an unbiased estimate of the squared expected-response error under that independence; it can be negative and is never truncated. Its square root, the expected RMSE, is in percentage points like the signed error. Raw absolute errors also contain inner simulation noise and are reported separately.

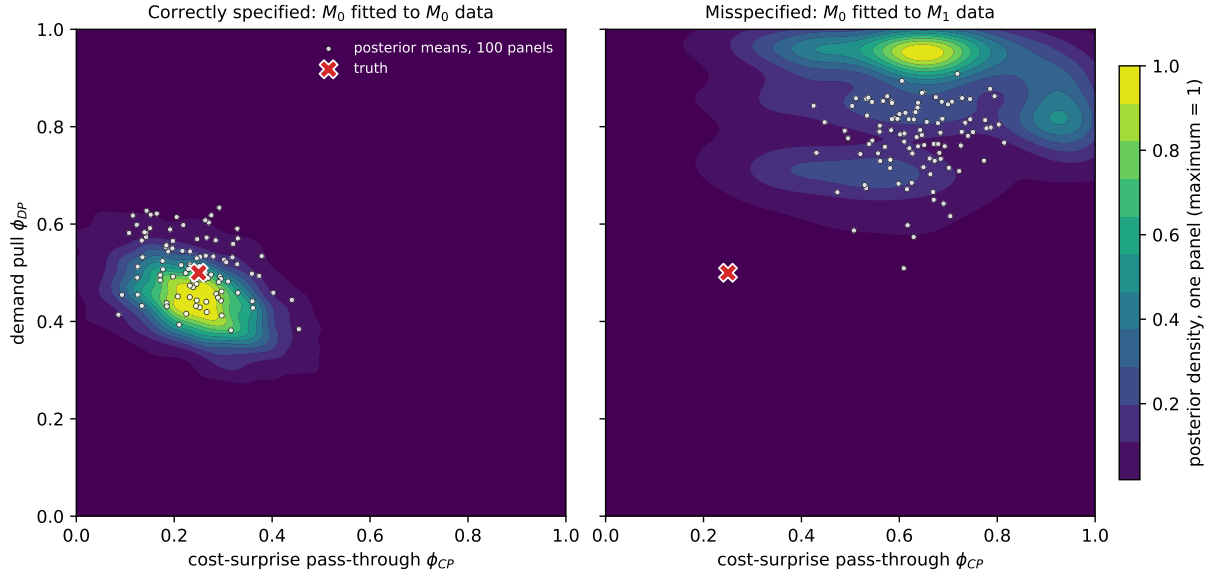


Figure 3: **Posteriors under the correct and the wrong wage rule.** Joint posterior density of cost-surprise pass-through ϕ_{CP} and demand pull ϕ_{DP} for one representative panel in the initial phase (the panel with the median correct-rule ϕ_{CP} estimate; the same generating seed in both panels), with the posterior means of all 100 panels (white dots) and the truth (red cross). Left: M_0 fitted to M_0 data. Right: M_0 fitted to M_1 data.

Table 7: **Error in the response to a common 5% wage reset.** Fitted minus generating twelve-quarter consumer-price response: mean signed error and expected root mean squared error (RMSE), by phase, in percentage points.

Fitted model and generating data	Initial phase		Post, no break		Post, break	
	Signed error	Expected RMSE	Signed error	Expected RMSE	Signed error	Expected RMSE
<i>A. Indexation model M_0 fitted (validated estimator)</i>						
M_0 on M_0 data (correct rule)	+0.01 (0.02)	0.18	+0.01 (0.02)	0.17	+0.06 (0.03)	0.29
M_0 on M_1 data	+1.04 (0.03)	1.08	+1.09 (0.04)	1.15	+1.10 (0.04)	1.16
<i>B. Error-correction model M_1 fitted (exploratory estimator)</i>						
M_1 on M_1 data (correct rule)	+0.08 (0.03)	0.26	+0.10 (0.02)	0.23	+0.07 (0.04)	0.33
M_1 on M_0 data	+0.32 (0.03)	0.41	+0.27 (0.03)	0.39	-0.05 (0.04)	0.36

Notes: Averages over 100 independent panels, with Monte Carlo standard errors of the signed error in parentheses. Positive errors mean the fitted model overstates transmission. The expected RMSE is the square root of the mean product $e^A e^B$ of errors from two independent halves of the response simulations, so inner simulation noise is removed; the report gives no Monte Carlo standard error for it.

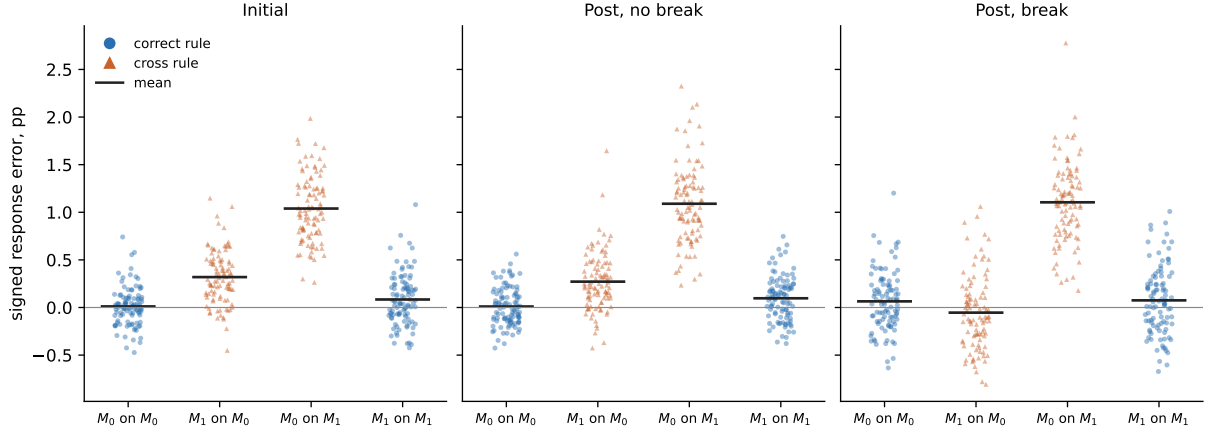


Figure 4: Signed error of the twelve-quarter CPI response to a common 5% wage reset across the 100 panels of each cell, fitted minus generating model (percentage points). Circles: correct rule; triangles: cross rule; bars: means.

Fitting M_0 to error-correction data overstates transmission. Its twelve-quarter CPI response exceeds the generating response by +1.04 percentage points in the initial phase, +1.09 without a break and +1.10 after it (MCSE at most 0.04), against +0.01 to +0.06 in the correct-rule control. The expected RMSE, 1.08–1.16 percentage points, is several times the control’s 0.17–0.29, so inner simulation noise does not drive the result (Figure 4). The error does not grow after the break, consistent with the stable level bias in ϕ_{CP} . The reverse omission is smaller and changes sign: M_1 fitted to indexation data overstates the response by +0.32 percentage points before the break, while the error after the break, -0.05 , is not distinguishable from zero; it follows its projection bias.

6.5 Do fit, forecasts or constancy warn?

Three features are computed for each panel before its response is evaluated: the maximum origin-level joint discrepancy of the fitted data, the energy-score loss of fresh held-out forecasts on a fixed common scale, and the posterior probability that ϕ_{CP} rose between phases. Panels 0–49 of the correct-rule, no-break cell of each fitted family calibrate the warnings; panels 50–99 of every cell evaluate them. Each feature is standardized by its calibration mean and standard deviation. The base warning is the larger of the standardized fit and forecast features; the augmented warning also takes the constancy feature. Thresholds are calibration quantiles fixed before evaluation, with the 0.95 quantile primary and 0.50–0.98 tracing a curve; an alarm requires the composite to exceed the threshold strictly. The constancy feature targets the prespecified positive break and is not a general constancy test.

We report achieved control alarm rates with their uncertainty, not a guaranteed size. For unflagged panels we report the mean absolute response error and, separately, the residual error mass per panel; more alarms mechanically lower the latter. Base and augmented warnings are compared on the same panels. These are retrospective warnings from completed forecast paths, not real-time early warnings or certified tests.

Table 8: **What the warnings catch.** For each cell after the boundary: the mean absolute error of the twelve-quarter consumer-price response, the share of panels flagged, and the absolute response error left in unflagged panels, for the fit-and-forecast warning (F) and the warning augmented with parameter constancy (F+C), at a strict and a loose alarm threshold.

			Strict (0.95)				Loose (0.80)	
		Mean	Flagged		Left		Flagged	
Cell	Phase	error	F	F+C	F	F+C	F	F+C
<i>A. Indexation model M_0 fitted (validated estimator)</i>								
M_0 on M_0 data (correct rule)	No break (control)	0.15	0.10	0.10	0.13	0.13	0.38	0.34
	Break	0.20	0.10	0.10	0.16	0.16	0.26	0.92
M_0 on M_1 data	No break	1.04	1.00	1.00	0.00	0.00	1.00	1.00
	Break	1.10	1.00	1.00	0.00	0.00	1.00	1.00
<i>B. Error-correction model M_1 fitted (exploratory estimator)</i>								
M_1 on M_1 data (correct rule)	No break (control)	0.19	0.00	0.00	0.19	0.19	0.08	0.18
	Break	0.31	0.04	0.58	0.29	0.10	0.22	0.88
M_1 on M_0 data	No break	0.29	0.98	0.98	0.01	0.01	0.98	0.98
	Break	0.33	1.00	1.00	0.00	0.00	1.00	1.00

Notes: Flagged is the share of panels flagged. Errors are in percentage points; left is the absolute response error of unflagged panels summed and divided by the number of panels, so it falls as more panels are flagged. Thresholds are quantiles of the warning score in panels 0–49 of each family’s correct-rule, no-break cell; all entries use the independent panels 50–99. Control rows give the achieved false-alarm rate, against nominal 0.05 and 0.20. In correct-rule break cells an alarm flags a genuine change in ϕ_{CP} , elsewhere a wrong wage rule. Monte Carlo standard errors of the shares are at most 0.07.

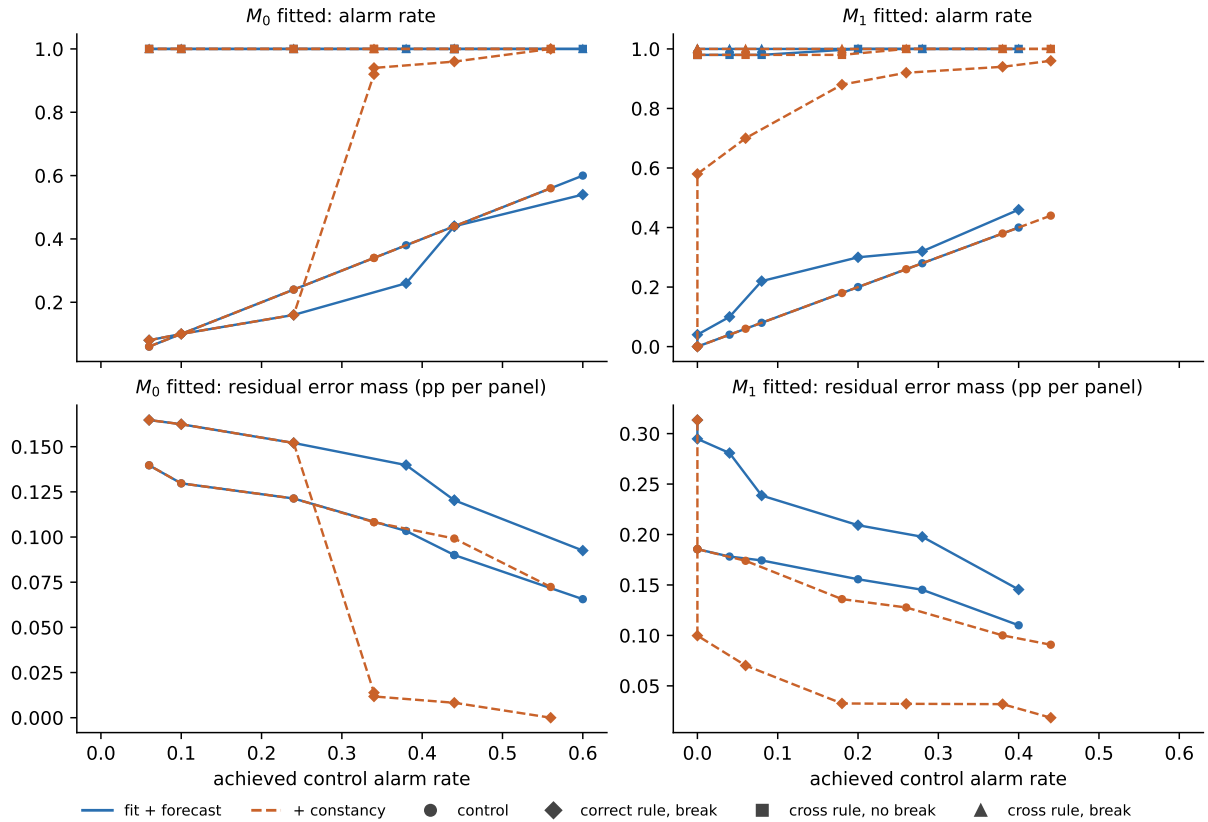


Figure 5: Warning operating characteristics over calibration quantiles 0.50–0.98. Top: alarm rate in each evaluation cell; bottom: residual absolute response error per panel left unflagged, for the correct-rule cells. Cross-rule cells are flagged almost entirely at every threshold, so their residual error mass is near zero and not shown. The horizontal axis is the achieved independent control alarm rate.

The warnings separate the wrong-rule panels. At the primary threshold, fit and forecast losses flag every panel of M_0 fitted to error-correction data, in both post phases, and 0.98 to 1.00 of the reverse cell. The achieved control alarm rates are 0.10 for M_0 and 0.00 for M_1 , not the nominal 0.05, because calibration and evaluation panels are independent finite samples. Because cross-rule panels are flagged almost without exception, almost no unflagged panel carries a cross-rule response error; the residual risk lies in correct-rule panels, whose errors are an order of magnitude smaller. Constancy adds nothing for M_0 at the primary threshold: the break cell is flagged at the control rate, 0.10. Its contribution appears only at looser thresholds, where it flags 0.92 of break panels at a control alarm rate of 0.34 (Figure 5). For the exploratory M_1 fits it raises break detection from 0.04 to 0.58 at no control alarms and cuts the residual error mass from 0.29 to 0.10. In both families the constancy feature flags a genuine change in the correct rule, not a transmission error.

6.6 Summary

Within this design, an estimated cost coefficient can track a change in pass-through while misstating its level. Fitting the indexation model to error-correction data biases ϕ_{CP} by +0.35 to +0.39 and overstates the twelve-quarter response to a common wage reset by +1.09 percentage points, yet recovers the break. Joint fit and forecast warnings flag every such panel, and constancy adds little to them for the indexation estimator. These operating characteristics are local to one omitted wage rule, one break and one shock.

7 Dutch evidence

The Dutch application compares three versions of the same markup economy on revised-data origin states: fixed CANVAS with all six response coefficients at one, the estimated indexation model M_0 , and the exploratory error-correction model M_1 . It asks whether estimation makes joint aggregate forecasts more adequate and how it changes the model's common-wage response. Conditional forecasts supplied with realized energy paths are not included.

7.1 Prior predictive check

Before estimation, each wage model's prior is confronted with the observed panel. Ranks compare the maximum origin-level joint discrepancy of the observed paths with the same maximum over 25,000 reference paths from 2,500 disjoint prior parameter groups. None of the 1,200,000 prior paths of either model is numerically invalid. Table 9 shows that the observed panels are already atypical before estimation under both wage rules. The fitted panels have small but non-floor ranks, between 0.0035 and 0.0165, and the calm held-out panel ranks 0.0124 under M_0 and 0.0784 under M_1 . The held-out panel of the pre-2020 window, which covers 2020 onward, is the extreme case: under both priors its largest discrepancy, at the 2020Q1 origin whose window covers the pandemic and the start of the inflation episode, lies at the rank floor. The error-correction rule widens the prior predictive distribution of wages without removing these discrepancies.

Figure 6 locates the discrepancy. At a calm origin, 2016Q1, the observed paths of all five variables lie inside the central 90% prior bands. At the episode origin 2021Q3, consumer prices

Table 9: **Prior predictive checks of both wage models.** For each Dutch panel, the largest origin-level joint discrepancy between the observed paths and the model’s prior predictive distribution, the origin where it occurs, and its upper-tail rank among prior draws.

Panel	Indexation, M_0			Error correction, M_1		
	Discrepancy	Origin	Rank	Discrepancy	Origin	Rank
Early, fitted	262.2	2010Q3	0.0042	243.5	2010Q3	0.0035
Pre-2020, fitted	262.2	2010Q3	0.0165	243.5	2010Q3	0.0109
Early, held-out	123.2	2016Q3	0.0124	48.3	2016Q3	0.0784
Pre-2020, held-out	2534.5	2020Q1	<0.0001	1995.9	2020Q1	<0.0001

Notes: Each model has its own prior bank of 5,000 parameter points with ten simulations each. Ranks compare the observed maximum with the same maximum over 25,000 reference paths from 2,500 disjoint prior parameter groups; the floor is 0.00004. Each origin contributes five variables over twelve quarters. A small rank means the observed panel is atypical of the prior; ranks are descriptive and carry no significance level.

and the value-added deflator leave the M_0 band within two quarters and stay above it for most of the horizon; twelve-quarter consumer-price growth exceeds 98% of M_0 prior paths and 92% of M_1 prior paths. Observed wage growth reaches the upper edge of the M_0 band only in the last quarters and lies inside the wider M_1 band, and unemployment stays inside both. The discrepancy is therefore mainly one of prices: across the prior range of the price coefficients, the model rarely produces a price increase as fast as the episode’s. These ranks describe the prior mixtures, not every admissible parameter value; that stronger statement needs the model-family test of Section 5.

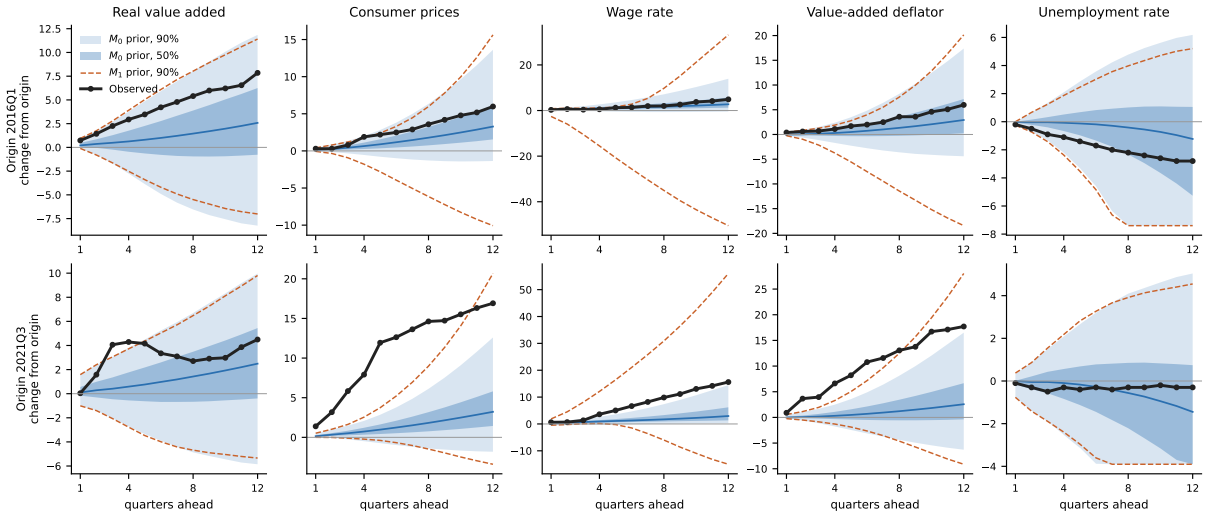


Figure 6: Prior predictive distributions and observed Dutch paths at a calm origin (2016Q1, top) and the inflation-episode origin (2021Q3, bottom). Blue bands are the central 50% and 90% of M_0 prior predictive paths and the blue line their median; dashed orange lines bound the central 90% of M_1 prior paths. Black dots are the observed paths. Each band uses 50,000 prior paths (5,000 parameter points, ten simulations each). Values are cumulative percent changes from the origin; unemployment is in percentage points.

7.2 Estimates

The baseline five-member neural estimator passes its recorded simulation-based calibration gate for the early and pre-2020 observation designs. Under this pricing rule, the posterior cost-surprise coefficient is approximately 0.01–0.02 in the two windows, far below the fixed value of one. This is a conditional estimate within the specified model and information set. It is

not a direct estimate of economy-wide labor-cost pass-through.

Table 10 reports all six coefficients, their central posterior intervals and the fixed benchmark. It also separates descriptive window changes from hypothesis tests. The wage-indexation coefficient and the quantity demand-response coefficient display larger standardized changes than the cost-surprise coefficient. The windows overlap, so neither a standardized difference nor interval overlap supplies a valid independent-sample significance test.

Table 10: **Posterior estimates of the response coefficients.** Posterior means and central 90% intervals in the early and pre-2020 estimation windows, against the fixed benchmark in which every M_0 coefficient equals one.

	Fixed	Early window		Pre-2020 window		Shift	Overlap
Coefficient		Mean	90% interval	Mean	90% interval		
<i>A. Indexation wages, M_0 (estimator passes calibration)</i>							
Demand pull, ϕ_{DP}	1	0.761	[0.433, 0.987]	0.765	[0.239, 0.994]	0.01	1.00
Cost surprise, ϕ_{CP}	1	0.012	[0.001, 0.055]	0.023	[0.000, 0.097]	0.25	1.00
Expected inflation, ϕ_{AE}	1	0.792	[0.621, 0.961]	0.687	[0.514, 0.838]	−0.69	0.67
Quantity demand, ϕ_{DP_Q}	1	0.217	[0.043, 0.602]	0.575	[0.274, 0.889]	1.30	0.59
Quantity growth, ϕ_{G_Q}	1	0.681	[0.504, 0.904]	0.737	[0.622, 0.871]	0.38	1.00
Wage indexation, ϕ_{UNION}	1	0.190	[0.090, 0.236]	0.318	[0.288, 0.366]	2.18	0.00
<i>B. Error-correction wages, M_1 (exploratory: estimator fails calibration)</i>							
Demand pull, ϕ_{DP}	−	0.099	[0.012, 0.204]	0.076	[0.002, 0.237]	−0.20	1.00
Cost surprise, ϕ_{CP}	−	0.084	[0.014, 0.230]	0.059	[0.003, 0.173]	−0.29	0.94
Expected inflation, ϕ_{AE}	−	0.814	[0.696, 0.921]	0.908	[0.779, 0.991]	0.90	0.67
Quantity demand, ϕ_{DP_Q}	−	0.107	[0.008, 0.329]	0.235	[0.120, 0.402]	0.87	0.74
Quantity growth, ϕ_{G_Q}	−	0.663	[0.466, 0.877]	0.710	[0.590, 0.843]	0.30	1.00
Wage indexation, β_{IDX}	−	0.275	[0.212, 0.409]	0.430	[0.366, 0.497]	1.76	0.33
Real-wage correction, λ_{RW}	−	0.006	[0.000, 0.021]	0.431	[0.310, 0.500]	5.40	0.00
Unemployment gap, δ_U	−	0.008	[0.000, 0.034]	0.182	[0.121, 0.308]	2.41	0.00

Notes: Shift is $(\bar{\theta}_{pre} - \bar{\theta}_{early}) / \sqrt{s_{pre}^2 + s_{early}^2}$; overlap is the length of the intersection of the two intervals divided by the shorter one. The windows share data, so both are descriptive, not tests. M_1 intervals are not calibrated uncertainty statements.

The error-correction estimator fails the calibration gate on wage-block coefficients in both windows (Table 18); Appendix D reports the local-information, representation and density-head diagnostics. Table 10 reports its saved posterior pools. The cost-surprise coefficient is again small, 0.084 and 0.059 in the two windows, so replacing the wage rule does not move the estimate away from zero (Section 7.6 discusses what this implies). The wage block is unstable across the overlapping windows: the real-wage correction speed moves from 0.006 to 0.431, consistent with the weak identification documented in the appendix.

7.3 Joint predictive adequacy

The origin-averaged diagnostic gives ranks near 0.008 (Appendix D). The checks in Table 11 preserve each origin's full five-variable, twelve-horizon path. Both held-out regimes reach the reference rank floor for the maximum discrepancy. This is descriptive predictive criticism, not a calibrated rejection of the whole model family, and does not isolate the wage equation as the cause.

Table 11 shows the panel maximum and Table 12 the separate origin-level twelve-quarter component errors. The held-out pre-window panel includes the inflation episode; its price and

wage discrepancies are no longer concealed by an average over origins. These component errors give the economic content of the joint discrepancy without turning variables or overlapping forecast origins into independent tests.

Table 11: **Posterior predictive adequacy on the Dutch panels.** Largest origin-level joint discrepancy between observed and posterior predictive paths, the origin where it occurs and its upper-tail rank, for the estimated indexation (M_0) and error-correction (M_1) models.

Panel	M_0			M_1 (exploratory)		
	Discrepancy	Origin	Rank	Discrepancy	Origin	Rank
Early window, fitted panel	216.4	2012Q3	0.0310	443.4	2010Q3	0.0078 [†]
Pre-2020 window, fitted panel	256.0	2013Q1	0.0620	277.3	2013Q1	0.0078 [†]
Early window, held-out panel	170.4	2016Q3	0.0078 [†]	156.9	2016Q3	0.0078 [†]
Pre-2020 window, held-out panel	2638.6	2020Q1	0.0078 [†]	2148.7	2020Q1	0.0078 [†]

Notes: The discrepancy keeps all five variables and twelve quarters of every origin. Ranks compare the observed panel with 128 posterior predictive paths, so the floor is $1/129 \approx 0.0078$; [†] marks a rank at the floor, which is reached in every panel except the two M_0 fitted panels. Early held-out origins (2016Q1, 2016Q3) are calm; pre-2020 held-out origins (2020Q1–2021Q3) span the inflation episode. Ranks are descriptive, not model-family tests.

Table 12: **Held-out forecast errors at twelve quarters.** Observed minus posterior predictive mean, by held-out origin and variable.

Origin	Model	Real VA	CPI	Wage	VA defl.	Unemp.	Rank
<i>A. Calm held-out origins</i>							
2016Q1	M_0	5.41	0.51	2.65	0.13	−1.34	0.0155
	M_1	6.40	1.69	−1.22	1.74	−2.13	0.0078 [†]
2016Q3	M_0	4.45	1.32	3.51	0.62	−0.76	0.0078 [†]
	M_1	5.62	2.95	2.18	2.83	−1.65	0.0078 [†]
<i>B. Episode held-out origins</i>							
2020Q1	M_0	5.07	11.62	5.88	8.18	0.63	0.0078 [†]
	M_1	7.97	12.66	4.15	9.59	−0.77	0.0078 [†]
2020Q3	M_0	7.40	13.93	10.69	10.45	0.22	0.0078 [†]
	M_1	8.60	14.06	7.50	10.30	−0.52	0.0078 [†]
2021Q1	M_0	6.14	12.85	12.28	14.66	0.49	0.0078 [†]
	M_1	6.59	12.63	9.08	13.88	0.35	0.0078 [†]
2021Q3	M_0	1.58	12.84	12.30	14.39	0.94	0.0078 [†]
	M_1	2.25	11.97	8.33	12.69	0.53	0.0078 [†]

Notes: Cumulative log-percent changes from the origin, except unemployment in percentage points. Positive values mean the model under-predicts. Rank is the origin’s joint upper-tail rank; [†] marks the floor $1/129$.

The error-correction model does not remove these discrepancies (Table 11). All four of its panels reach the rank floor, and its largest discrepancy is again the episode origin 2020Q1 (statistic 2149). Neither wage rule reproduces the joint wage–price behavior of the episode.

7.4 Forecasts

Tables 13 and 14 compare fixed CANVAS, estimated M_0 , M_1 and AR(1) on identical eligible origin–horizon cells. Each regime uses eight semiannual origins, with 60 scored cells per variable in QOOS and 80 in WOOS. Every model arm uses 512 paths per origin, with no cap hits or nonfinite scored paths. The AR(1) reference is an analytic point forecast, not a predictive ensemble with a nominal 90% interval.

Estimation does not give a uniform forecast gain. In QOOS, consumer-price RMSE falls from 1.96 under fixed CANVAS to 1.24 under estimated M_0 , and value-added-deflator RMSE falls from 2.66 to 1.67. Real-value-added RMSE instead rises from 1.46 to 3.20, and wage RMSE rises from 1.55 to 2.87. In WOOS, price RMSE improves relative to fixed CANVAS but remains above AR(1), while wage RMSE increases from 4.10 to 6.60. These are realized matched losses, not significance tests establishing a population ranking.

Table 13: **Point forecast accuracy.** Root mean squared error (RMSE) of ensemble-mean forecasts over all origins and horizons of one to twelve quarters, by variable and model, in levels (panel A) and as a ratio to the AR(1) RMSE (panel B).

Variable	QOOS, calm				WOOS, episode			
	Fixed	M_0	M_1	AR(1)	Fixed	M_0	M_1	AR(1)
<i>A. Root mean squared error</i>								
Real value added	1.46	3.20	3.62	3.02	4.63	4.92	5.31	5.58
Consumer prices	1.96	1.24	1.25	1.36	8.61	8.23	7.90	7.58
Wage rate	1.55	2.87	1.97	2.16	4.10	6.60	4.34	5.54
Value-added deflator	2.66	1.67	1.59	2.03	8.11	7.36	6.70	6.90
Unemployment rate	0.48	1.31	1.57	1.32	0.79	0.67	0.71	0.71
<i>B. Relative to AR(1)</i>								
Real value added	0.48	1.06	1.20		0.83	0.88	0.95	
Consumer prices	1.44	0.92	0.92		1.13	1.09	1.04	
Wage rate	0.72	1.33	0.91		0.74	1.19	0.78	
Value-added deflator	1.31	0.83	0.78		1.18	1.07	0.97	
Unemployment rate	0.36	0.99	1.19		1.11	0.94	0.99	

Notes: Errors are in cumulative percent changes from the origin, unemployment in percentage points; lower is better. In panel B a value below one beats AR(1). Every model is scored on the same cells: 60 per variable in QOOS (eight origins) and 80 in WOOS. Fixed sets all six coefficients to one; M_0 and M_1 are estimated, M_1 exploratorily. Each model uses 512 simulated paths per origin, with no capped or invalid paths. AR(1) is an analytic benchmark. Differences are descriptive, not tests.

CRPS assesses a different object: the estimated predictive distributions improve on the point-mass AR(1) reference for all four non-wage variables in QOOS and three in WOOS. Wage CRPS is worse in both regimes; WOOS unemployment is also worse than AR(1), at 0.606 versus 0.497. Lower CRPS does not imply lower RMSE or calibrated predictive intervals. Estimated wage coverage in QOOS is 0.233 for nominal 90% intervals, versus 0.867 under fixed CANVAS. In WOOS, estimated M_0 coverage is 0.100 for wages, 0.175 for consumer prices and 0.275 for the value-added deflator. Its corresponding intervals are narrower than the fixed model's. These are empirical fractions over overlapping forecasts, not independent-binomial estimates of population coverage.

The combination is economically important: a forecast can improve on a benchmark and still materially understate uncertainty or miss the joint behavior of wages and prices. An inference-calibrated posterior can also yield poor predictive coverage when its economic or observation model is inadequate. Forecast rankings alone therefore do not validate the small modeled wage-shock response.

The error-correction model mainly changes the wage forecasts. Its wage RMSE is 1.97 in QOOS and 4.34 in WOOS, below the estimated indexation model in both regimes, and its wage interval coverage rises to 0.450 and 0.500. Episode price forecasts barely change: WOOS CPI RMSE is 7.90. Its posterior-mean plug-in forecasts, which are not a predictive distribution, have overall RMSE 1.85 in QOOS and 5.02 in WOOS. No forecast-comparison p -values are reported; overlapping origins do not support a small-sample test.

Table 14: **Density forecast accuracy and calibration.** Continuous ranked probability score (CRPS), empirical coverage of central 90% predictive intervals and their average width, by variable and model.

Variable	CRPS				90% coverage			90% width		
	Fixed	M_0	M_1	AR(1)	Fixed	M_0	M_1	Fixed	M_0	M_1
<i>QOOS: origins 2016Q1–2019Q3, calm targets</i>										
Real value added	1.08	1.64	1.92	2.48	1.00	0.82	0.48	13.7	7.2	6.8
Consumer prices	1.38	0.76	0.67	1.08	0.60	0.98	0.93	8.2	6.2	3.0
Wage rate	1.09	1.85	1.12	1.63	0.87	0.23	0.45	10.2	2.5	3.0
Value-added deflator	1.84	1.01	0.83	1.61	0.60	1.00	0.95	12.2	9.7	4.5
Unemployment rate	0.83	0.66	0.76	1.10	0.98	0.98	0.87	8.8	5.0	4.6
<i>WOOS: origins 2020Q1–2023Q3, episode targets</i>										
Real value added	2.59	2.82	3.12	4.13	0.75	0.69	0.56	11.6	9.6	8.2
Consumer prices	5.71	5.75	5.69	6.01	0.33	0.17	0.20	11.4	5.2	3.6
Wage rate	3.05	4.86	2.53	4.43	0.59	0.10	0.50	12.6	3.8	7.0
Value-added deflator	5.14	4.58	4.45	5.31	0.46	0.28	0.29	17.7	8.7	5.4
Unemployment rate	0.78	0.61	0.51	0.50	0.99	0.99	0.99	7.0	6.1	5.6

Notes: Lower CRPS is better; well-calibrated intervals have coverage near 0.90. Coverage is a share of overlapping forecast targets, not of independent trials. Units as in Table 13. AR(1) is a point forecast and has no interval.

7.5 Responses to the common wage reset

With a small coefficient the estimated baseline absorbs most unexpected cost increases through its markup; Section 7.6 explains why we read this as a specification symptom rather than a property of Dutch firms. Figure 7 shows the responses. The paired 5% wage-base reset gives estimated mean twelve-quarter CPI responses of about -0.291 percentage points at 2019Q4 and -0.087 at 2021Q4, with conditional Monte Carlo standard errors 0.083 and 0.082. These are small negative model-implied means, not a demonstrated zero response or evidence that a common Dutch wage increase has no inflation effect. They also differ from a median response and from a mean level ratio.

Table 15 reports percent deviations from the unshocked path at quarters 1, 4 and 12 for the fixed-all-ones model and both estimated models, with conditional Monte Carlo standard errors. All three arms use the same seeds and origin inputs. Wages are compensation per employee and include the shocked firm base wage; the wage-index multiplier alone is not that outcome. Each origin has 128 attempted pairs per arm and zero failures.

Under the fixed coefficients, mean twelve-quarter additional log CPI inflation is 5.710 percentage points at 2019Q4 and 5.536 at 2021Q4. Relative to the estimated model, the paired differences are 6.000 and 5.624 points, with conditional Monte Carlo standard errors 0.343 and 0.329. All 128 fixed pairs per origin are complete with zero failures. This is substantial sensitivity to calibration: all six coefficients change together, so the contrast is not an isolated effect of ϕ_{CP} . Neither model's response is observed Dutch counterfactual truth.

Figures 8 and 9 compare all five response outcomes. The same economic equations, initialization and shock protocol are used in both arms. The retained aggregate cost records do not contain the firm-level allocations needed to reconstruct an exact five-channel markup/CPI decomposition; no such decomposition is inferred from averaged cost terms.

The exploratory error-correction model gives 0.048 percentage points at 2019Q4 and 0.093 at

Table 15: **Responses to a common 5% wage reset.** Deviation from the unshocked path at quarters 1, 4 and 12 under fixed coefficients and the estimated indexation (M_0) and error-correction (M_1) models, in percent (unemployment in percentage points).

Quarter	Fixed			M_0			M_1		
	1	4	12	1	4	12	1	4	12
<i>Origin 2019Q4</i>									
Consumer prices	1.03 (0.00)	2.77 (0.05)	5.94 (0.34)	0.03 (0.00)	-0.07 (0.02)	-0.29 (0.08)	0.06 (0.01)	0.08 (0.01)	0.05 (0.02)
Wage per employee	6.48 (0.00)	8.27 (0.10)	12.07 (0.47)	5.01 (0.00)	5.04 (0.01)	4.96 (0.08)	5.00 (0.00)	5.25 (0.02)	4.93 (0.06)
Real value added	0.00 (0.00)	-1.04 (0.06)	0.60 (0.17)	0.00 (0.00)	0.02 (0.04)	0.03 (0.12)	0.00 (0.00)	-0.08 (0.03)	-0.39 (0.10)
Employment	0.00 (0.00)	-0.98 (0.07)	0.23 (0.15)	0.00 (0.00)	0.00 (0.04)	0.08 (0.11)	0.00 (0.00)	-0.10 (0.03)	-0.30 (0.09)
Unemployment rate (pp)	0.00 (0.00)	0.94 (0.07)	-0.18 (0.14)	0.00 (0.00)	0.00 (0.03)	-0.06 (0.10)	0.00 (0.00)	0.10 (0.03)	0.29 (0.09)
<i>Origin 2021Q4</i>									
Consumer prices	1.01 (0.00)	2.74 (0.05)	5.76 (0.34)	0.03 (0.00)	-0.02 (0.03)	-0.08 (0.08)	0.05 (0.00)	0.08 (0.01)	0.09 (0.01)
Wage per employee	6.49 (0.00)	8.24 (0.09)	11.33 (0.48)	5.01 (0.00)	5.02 (0.01)	5.02 (0.07)	5.00 (0.00)	5.23 (0.02)	5.18 (0.08)
Real value added	0.00 (0.00)	-0.36 (0.04)	0.50 (0.11)	0.00 (0.00)	0.03 (0.04)	-0.10 (0.09)	0.00 (0.00)	0.01 (0.03)	0.05 (0.09)
Employment	0.00 (0.00)	-0.36 (0.05)	0.34 (0.10)	0.00 (0.00)	0.02 (0.04)	-0.16 (0.09)	0.00 (0.00)	0.00 (0.04)	0.05 (0.09)
Unemployment rate (pp)	0.00 (0.00)	0.35 (0.05)	-0.32 (0.10)	0.00 (0.00)	-0.02 (0.04)	0.16 (0.08)	0.00 (0.00)	0.00 (0.04)	-0.04 (0.08)

Notes: Means over 128 paired shocked and unshocked paths per origin and model, with Monte Carlo standard errors in parentheses. All models share origin states, seeds and the once-only reset before the first quarter's decisions. Fixed sets all six coefficients to one; M_0 and M_1 condition on their saved posterior draws, M_1 exploratorily. Consumer prices are percent deviations of the price level, not the cumulative log inflation quoted in the text. Wages are compensation per employee and include the reset. Responses measure specification sensitivity, not Dutch causal effects.

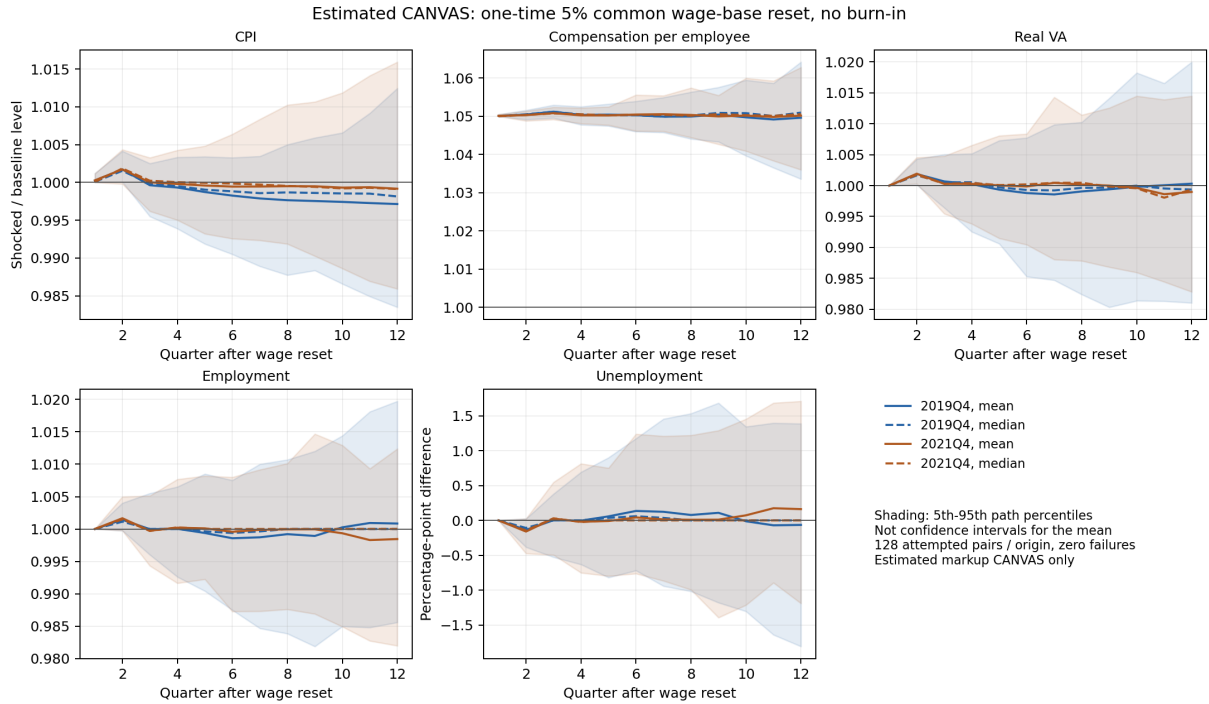


Figure 7: Estimated-baseline response to the common wage-base reset at two Dutch origins. Solid and dashed lines are paired-response means and medians; shading shows the fifth to ninety-fifth path percentiles, not confidence intervals for the mean. Wages measure compensation per employee. Unemployment is a percentage-point difference; the other outcomes are shocked/baseline level ratios. Each origin has 128 attempted pairs and zero failures.

2021Q4, with conditional Monte Carlo standard errors 0.016 and 0.013. Paired with the same seeds, it exceeds the estimated indexation response by 0.338 and 0.180 points (standard errors 0.085 and 0.081). Both estimated specifications therefore imply responses close to zero, far below the fixed-all-ones model; Table 15 shows the same ordering in the consumer-price level and the other outcomes.

7.6 A specification caveat

Estimation moves the modeled response to the common wage reset from about 5.5–5.7 percentage points under fixed coefficients to near zero, and the error-correction wage rule leaves it near zero. Both estimated models nonetheless fail to reproduce the joint wage–price paths of the inflation episode, and both priors already fail to generate them (Section 7.1). The Dutch comparison therefore measures specification sensitivity; it cannot say which response is correct.

We do not interpret the near-zero cost-surprise coefficient as evidence that Dutch firms absorb wage-cost increases. Within the model the coefficient is identified: fitted to data the model generates, the indexation estimator recovers a true value of 0.25 with bias -0.01 and 90% coverage 0.94 (Table 6). The one misspecification studied in the experiment, an omitted error-correction wage rule, biases the estimate upward by $+0.35$ to $+0.39$, not toward zero. A near-zero Dutch estimate that coexists with prior and posterior predictive failures concentrated in the episode therefore points to a misspecification the experiment does not cover, most plausibly in the price block itself: with the estimated coefficients, the model does not generate the episode’s price path, and a small cost coefficient is where the fit settles. We have not isolated the missing mechanism.

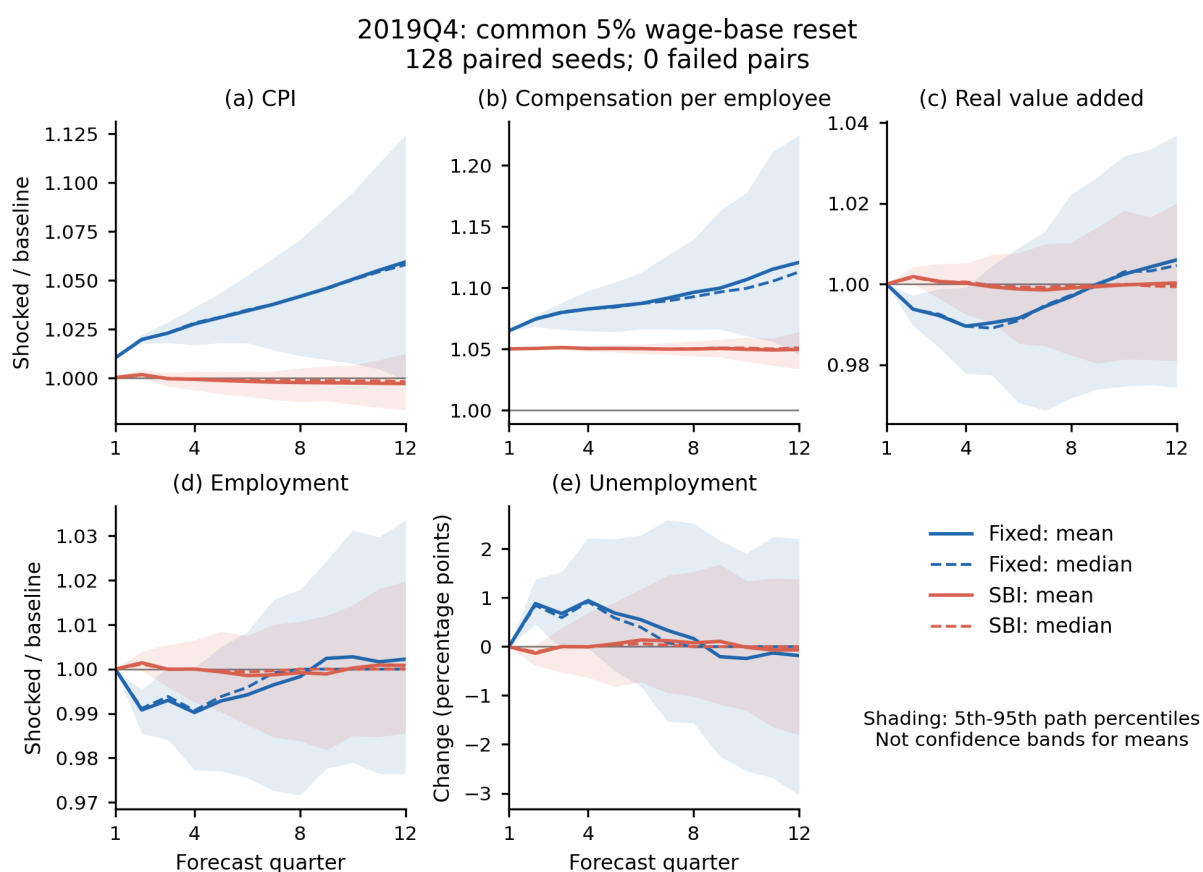


Figure 8: Fixed and estimated baseline wage responses at 2019Q4. Solid and dashed lines are means and medians; shaded bands are fifth to ninety-fifth path percentiles, not mean-confidence bands. Compensation per employee includes the wage-base reset. Unemployment uses percentage-point differences; other outcomes use shocked/baseline level ratios. All six fixed coefficients equal one. There are 128 attempted pairs per arm and zero failures.

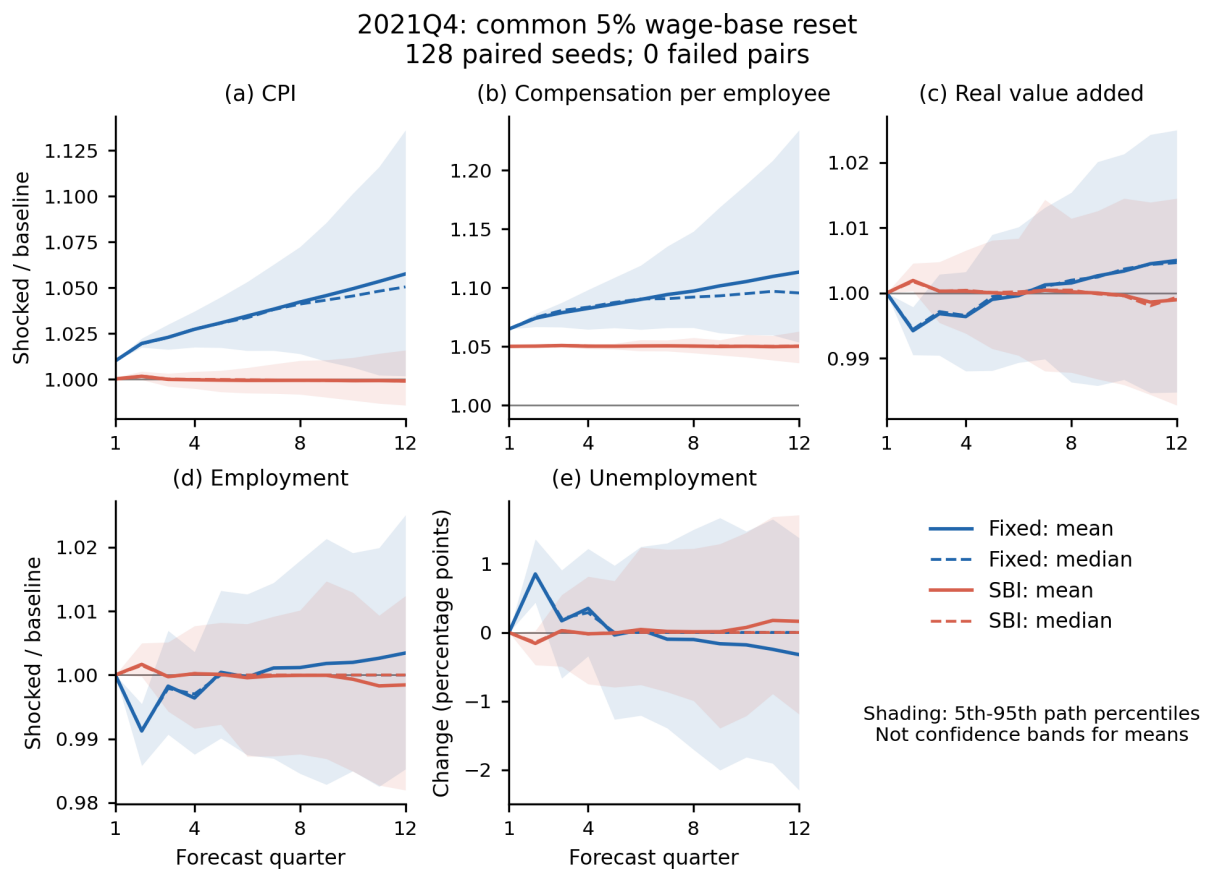


Figure 9: The same matched fixed-versus-estimated response comparison at 2021Q4, with the definitions and uncertainty interpretation of Figure 8. There are 128 attempted pairs per arm and zero failures. The comparison measures calibration sensitivity, not Dutch response accuracy.

The consequence for use is direct. The Dutch responses in Table 15 are diagnostics of this specification, not estimates for policy analysis, and the forecasts in Section 7.4 should be read with the low interval coverage reported there. The methodological results do not depend on this caveat: the known-truth experiment evaluates the estimator and the warnings on model-generated data, and the Dutch application shows those warnings flagging a specification that is not yet adequate.

8 Conclusion

The paper asks whether an estimated cost coefficient records wage–price transmission or compensates for an inadequate wage-setting rule. In a controlled known-truth experiment with 100 independent panels per cell, fitting indexation to error-correction data leaves the estimated break nearly unbiased (-0.03 against $+0.02$ in the correct-rule control) but biases the level of the coefficient by $+0.35$ to $+0.39$ and overstates the twelve-quarter response by $+1.04$ to $+1.10$ percentage points, against at most $+0.06$ in the control. Joint fit and forecast warnings flag every such panel at an achieved control alarm rate of 0.10, and parameter constancy adds nothing to them at that rate for the validated estimator.

The Dutch evidence shows why inference credibility, predictive adequacy and counterfactual reliability must be kept apart. The estimated indexation model passes its calibration check and puts the cost-surprise coefficient near 0.01–0.02, yet both wage models are atypical of the observed panel before and after estimation, and their wage and episode-price intervals have low coverage. Their mean twelve-quarter CPI responses to a common 5% wage reset are close to zero, against about 5.5–5.7 percentage points with all coefficients fixed at one; the two estimated wage rules differ by only 0.180–0.338 percentage points.

The main limit of the empirical results is the specification itself. The cost coefficient is recovered in model-generated data, and the omitted wage rule biases it upward, so its near-zero Dutch value is best read as a symptom of a misspecified price block, not as a measurement of Dutch pass-through. The Dutch wage-shock responses are therefore not suitable for policy analysis. Identifying the missing mechanism, and confronting the price block with direct evidence on margins and unit costs during the episode, is the next step.

Two further limits bound the conclusions. The error-correction estimator fails simulation-based calibration, so its Dutch and synthetic results are exploratory; the main synthetic result rests on the indexation estimator, which passes its calibration screen. And the operating characteristics are local to one wage-rule omission, one break and one shock; they do not establish validity against other misspecifications, including the one the Dutch data point to. A certified model-family test, which would require a valid upper bound on the maximized p -value, remains open.

References

- J. F. Adolfson, M. Ferrari Minesso, J. E. Mork, and I. Van Robays. Gas price shocks and euro area inflation. Technical report, European Central Bank, 2024.
- J. Alvarez, J. C. Bluedorn, N.-J. Hansen, Y. Huang, E. Pugacheva, and A. Sollaci. Wage–price spirals: What is the historical evidence? *Economica*, 91(364):1291–1319, July 2024.

- R. Asghar, J. Fudurich, and J. Voll. Firms' Inflation Expectations and Price-Setting Behaviour in Canada: Evidence from a Business Survey. Technical report, Bank of Canada, 2023.
- R. L. Axtell and J. D. Farmer. Agent-Based Modeling in Economics and Finance: Past, Present, and Future. *Journal of Economic Literature*, 63(1):197–287, Mar. 2025.
- B. Bernanke and O. Blanchard. What Caused the US Pandemic-Era Inflation? Working Paper 31417, National Bureau of Economic Research, June 2023.
- O. Blanchard and L. F. Katz. Wage Dynamics: Reconciling Theory and Evidence. *American Economic Review*, 89(2):69–74, 1999.
- A. Borsos, A. Carro, A. Glielmo, M. Hinterschweiger, J. Kaszowska-Mojas, and A. Uluc. Agent-based modeling at central banks: Recent developments and new challenges. Technical report, Banco de España, Madrid, Feb. 2025.
- A. Caiani, A. Godin, E. Caverzasi, M. Gallegati, S. Kinsella, and J. E. Stiglitz. Agent based-stock flow consistent macroeconomics: Towards a benchmark model. *Journal of Economic Dynamics and Control*, 69:375–408, Aug. 2016.
- M. Clements and D. Hendry. *Forecasting Economic Time Series*. Cambridge University Press, 1998.
- K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, May 2020.
- H. Dawid and D. Delli Gatti. Agent-Based Macroeconomics. In *Handbook of Computational Economics*, volume 4, pages 63–156. Elsevier, 2018.
- H. Dawid, S. Gemkow, P. Harting, S. Van Der Hoog, and M. Neugart. The Eurace@Unibi Model: An Agent-Based Macroeconomic Model for Economic Policy Analysis. *Bielefeld Working Papers in Economics and Management*, 2012.
- H. Dawid, D. Delli Gatti, L. E. Fierro, and S. Poledna. Implications of Behavioral Rules in Agent-Based Macroeconomics, 2024.
- A. Delaunoy, J. Hermans, F. Rozet, A. Wehenkel, and G. Louppe. Towards Reliable Simulation-Based Inference with Balanced Neural Ratio Estimation. In *Advances in Neural Information Processing Systems*, 2022.
- D. Delli Gatti and J. Grazzini. Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic Agent Based Models. *Journal of Economic Behavior & Organization*, 178:875–902, Oct. 2020.
- F. X. Diebold. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests. *Journal of Business & Economic Statistics*, 33(1): 1–9, 2015.
- G. Dosi, G. Fagiolo, and A. Roventini. Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics and Control*, 34(9): 1748–1767, June 2010.
- G. Dosi, G. Fagiolo, M. Napoletano, A. Roventini, and T. Treibich. Fiscal and monetary policies in complex evolving economies. *Journal of Economic Dynamics and Control*, 52:166–189, Mar. 2015.
- J.-M. Dufour. Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, 133(2):443–477, Aug. 2006.

- C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. Neural Spline Flows, June 2019.
- J. Dyer, P. Cannon, J. D. Farmer, and S. M. Schmon. Black-box Bayesian inference for agent-based models. *Journal of Economic Dynamics and Control*, 161:104827, Apr. 2024.
- D. T. Frazier, D. J. Nott, and C. Drovandi. Synthetic Likelihood in Misspecified Models. *Journal of the American Statistical Association*, 120(550):884–895, Sept. 2024.
- J. Grazzini, M. G. Richiardi, and M. Tsionas. Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, 2017.
- J. Hermans, V. Begy, and G. Louppe. Likelihood-free MCMC with Amortized Approximate Ratio Estimators. In *Proceedings of the 37th International Conference on Machine Learning*. PMLR, 2020.
- J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, V. Begy, and G. Louppe. A Trust Crisis In Simulation-Based Inference? Your Posterior Approximations Can Be Unfaithful. *Transactions on Machine Learning Research*, 2022.
- C. Hommes and S. Poledna. Forecasting economic crises: The Great Recession, the sovereign debt crisis, and COVID-19 in the euro area. *Economic Modelling*, 158:107497, May 2026.
- C. Hommes, M. He, S. Poledna, M. Siqueira, and Y. Zhang. CANVAS: A Canadian behavioral agent-based model for monetary policy. *Journal of Economic Dynamics and Control*, 172:104986, Mar. 2025.
- C. Hommes, S. Kozicki, S. Poledna, and Y. Zhang. How an Agent-Based Model Can Support Monetary Policy in a Complex Evolving Economy. In R. M. del Rio-Chanona, M. Pangallo, J. Bednar, E. D. Beinhocker, J. Kaszowska-Mojas, F. Lafond, P. Mealy, A. Pichler, and J. D. Farmer, editors, *The Economy as an Evolving Complex System IV*, pages 449–481. SFI Press, Santa Fe, NM, 2026.
- C. H. Hommes and S. Poledna. Analyzing and Forecasting Economic Crises with An Agent-Based Model of the Euro Area. *SSRN Electronic Journal*, 2023.
- B. J. K. Kleijn and A. W. van der Vaart. Misspecification in infinite-dimensional Bayesian statistics. *The Annals of Statistics*, 34(2):837–877, Apr. 2006.
- R. E. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46, Jan. 1976.
- J.-M. Lueckmann, J. Boelts, D. Greenberg, P. Gonçalves, and J. Macke. Benchmarking simulation-based inference. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*. PMLR, 2021.
- E. Nakamura and J. Steinsson. Five Facts about Prices: A Reevaluation of Menu Cost Models. *Quarterly Journal of Economics*, 123(4):1415–1464, Nov. 2008.
- R. R. Nelson and S. G. Winter. *An Evolutionary Theory of Economic Change*. Harvard University Press, 1982.
- D. Platt. A comparison of economic agent-based model calibration methods. *Journal of Economic Dynamics and Control*, 113:103859, Apr. 2020.
- S. Poledna, M. G. Miess, C. Hommes, and K. Rabitsch. Economic forecasting with an agent-based model. *European Economic Review*, 151:104306, Jan. 2023.
- E. Rubbo. Networks, Phillips Curves, and Monetary Policy. *Econometrica*, 91(4):1417–1455, 2023.

- I. Salle, Y. Gorodnichenko, and O. Coibion. Lifetime Memories of Inflation: Evidence from Surveys and the Lab. Working Paper 31996, National Bureau of Economic Research, Dec. 2023.
- S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman. Validating Bayesian Inference Algorithms with Simulation-Based Calibration, 2018.
- J. Visser. *Wage Bargaining Institutions - from Crisis to Crisis*. European Commission, 2016.
- I. M. Weber and E. Wasner. Sellers' inflation, profits and conflict: Why can large firms hike prices in an emergency? *Review of Keynesian Economics*, 11(2):183–213, Apr. 2023.
- H. White. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25, Jan. 1982.
- S. Wiese, J. Kaszowska-Mojša, J. Dyer, J. Moran, M. Pangallo, F. Lafond, J. Muellbauer, A. Calinescu, and J. D. Farmer. Forecasting Macroeconomic Dynamics using a Calibrated Data-Driven Agent-based Model, 2024.

A Full model

This appendix records the CANVAS accounting, markets and the published CANVAS pricing equation. CANVAS (Hommes et al., 2025) extends Poledna et al. (2023); we recalibrate it for the Netherlands. The quarterly small open economy contains heterogeneous firms, households, a commercial bank, a central bank, government, and the rest of the world. Section 3 gives the pricing, wage and labour-market rules estimated in this paper.

A.1 Firms

The I firms, indexed by i and sector $g \in \{1, \dots, G\}$, carry capital K_i , materials M_i , labour N_i , baseline labour productivity $\bar{\alpha}_i$, capital and materials productivity (κ_i, β_i) , and depreciation δ_i .

A.1.1 Production

Production follows a Leontief technology with capacity utilisation adjustment. Effective labour productivity adjusts for the binding constraint among planned supply, capital capacity, and materials:

$$\alpha_i = \bar{\alpha}_i \min \left(1.5, \frac{\min(Q_i^s, K_i \kappa_i, M_i \beta_i)}{N_i \bar{\alpha}_i} \right), \quad (22)$$

where the factor 1.5 caps the utilisation adjustment to prevent unrealistic productivity spikes. Output is then determined by the Leontief production function:

$$Y_i = \min(Q_i^s, N_i \alpha_i, K_i \kappa_i, M_i \beta_i). \quad (23)$$

A.1.2 Wages

Firm-level wages adjust for the same capacity utilisation factor as productivity:

$$w_i = \bar{w}_i \min \left(1.5, \frac{\min(Q_i^s, K_i \kappa_i, M_i \beta_i)}{N_i \bar{\alpha}_i} \right), \quad (24)$$

where $\bar{w}_i$ is the baseline wage rate for firm i .

A.1.3 Desired Investment, Intermediate Goods, and Employment

Desired capital investment ensures replacement of depreciated capital up to the minimum of planned supply and capital capacity:

$$I_i^d = \frac{\delta_i}{\kappa_i} \min(Q_i^s, K_i \kappa_i). \quad (25)$$

Desired intermediate goods purchases cover the materials needed for the target production level:

$$DM_i^d = \frac{\min(Q_i^s, K_i \kappa_i)}{\beta_i}. \quad (26)$$

Desired employment is the integer-rounded labour requirement:

$$N_i^d = \max \left(1, \left\lfloor \frac{\min(Q_i^s, K_i \kappa_i)}{\bar{\alpha}_i} \right\rfloor \right), \quad (27)$$

where $\lfloor \cdot \rfloor$ denotes rounding to the nearest integer.

A.1.4 Quantity and Price Setting

Firms set planned production Q_i^s and price P_i from the previous period's demand–supply signals, where Q_i^d is the demand firm i faced, $Q_i^o = Y_i + S_i$ is offered output including inventories, and $\bar{P}_g^{\text{dom}}$ is the domestic realised-sales price index for the sector. Each firm forms a demand-pull signal π_i^d and a quantity-growth signal γ_i^d according to whether offered output ran above or below demand and whether its price sat above or below the sector average:

$$\begin{cases} \gamma_i^d = Q_i^d / Q_i^o - 1, & \pi_i^d = 0 & \text{if } (Q_i^o \leq Q_i^d \text{ and } P_i \geq \bar{P}_g^{\text{dom}}) \text{ or } (Q_i^o > Q_i^d \text{ and } P_i < \bar{P}_g^{\text{dom}}), \\ \gamma_i^d = 0, & \pi_i^d = Q_i^d / Q_i^o - 1 & \text{if } (Q_i^o \leq Q_i^d \text{ and } P_i < \bar{P}_g^{\text{dom}}) \text{ or } (Q_i^o > Q_i^d \text{ and } P_i \geq \bar{P}_g^{\text{dom}}). \end{cases} \quad (28)$$

Planned supply incorporates both aggregate expected growth γ^e and firm-specific demand signals:

$$Q_i^s = \max \left\{ 0, Q_i^o \left(1 + \phi_Q^G \gamma^e \right) \left(1 + \phi_Q^{DP} \gamma_i^d \right) - S_i \right\}, \quad (29)$$

where ϕ_Q^G and ϕ_Q^{DP} are sensitivity parameters.

The absolute unit cost of production at current input prices is:

$$AC_i = \underbrace{\frac{(1 + \tau^{SIF})\bar{w}_i}{\bar{\alpha}_i} P^W}_{\text{labour cost}} + \underbrace{\frac{1}{\beta_i} \sum_g a_{sg} \bar{P}_g}_{\text{material cost}} + \underbrace{\frac{\delta_i}{\kappa_i} P^{CF}}_{\text{capital cost}}, \quad (30)$$

where a_{sg} is the input–output coefficient from sector g to sector s , the sector of firm i , P^W is the wage price index, P^{CF} is the capital formation price index, and τ^{SIF} is the employer social insurance contribution rate.

$$AC_i^e = AC_i \left(1 + \phi^{AE} \pi^e \right), \quad (31)$$

where π^e is the aggregate expected inflation rate and ϕ^{AE} is a sensitivity parameter.

For reference, the published sector-relative CANVAS cost-pressure signal is

$$\pi_i^c = \frac{(1 + \tau^{SIF})\bar{w}_i}{\bar{\alpha}_i} \left(\frac{P^{HH}}{\bar{P}_s} - 1 \right) + \frac{1}{\beta_i} \left(\frac{\sum_g a_{sg} \bar{P}_g}{\bar{P}_s} - 1 \right) + \frac{\delta_i}{\kappa_i} \left(\frac{P^{CF}}{\bar{P}_s} - 1 \right), \quad (32)$$

where s is firm i 's sector. This reference uses input-price indices relative to the sector price, not the individual firm's output price. The corresponding multiplicative price rule is

$$P_{i,t+1} = P_i (1 + \phi^{CP} \pi_i^c) (1 + \phi^{DP} \pi_i^d) (1 + \pi^e). \quad (33)$$

A wage-base reset changes the coefficient on the labour-price wedge, not its reference-price ratio. A small wedge can therefore produce a weak direct response, without implying zero response in every economy. Neither controlled model uses (33): both use the decision-cost markup rule (6). The expected-cost variable (31) does not add another expectation multiplier to that operational price rule.

A.1.5 Profits

Firm profits are:

$$\begin{aligned} \Pi_i = & P_i Q_i + P_i \Delta S_i + r^{bar} \max(0, D_i) \\ & - (1 + \tau^{SIF}) w_i N_i P^W - \frac{1}{\beta_i} \bar{P}_i Y_i - \frac{\delta_i}{\kappa_i} P_i^{CF} Y_i \\ & - \tau_i^Y P_i Y_i - \tau_i^K P_i Y_i - r(L_i + \max(0, -D_i)), \end{aligned} \quad (34)$$

where Q_i denotes actual sales, $\Delta S_i = Y_i - Q_i$ is the change in inventories, P^W is the wage index used in the paper's wage-ratchet configuration, $\bar{P}_i$ is the effective input price paid by firm i , P_i^{CF} is the effective capital goods price paid, τ_i^Y and τ_i^K are product and production tax rates, r is the bank lending rate, r^{bar} is the central bank policy rate, L_i is outstanding loans, and D_i is deposits.

A.1.6 Deposit Dynamics

The flow-of-funds identity for firm deposits is:

$$\begin{aligned}
D_{i,t+1} = & D_i + P_i Q_i - (1 + \tau^{SIF}) w_i N_i P^W - DM_i \bar{P}_i \\
& - \tau_i^Y P_i Y_i - \tau_i^K P_i Y_i \\
& - \tau^{FIRM} \max(0, \Pi_i) - \theta^{DIV} (1 - \tau^{FIRM}) \max(0, \Pi_i) \\
& - r(L_i + \max(0, -D_i)) + r^{bar} \max(0, D_i) \\
& - P_i^{CF} I_i + DL_i - \theta L_i,
\end{aligned} \tag{35}$$

where τ^{FIRM} is the corporate tax rate, θ^{DIV} is the dividend payout ratio, θ is the loan repayment rate, DL_i is new credit obtained, I_i is actual investment, and DM_i is actual intermediate goods purchased.

A.1.7 Capital, Materials, and Inventories

Stocks evolve according to:

$$K_{i,t+1} = K_i - \frac{\delta_i}{\kappa_i} Y_i + I_i, \tag{36}$$

$$M_{i,t+1} = M_i - \frac{Y_i}{\beta_i} + DM_i, \tag{37}$$

$$S_{i,t+1} = S_i + Y_i - Q_i. \tag{38}$$

A.1.8 Equity

Firm equity is:

$$E_i = D_i + M_i \sum_g a_{s_i g} \bar{P}_g + P_i S_i + P^{CF} K_i - L_i, \tag{39}$$

where s_i denotes the sector of firm i , so $a_{s_i g}$ is the coefficient from sector g into that sector, and P^{CF} is the aggregate capital price index.

A.1.9 Loans

Outstanding loans evolve as:

$$L_{i,t+1} = (1 - \theta) L_i + DL_i. \tag{40}$$

A.1.10 Expectations and Desired Loans

Expected profits scale current profits by expected nominal growth:

$$\Pi_i^e = \Pi_i (1 + \pi^e) (1 + \gamma^e). \tag{41}$$

Expected changes in deposits, capital, and loans are:

$$DD_i^e = \Pi_i^e - \theta L_i - \tau^{FIRM} \max(0, \Pi_i^e) - \theta^{DIV} (1 - \tau^{FIRM}) \max(0, \Pi_i^e), \quad (42)$$

$$K_i^e = P^{CF} (1 + \pi^e) K_i, \quad (43)$$

$$L_i^e = L_i. \quad (44)$$

Desired new loans close any expected financing gap:

$$DL_i^d = \max(0, -DD_i^e - D_i). \quad (45)$$

A.1.11 Insolvency

A firm is declared insolvent when both $D_i < 0$ and $E_i < 0$. Upon insolvency, the bank absorbs the loss and recapitalises the firm:

$$\Delta E_k = -\left(L_i - D_i - \zeta_b P^{CF} K_i\right), \quad (46)$$

where ζ_b is the loan-to-capital ratio for the recapitalised firm. The firm's deposits are reset to zero and its loans to $\zeta_b P^{CF} K_i$.

A.2 Households

The H persons comprise H^W workers, I firm investors, H^{inact} inactive persons, and one bank owner. Labour matching endogenously partitions workers into employed H_t^E and unemployed H_t^U . Investors and the bank owner receive dividends; everyone consumes.

A.2.1 Activity Status and the Reference Wage

Each worker h carries a reference wage w_h , set to her employer's wage and held fixed at the last wage earned while she is unemployed:

$$w_h(t) = \begin{cases} w_i(t) & \text{if employed by firm } i, \\ w_h(t-1) & \text{if unemployed.} \end{cases} \quad (47)$$

The unemployment benefit pays a fraction θ^{UB} of this reference wage, and re-employment resets it to the new employer's wage (Equation (48)).

A.2.2 Income

Income depends on employment status. Let $O_h \in \{0, 1, \dots, I\}$ denote the employment status of worker h , where $O_h = 0$ indicates unemployment and $O_h = i > 0$ indicates employment at firm i :

$$Y_h^{act} = \begin{cases} w_h(1 - \tau^{SIW} - \tau^{INC}(1 - \tau^{SIW}))P^W + sb^{other} P^{HH} & \text{if } O_h \neq 0, \\ [\theta^{UB} w_h + sb^{other}] P^{HH} & \text{if } O_h = 0, \end{cases} \quad (48)$$

where τ^{SIW} is the employee social insurance contribution rate, τ^{INC} is the income tax rate, θ^{UB} is the unemployment benefit replacement rate, and sb^{other} is a general social benefit.

$$Y_h^{inact} = (sb^{inact} + sb^{other}) P^{HH}, \quad (49)$$

where sb^{inact} is the social benefit for inactive individuals.

Firm owner h (associated with firm i) receives dividend income:

$$Y_h^{firm} = \theta^{DIV} (1 - \tau^{INC}) (1 - \tau^{FIRM}) \max(0, \Pi_i) + sb^{other} P^{HH}. \quad (50)$$

$$Y_h^{bank} = \theta^{DIV} (1 - \tau^{INC}) (1 - \tau^{FIRM}) \max(0, \Pi^k) + sb^{other} P^{HH}, \quad (51)$$

where Π^k denotes bank profits.

A.2.3 Consumption and Investment Demand

All household types share the same consumption heuristic. Expected income Y_h^e applies the same type-specific formula using expected nominal wages and transfers and, for firm and bank owners, expected profits. Consumption and housing investment demand are fixed fractions of expected income:

$$C_h^d = \frac{\psi Y_h^e}{1 + \tau^{VAT}}, \quad (52)$$

$$I_h^d = \frac{\psi_H Y_h^e}{1 + \tau^{CF}}, \quad (53)$$

where ψ is the propensity to consume, ψ_H is the propensity to invest in housing, τ^{VAT} is the value-added tax rate, and τ^{CF} is the capital formation tax rate.

A.2.4 Household Allocation and Capital Stock

Household nominal budgets are allocated across goods using price-normalised expenditure shares,

$$\omega_g^{HH} = \frac{\bar{P}_g b_g^{HH}}{\sum_{g'} \bar{P}_{g'} b_{g'}^{HH}}, \quad \omega_g^{CFH} = \frac{\bar{P}_g b_g^{CFH}}{\sum_{g'} \bar{P}_{g'} b_{g'}^{CFH}}, \quad (54)$$

where b_g^{HH} and b_g^{CFH} are the underlying real-basket coefficients. Consumption and housing-investment demands by good are

$$C_{h,g}^d = \omega_g^{HH} C_h^d, \quad I_{h,g}^d = \omega_g^{CFH} I_h^d, \quad (55)$$

and realised purchases $I_h = \sum_g I_{h,g}$ are the outcome of search-and-matching on the capital-goods market, which may ration them below demand. The household's stock of housing capital then accumulates without depreciation:

$$K_{h,t+1} = K_h + I_h. \quad (56)$$

A.2.5 Deposit Dynamics

Household deposits evolve according to:

$$D_{h,t+1} = D_h + Y_h - (1 + \tau^{VAT})C_h - (1 + \tau^{CF})I_h + r^{bar} \max(0, D_h) - r \max(0, -D_h), \quad (57)$$

where C_h and I_h denote actual consumption and housing investment (which may be rationed relative to demand).

A.3 Banking Sector

The economy contains a single commercial bank.

A.3.1 Bank Rate

The bank sets its lending rate as a fixed spread μ above the policy rate:

$$r = r^{bar} + \mu. \quad (58)$$

A.3.2 Bank Profits

Bank profits arise from the spread between lending and deposit rates:

$$\begin{aligned} \Pi^k = & r \sum_i (L_i + \max(0, -D_i)) + r \sum_h \max(0, -D_h) \\ & + r^{bar} \max(0, D_k) - r^{bar} \sum_i \max(0, D_i) \\ & - r^{bar} \sum_h \max(0, D_h) - r^{bar} \max(0, -D_k), \end{aligned} \quad (59)$$

where D_k is the bank's own deposit position and the sums over h include all household types (active workers, inactive workers, firm owners, and bank owner).

A.3.3 Bank Equity

Bank equity evolves as:

$$E_{k,t+1} = E_k + \Pi^k - \theta^{DIV}(1 - \tau^{FIRM}) \max(0, \Pi^k) - \tau^{FIRM} \max(0, \Pi^k). \quad (60)$$

A.3.4 Bank Deposits

The bank's deposit position is determined by the balance sheet identity:

$$D_k = \sum_i D_i + \sum_h D_h + E_k - \sum_i L_i, \quad (61)$$

where the sum over h includes all household types.

A.3.5 Expected Profits

$$\Pi^{e,k} = \Pi^k(1 + \pi^e)(1 + \gamma^e). \quad (62)$$

A.4 Central Bank

A.4.1 Taylor Rule

The central bank sets the policy rate according to a Taylor rule with interest rate smoothing:

$$r_t^{bar} = \rho r_{t-1}^{bar} + (1 - \rho) \left(r^* + \pi^* + \xi_\pi (\pi_t^{EA} - \pi^*) + \xi_\gamma \gamma_t^{EA} \right), \quad (63)$$

where $\rho \in [0, 1)$ is the smoothing parameter, r^* is the real equilibrium rate, π^* is the inflation target, ξ_π and ξ_γ are the weights on inflation and output growth respectively, π_t^{EA} is Euro Area inflation, and γ_t^{EA} is Euro Area output growth. The policy rate is floored at zero: $r_t^{bar} = \max(0, r_t^{bar})$.

A.4.2 Central Bank Profits and Equity

Central bank profits are:

$$\Pi^{CB} = r^G L^G - r^{bar} D_k, \quad (64)$$

where r^G is the interest rate on government bonds and L^G is government debt held by the central bank.

Central bank equity evolves as:

$$E_{t+1}^{CB} = E_t^{CB} + \Pi^{CB}. \quad (65)$$

A.5 Government

A.5.1 Government Consumption

Government consumption follows an AR(1) process in logs:

$$\log C_t^G = \alpha_G \log C_{t-1}^G + \beta_G + \epsilon_t^G, \quad \epsilon_t^G \sim \mathcal{N}(0, \sigma_G^2), \quad (66)$$

where α_G , β_G , and σ_G are estimated from historical data. The AR(1) state C^G is converted into total nominal demand for each government entity j :

$$C_j^d = \frac{C^G}{J} \sum_{g'} c_{g'}^G \bar{P}_{g'} (1 + \pi^e), \quad (67)$$

where c_g^G are real basket weights and J is the number of government entities. The retail matcher subsequently allocates this nominal total across goods as $C_{j,g}^d = \omega_g^G C_j^d$, with price-weighted share

$$\omega_g^G = \frac{\bar{P}_g c_g^G}{\sum_{g'} \bar{P}_{g'} c_{g'}^G}. \quad (68)$$

A.5.2 Government Revenue

Total government revenue comprises:

$$\begin{aligned}
Y^G = & (\tau^{SIF} + \tau^{SIW}) \cdot W^{emp} \cdot P^W + \tau^{INC}(1 - \tau^{SIW}) \cdot W^{emp} \cdot P^W \\
& + \tau^{VAT} \sum_h C_h + \tau^{CF} \sum_h I_h \\
& + \tau^{INC}(1 - \tau^{FIRM})\theta^{DIV} \left(\sum_i \max(0, \Pi_i) + \max(0, \Pi^k) \right) \\
& + \tau^{FIRM} \left(\sum_i \max(0, \Pi_i) + \max(0, \Pi^k) \right) \\
& + \sum_i (\tau_i^Y + \tau_i^K) P_i Y_i + \tau^{EXP} C_l^{RoW},
\end{aligned} \tag{69}$$

where $W^{emp} = \sum_{h:O_h \neq 0} w_h$ is the total wage bill of employed workers and τ^{EXP} is the export tax rate.

A.5.3 Social Benefits

Social benefit rates grow with expected real growth:

$$sb_{t+1}^{other} = sb_t^{other} (1 + \gamma^e), \tag{70}$$

$$sb_{t+1}^{inact} = sb_t^{inact} (1 + \gamma^e). \tag{71}$$

A.5.4 Government Debt

Government debt evolves according to:

$$L_{t+1}^G = L_t^G + H^{inact} \cdot sb^{inact} \cdot P^{HH} + \theta^{UB} W^{unemp} P^{HH} + H \cdot sb^{other} \cdot P^{HH} + C_{real}^G + r^G L_t^G - Y^G, \tag{72}$$

where $W^{unemp} = \sum_{h:O_h=0} w_h$ is the total reference wage of unemployed workers and $C_{real}^G = \sum_j C_j^G$ is realised aggregate government consumption expenditure.

A.6 Rest of the World

The rest of the world (RoW) sector captures international trade and links the domestic economy to the Euro Area.

A.6.1 Euro Area Growth and Inflation

Euro Area GDP and inflation follow AR(1) processes in logs:

$$\log Y_t^{EA} = \alpha_Y^{EA} \log Y_{t-1}^{EA} + \beta_Y^{EA} + \epsilon_{Y,t}^{EA}, \tag{73}$$

$$\log(1 + \pi_t^{EA}) = \alpha_\pi^{EA} \log(1 + \pi_{t-1}^{EA}) + \beta_\pi^{EA} + \epsilon_{\pi,t}^{EA}, \tag{74}$$

with corresponding growth rate $\gamma_t^{EA} = Y_t^{EA} / Y_{t-1}^{EA} - 1$.

A.6.2 Export Demand

Aggregate export demand follows an AR(1) process in logs:

$$\log C_t^E = \alpha_E \log C_{t-1}^E + \beta_E + \epsilon_t^E, \quad (75)$$

where the export innovation is drawn jointly with the euro-area GDP and import-supply innovations. The export state C^E is converted into total nominal demand for each export partner l :

$$C_l^d = \frac{C^E}{L} \sum_{g'} c_{g'}^E \bar{P}_{g'} (1 + \pi^e), \quad (76)$$

where c_g^E are real export-basket weights and L is the number of export partners. The retail matcher allocates this total across goods as $C_{l,g}^d = \omega_g^E C_l^d$, where

$$\omega_g^E = \frac{\bar{P}_g c_g^E}{\sum_{g'} \bar{P}_{g'} c_{g'}^E}. \quad (77)$$

A.6.3 Import Supply

Aggregate import supply also follows an AR(1) process:

$$\log Y_t^I = \alpha_I \log Y_{t-1}^I + \beta_I + \epsilon_t^I. \quad (78)$$

Import supply per sector is:

$$Y_{m,g,t}^I = c_g^I Y_t^I, \quad P_{m,g,t}^I = \begin{cases} \bar{P}_{g,t} (1 + \pi_t^{EA}), & \text{when the import-price state is initialised,} \\ P_{m,g,t-1}^I (1 + \pi_t^{EA}), & \text{thereafter,} \end{cases} \quad (79)$$

where c_g^I are import composition shares. Import prices are seeded from the domestic sector price only when their state is zero and subsequently compound at euro-area inflation π^{EA} rather than the domestic expectation, so the domestic expectations channel does not double as an import-price channel.

A.6.4 Rest of World Deposits

The RoW net financial position evolves as:

$$D_{t+1}^{RoW} = D_t^{RoW} + \sum_g P_{m,g}^I Q_{m,g}^I - (1 + \tau^{EXP}) C_l^{RoW}, \quad (80)$$

where $Q_{m,g}^I$ is the actual quantity of imports sold and C_l^{RoW} is total export expenditure.

A.6.5 Financial Closure

Stock-flow consistency closes the financial system through a single accounting identity that links central-bank equity E^{CB} , the net foreign position D^{RoW} , government debt L^G , and the

bank's reserve position D_k :

$$E_t^{CB} + D_t^{RoW} = L_t^G - D_{k,t} = L_t^G - \sum_i D_i - \sum_h D_h - E_k + \sum_i L_i, \quad (81)$$

where the second equality substitutes the bank's balance-sheet position (61). Because every nominal flow has a balance-sheet counterpart, this identity holds in every period: a government deficit or a trade imbalance must be matched, euro for euro, by an offsetting change in private deposits, bank reserves, or the net foreign asset position.

A.7 Market Clearing

The model features three markets, each cleared through decentralised search-and-matching protocols.

A.7.1 Goods Market

Goods trade first among firms for intermediate goods and capital, then at retail for consumption, government, and exports. For each $g \in \{1, \dots, G\}$, domestic firms and imports $Y_{m,g}^I$ are sampled by price competitiveness and size:

$$w_f = \frac{\exp(-2P_f)}{\sum_{f' \in \mathcal{F}_g} \exp(-2P_{f'})} + \frac{A_f}{\sum_{f' \in \mathcal{F}_g} A_{f'}}, \quad (82)$$

where $\mathcal{F}_g$ is the set of suppliers with positive availability, P_f is the supplier's price and A_f is available supply. For a domestic firm $A_f = Y_f + S_f$; for the foreign supplier it is import supply. The two components are normalised separately before the sampler normalises their sum.

Firm demand is $a_{sig} DM_i^d + b_g^{CF} I_i^d$. Sequential buyers resample when stock is exhausted; a second round accesses excess capacity. Retail demand is $\omega_g^{HH} C_h^d + \omega_g^{CFH} I_h^d$ for households, $\omega_g^G C_j^d$ for government, and $\omega_g^E C_l^d$ for exports. Purchases from remaining supply determine possibly rationed consumption and investment.

Firm-level sales are determined as:

$$Q_i = \min(Y_i + S_i, Q_i^d), \quad (83)$$

where Q_i^d is the total demand directed to firm i from all sources.

A.7.2 Labour Market

Firms with $V_i = N_i^d - N_i < 0$ dismiss randomly selected workers; those with $V_i > 0$ post vacancies. In M_0 , random matching continues until vacancies or unemployed workers are exhausted, then updates N_i and O_h . The controlled M_1 uses the same matching rule. The broader search variant instead uses the urn-ball market with recall of Appendix B, in which each unemployed worker sends on average n_{SEARCH} applications.

A.7.3 Credit Market

Firms with positive desired loan demand ($DL_i^d > 0$) approach the bank in random order. The bank provides credit subject to two constraints:

Under the loan-to-value constraint the firm's total expected debt cannot exceed a fraction ζ^{LTV} of its expected capital value:

$$DL_i \leq \zeta^{LTV} K_i^e - L_i^e. \quad (84)$$

Under the capital adequacy constraint the bank's total lending cannot exceed a multiple of its equity:

$$\sum_i L_i^e + \sum_i DL_i \leq \frac{E_k}{\zeta}. \quad (85)$$

Actual credit received by firm i is:

$$DL_i = \max \left(0, \min \left(DL_i^d, \zeta^{LTV} K_i^e - L_i^e, \frac{E_k}{\zeta} - \sum_j L_j^e - \sum_{j < i} DL_j \right) \right), \quad (86)$$

where the ordering over $j < i$ reflects the sequential (random) processing of firms. Any difference $DL_i^d - DL_i > 0$ is recorded as a denied loan.

A.8 Aggregate Variables and Price Indices

A.8.1 Expectations

Aggregate growth and inflation expectations are formed via AR(1) estimation on the history of log-transformed variables:

$$\log(1 + \gamma_t^e) = \hat{\alpha}_\gamma \log(1 + \gamma_{t-1}) + \hat{\beta}_\gamma + \epsilon_{\gamma,t}, \quad (87)$$

$$\log(1 + \pi_t^e) = \hat{\alpha}_\pi \log(1 + \pi_{t-1}) + \hat{\beta}_\pi + \epsilon_{\pi,t}, \quad (88)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are re-estimated each period from all available historical data, and the error terms $\epsilon_{\gamma,t}$ and $\epsilon_{\pi,t}$ are drawn jointly to preserve their empirical correlation.

A.8.2 Aggregate Output Used for Expectations

The aggregate history used by the adaptive real-growth expectation is total current production:

$$Y_t^{\text{prod}} = \sum_i Y_{i,t}. \quad (89)$$

Its growth rate is:

$$\gamma_t = \frac{Y_t^{\text{prod}}}{Y_{t-1}^{\text{prod}}} - 1. \quad (90)$$

This state is an output proxy, not recorded nominal GDP. The forecast target for nominal GDP is constructed separately from the model's value-added, product-tax and expenditure identities.

A.8.3 Price Level and Inflation

The aggregate price level used for inflation expectations is the domestic production-weighted price index:

$$\bar{P}_t = \frac{\sum_i P_{i,t} Y_{i,t}}{\sum_i Y_{i,t}}, \quad (91)$$

and inflation is:

$$\pi_t = \frac{\bar{P}_t}{\bar{P}_{t-1}} - 1. \quad (92)$$

A.8.4 Sector Price Indices

The import-inclusive price index for sector g weights realised domestic sales and realised imports, while the quadrant rules in Equations (28) and (3) use the domestic-only index:

$$\bar{P}_{g,t} = \frac{\sum_{i \in g} P_{i,t} Q_{i,t} + P_{m,g,t}^I Q_{m,g,t}^I}{\sum_{i \in g} Q_{i,t} + Q_{m,g,t}^I}, \quad \bar{P}_{g,t}^{\text{dom}} = \frac{\sum_{i \in g} P_{i,t} Q_{i,t}}{\sum_{i \in g} Q_{i,t}}, \quad (93)$$

where the sum $\sum_{i \in g}$ runs over all firms in sector g .

A.8.5 Composite Price Indices

The capital formation and household consumption price indices are weighted averages of sector prices:

$$P^{CF} = \sum_g b_g^{CF} \bar{P}_g, \quad (94)$$

$$P^{HH} = \sum_g b_g^{HH} \bar{P}_g, \quad (95)$$

where b_g^{CF} and b_g^{HH} are the capital formation and household consumption basket weights, respectively.

A.8.6 Wage Index

The aggregate wage index updates after current-period inflation is observed and enters the next quarter's labour cost:

$$P_{t+1}^W = \left(1 + \phi^U \max(\pi_t, 0)\right) P_t^W, \quad (96)$$

where ϕ^U , written ϕ_{UNION} in the main text, governs the degree of wage indexation to prices. This is the wage rule of M_0 : higher prices lead to higher negotiated wages, which feed back into firm costs and hence into prices. In M_1 the wage index follows the error-correction equation (9) instead.

A.9 Summary of Notation

Table 16 summarises the key variables and parameters used throughout the model.

Table 16: **Summary of notation.** Agent-level variables, aggregate and policy variables, and parameters of the model.

Symbol	Description
<i>Firm variables</i>	
Y_i	Output
Q_i^s, Q_i^d, Q_i	Planned supply, demand, actual sales
K_i, M_i, S_i	Capital, materials, inventories
N_i, N_i^d	Employment, desired employment
P_i	Output price
D_i, L_i, E_i	Deposits, loans, equity
Π_i, Π_i^e	Profits, expected profits
I_i, I_i^d	Actual and desired investment
DM_i, DM_i^d	Actual and desired intermediate goods
DL_i, DL_i^d	Actual and desired new loans
$w_i, \bar{w}_i$	Actual and baseline wage
$\alpha_i, \bar{\alpha}_i$	Effective and baseline labour productivity
$\kappa_i, \beta_i, \delta_i$	Capital productivity, materials productivity, depreciation
<i>Household variables</i>	
Y_h, Y_h^e	Income, expected income
C_h, C_h^d	Actual and desired consumption
I_h, I_h^d	Actual and desired housing investment
D_h, K_h	Deposits, housing capital stock
w_h	Reference (last-employed) wage
O_h	Employment status
H_t^E, H_t^U	Employed and unemployed worker sets
<i>Aggregate and policy variables</i>	
$\bar{P}, \bar{P}_g$	Aggregate and sector price indices
p^{HH}, p^{CF}, p^W	Household, capital formation, wage price indices
π, π^e	Inflation, expected inflation
γ, γ^e	Aggregate-production growth and its expectation
r^{bar}, r	Central bank rate, bank lending rate
L^G, E^{CB}, D^{RoW}	Government debt, central bank equity, net foreign position
<i>Parameters</i>	
$\tau^{SIF}, \tau^{SIW}, \tau^{INC}$	Employer SI, employee SI, income tax rates
$\tau^{FIRM}, \tau^{VAT}, \tau^{CF}$	Corporate, VAT, capital formation tax rates
$\tau_i^Y, \tau_i^K, \tau^{EXP}$	Product, production, export tax rates
$\theta, \theta^{DIV}, \theta^{UB}$	Loan repayment, dividend payout, UB replacement rates
ψ, ψ_H	Consumption and housing investment propensities
$\phi_{CP}, \phi_{DP}, \phi_{AE}$	Cost-surprise, demand-pull and expectation responses
ϕ_{DP_Q}, ϕ_{G_Q}	Quantity responses to firm demand and expected growth
ϕ_{UNION}	Wage indexation (M_0)
$\beta_{IDX}, \lambda_{RW}, \delta_U$	Wage catch-up, labour-share correction, wage Phillips slope (M_1)
ζ_{RD}, n_{SEARCH}	Innovation intensity and search applications (broader variants)
ρ, r^*, π^*	Taylor rule smoothing, equilibrium rate, inflation target
ξ_π, ξ_γ	Taylor rule weights on inflation and growth
μ	Bank spread
$\zeta, \zeta^{LTV}, \zeta_b$	Capital adequacy, LTV, and recapitalisation parameters
a_{sg}	Input-output coefficients
b_{sg}^{HH}, b_{sg}^{CF}	Consumption and investment basket weights
b_{sg}^{CFH}	Household housing-investment basket weights
c_g^E, c_g^I, c_g^G	Export, import, government composition shares

B Diagnostic model variants

The controlled comparison changes only the wage equation. Two broader variants were estimated diagnostically to check whether search frictions and productivity growth change the wage-block identification problem. They are not part of the eight-cell design, and their estimators fail the same calibration gate as M_1 (Appendix D).

In the broader search-friction diagnostic variant, hiring passes through an urn-ball market with recall. Each unemployed worker sends on average n_{SEARCH} applications to distinct vacancies chosen at random, with fractional values rounded at random. Each vacancy with applicants makes an offer to one of them, and each worker holding several offers accepts one at random. A vacancy whose offer was declined then offers to another still-unemployed applicant, in further rounds, until it is filled or has no applicants left. The recall rounds reduce losses from multiple offers to the same worker. When n_{SEARCH} is low, many vacancies receive no applications and stay open, so unemployment and vacancies coexist. When it is high, the market approaches the frictionless limit of M_0 .

A broader diagnostic variant also adds productivity growth; it is not part of the one-wage-equation M_1 . In M_0 the productivity anchor $\bar{a}_i$ and the wage anchor $\bar{w}_i$ stay at their initial values, so over a multi-year window the price, quantity and wage coefficients must account for paths that also contain productivity and real-wage trends. Following the evolutionary tradition of Nelson and Winter (1982), each quarter firm i innovates with probability

$$p_i = 1 - \exp(-\zeta_{RD} I_i / K_i), \quad (97)$$

where I_i / K_i is the firm's gross investment rate and $\zeta_{RD} \geq 0$ the productivity content of investment. Technical change is embodied in installed capital, and research expenditure is already part of gross fixed capital formation under ESA 2010, so the channel adds no expenditure flow. On success the productivity anchor rises permanently, $\bar{a}_i \leftarrow (1 + x)\bar{a}_i$ with $x \sim U(0, 0.10)$, and the wage anchor rises by the same share, so the firm-level labour share is unchanged, as in the Keynes-meets-Schumpeter family (Dosi et al., 2010). Setting $\zeta_{RD} = 0$ reproduces the model without growth.

The second broader variant adds a fixed labour-force-growth rule. Both broader variants estimate ten coefficients: the eight of the controlled wage model plus $\zeta_{RD} \in [0, 5]$ and $n_{SEARCH} \in [0.05, 1.5]$. Their additional rules are not treated as a validated Dutch remedy.

C Calibration details

C.1 Data Sources

Table 2 in Section 4 lists the principal sources.

All series are taken from one fully revised vintage. Two corrections are required for exact replication. The Netherlands initial conditions distributed on Zenodo contain a stale fixed census count for unemployed workers; the analysis replaces it with the value verified against Eurostat. The observed-series dates are Julia date numbers rather than Python ordinals. The estimation grid therefore uses exact quarter lookup and verifies the lead-zero identity for every table, avoiding the four-quarter shift caused by treating the two date conventions as equivalent.

C.2 Macroeconomic Parameters

Table 17 reports the key macroeconomic parameters together with their definitions and sources.

Table 17: **Macroeconomic parameters.** Definitions of the parameters computed from external data and the sources from which they are calculated.

Parameter	Description	Source
<i>Production technology</i>		
s^{pop}	Simulation population scale	0.001 of the national population
a_{sg}	Input-output coefficient from sector g to sector s	FIGARO tables
$\bar{\alpha}_s$	Labour productivity, sector s	Quarterly gross output / employees
$\bar{\kappa}_s$	Capital productivity, sector s	Quarterly gross output / ($\omega \times$ fixed assets excluding dwellings)
$\bar{\beta}_s$	Materials productivity, sector s	Gross output / intermediate consumption
δ_s	Quarterly depreciation rate, sector s	Quarterly consumption of fixed capital / ($\omega \times$ fixed assets excluding dwellings)
<i>Tax rates</i>		
τ^{INC}	Personal income tax	National accounts
τ^{SIF}	Employer social insurance contributions	National accounts
τ^{SIW}	Employee social insurance contributions	National accounts
τ^{FIRM}	Corporate income tax	National accounts
τ^{VAT}	Value-added tax	National accounts
τ^{CF}	Tax or subsidy on capital formation	National accounts
τ_s^Y	Output tax, sector s	National accounts
τ_s^K	Capital tax, sector s	National accounts
<i>Financial parameters</i>		
θ	Loan repayment rate	Fixed at 0.05
θ^{DIV}	Dividend payout ratio	Sector accounts
θ^{UB}	Unemployment benefit replacement rate	Gross replacement of 0.55, net of τ^{INC} and τ^{SIW}
<i>Central bank</i>		
ρ	Interest-rate smoothing	Taylor-rule regression
r^*	Natural rate of interest	Taylor-rule regression
π^*	Inflation target	2% a year (ECB target)
ζ_π	Response to the inflation gap	Taylor-rule regression
ζ_γ	Response to output growth	Taylor-rule regression

Notes: Except for s^{pop} , θ , the gross benefit replacement and π^* , values are origin specific: they are recomputed from the sources in Table 2 with data up to each forecast origin. The Taylor-rule regression relates the three-month Euribor to euro-area inflation and real GDP growth from 1996Q4 to the origin. The scaling constant ω is fixed at 0.85.

C.3 Fixed constants of the two models

Both models form expected inflation and expected real growth by AR(1) learning re-estimated every quarter (Appendix A). In M_1 the equilibrium unemployment rate in (9) is fixed at $u^* = 0.052$ and the real-wage target k is fixed. In the broader productivity variant, the innovation step is drawn from $U(0, 0.10)$ and wage anchors rise by the full productivity gain. None of these constants is estimated, so differences between estimated M_0 and M_1 come from the coefficients in (10) and (11) and from the rules themselves.

C.4 Estimation and evaluation windows

The structural parameters and initial conditions in Tables 2 and 17 come from FIGARO and the national accounts. The behavioural coefficients used in baseline forecasts are estimated on the early and pre-2020 windows of Section 4. Evaluation uses eligible QOOS and WOOS targets absent from their corresponding estimation window.

D Inference validation and local diagnostics

D.1 Support, splits and calibration

Training, validation and held-out observations are grouped by their generating parameter draw. Stochastic siblings never cross these folds. The adopted baseline density uses a uniform-box-supported spline flow; the five-member mixture shares the data split and varies initialization. Support compliance and finite draws are checked separately from simulation-based calibration.

The empirical calibration check uses 200 held-out prior cases and a prespecified finite-draw rank calculation. Kolmogorov–Smirnov uniformity and rank-dispersion tests are corrected across parameters, both tests and the two estimation windows. The eight-coefficient error-correction gate uses threshold 0.0015625. Thresholds and finite-draw reference rules are not relaxed after viewing failures.

The baseline five-member markup estimator passes its recorded empirical gate. The error-correction model and both broader variants fail on wage-block calibration. Some individual densities are too concentrated, whereas pooling disagreeing networks can make the mixture too dispersed. Ensemble disagreement is not evidence of a smooth economic nonidentification ridge.

The baseline empirical report records no failures. Its SBC-screen pass is not an independent verification of training convergence, and it does not validate a different synthetic observation design.

The simulation-study screen (panel B of Table 18) and the empirical wage-block screen (panel C) are distinct records, with their own frozen multiplicity corrections. The simulation-study report records no baseline rank rejection, but does not verify training convergence; its combined replication gate fails. A rank-screen non-rejection is therefore not relabeled as complete estimator validation. Raw p-values and recorded decisions are displayed together rather than applying conventional per-parameter significance stars.

D.2 Origin averaged predictive diagnostics

The origin-averaged predictive calculation first averages the five-variable, twelve-horizon paths over origins. Opposing origin-specific errors can therefore cancel. Table 19 preserves these values with that explicit interpretation; the main results instead display origin-level errors and a maximum discrepancy calculated on complete reference panels. Neither diagnostic supplies a certified model-family rejection.

Table 18: **Simulation-based calibration screens.** Rank-variance ratio, raw Kolmogorov–Smirnov (KS) and rank-dispersion p -values, and the decision recorded in each saved report, for the empirical design and the simulation study.

Coefficient	Window or arm	Variance ratio	KS p	Dispersion p	Verdict
<i>A. Empirical design, M_0 (early and pre-2020 windows)</i>					
ϕ_{DP}	Early	0.961	0.8190	0.5373	No rejection
ϕ_{CP}		1.044	0.0512	0.4917	No rejection
ϕ_{AE}		1.035	0.0577	0.5820	No rejection
ϕ_{DPQ}		0.950	0.0788	0.4362	No rejection
ϕ_{GQ}		1.089	0.4601	0.1638	No rejection
ϕ_{UNION}		1.006	0.5508	0.9200	No rejection
ϕ_{DP}	Pre-2020	1.055	0.5400	0.3838	No rejection
ϕ_{CP}		1.148	0.0073	0.0212	No rejection
ϕ_{AE}		1.056	0.0422	0.3816	No rejection
ϕ_{DPQ}		0.918	0.2345	0.1966	No rejection
ϕ_{GQ}		0.954	0.8713	0.4699	No rejection
ϕ_{UNION}		0.992	0.6281	0.8991	No rejection
<i>B. Simulation study, M_0 and M_1</i>					
ϕ_{DP}	M_0 , six coefficients	1.026	0.9060	0.6805	No rejection
ϕ_{CP}		1.166	0.0023	0.0096	No rejection
ϕ_{AE}		1.067	0.0744	0.2913	No rejection
ϕ_{DPQ}		0.900	0.2432	0.1147	No rejection
ϕ_{GQ}		0.949	0.5653	0.4201	No rejection
ϕ_{UNION}		1.010	0.6021	0.8763	No rejection
ϕ_{DP}	M_1 , eight coefficients	0.963	0.7391	0.5593	No rejection
ϕ_{CP}		1.020	0.2802	0.7551	No rejection
ϕ_{AE}		1.035	0.5653	0.5810	No rejection
ϕ_{DPQ}		1.088	0.4083	0.1671	No rejection
ϕ_{GQ}		0.869	0.1363	0.0379	No rejection
β_{IDX}		0.490	1.79×10^{-6}	1.00×10^{-5}	Reject
λ_{RW}		0.619	0.0097	1.00×10^{-5}	Reject
δ_U		0.379	6.92×10^{-10}	1.00×10^{-5}	Reject
<i>C. Empirical design, M_1 wage block (early and pre-2020 windows)</i>					
β_{IDX}	Early	0.340	2.11×10^{-9}	1.00×10^{-5}	Reject
λ_{RW}		0.866	0.4635	0.0336	No rejection
δ_U		0.440	3.33×10^{-8}	1.00×10^{-5}	Reject
β_{IDX}	Pre-2020	0.481	1.01×10^{-5}	1.00×10^{-5}	Reject
λ_{RW}		0.637	0.0035	1.00×10^{-5}	Reject
δ_U		0.378	2.04×10^{-11}	1.00×10^{-5}	Reject

Notes: p -values are raw. Each verdict reproduces the saved report’s decision under its own frozen multiplicity threshold, not a 5% per-parameter rule. A variance ratio below one signals overly concentrated ranks, consistent with overly dispersed posterior intervals. Panel A: threshold 0.0020833 across six coefficients, two windows and two tests; the report records pass=true and no failures, but its training fields are null, so the table does not certify training convergence or Dutch structural correctness. Panel B: threshold 0.0017857 across 28 tests; each arm has 200 held-out parameter cases and 256 posterior draws per case; the rank-dispersion reference uses 200,000 simulations of IID discrete-uniform ranks, labelled nominal for the Sobol parameter design. The report does not verify training convergence and its combined diagnostic gate fails, so a no-rejection row is not full estimator adoption. Panel C: threshold 0.0015625 across eight coefficients, two windows and two tests; all eight coefficients enter the gate although only the wage block is shown, and the recorded overall verdict is fail. This is inference criticism, not a test of the Dutch wage equation.

Table 19: **Origin-averaged joint predictive discrepancies and component errors.** Joint path statistic and tail rank, and observed minus predictive mean at quarter twelve, after averaging over origins, by window and fitted or held-out origins.

	Early window		Pre-2020 window	
	Fitted	Held out	Fitted	Held out
<i>A. Origin-averaged path check</i>				
Origin range	2010Q1–2012Q3	2016Q1–2016Q3	2010Q1–2016Q3	2020Q1–2021Q3
Statistic	312.68	137.27	305.72	823.77
Tail rank	0.0078	0.0078	0.0078	0.0078
<i>B. Observed minus predictive mean at quarter twelve</i>				
Real value added	−0.20	4.93	1.32	5.05
Consumer prices	−1.60	0.92	−0.67	12.81
Wage rate	0.98	3.08	1.06	10.29
Value-added deflator	−4.87	0.38	−1.93	11.92
Unemployment rate	3.49	−1.05	2.09	0.57

Notes: Paths and component errors are averaged over origins before the diagnostic is calculated. Simulation batches A and B have disjoint theta groups, 128 paths each and covariance shrinkage 0.10. The tail-rank floor is 1/129; fitted data reuse and the limited held-out design do not justify a whole-family rejection. Component errors are cumulative log-percent except the unemployment rate, in percentage points. Held-out panel checks are distinct from the full rolling QOOS/WOOS forecast exercise. No binomial significance is inferred from overlapping targets.

D.3 Local mean-information

For one complete core-five forecast path, let $\mu(\theta)$ and $\Sigma(\theta)$ be its simulator mean and covariance. The diagnostic evaluates

$$I_\mu(\theta) = J_\mu(\theta)^\top \Sigma(\theta)^{-1} J_\mu(\theta), \quad J_\mu(\theta) = \frac{\partial \mu(\theta)}{\partial \theta^\top}, \quad (98)$$

using finite differences and regularized covariance estimates. It omits covariance derivatives and is not the Fisher information of the full simulator likelihood. Origins are evaluated separately, not summed as independent observations.

The initial grid covers three variants, five origins and three evaluation points: the prior center and the early/pre failed-estimator means. The latter are diagnostic points only. All 45 cases completed, with 116,160 recorded paths and no simulation failures. Independent derivative halves estimate cross-product information; negative eigenvalue estimates are retained as numerical uncertainty, not interpreted as negative true information or calibrated rank rejections. Finite-step and covariance-shrinkage sensitivity remains material. Table 20 reports selected wage entries.

Table 20: **Selected local wage sensitivity.** Cross-split diagonal entries of the local mean-information at 2016Q1, prior center, by model variant.

Variant	β_{IDX}	λ_{RW}	δ_U
Error-correction, eight coefficients	45.67	21.47	376.76
Growth and search, ten coefficients	47.71	27.11	392.15
Growth, search and fixed labor-force growth	54.46	37.71	451.23

Notes: Finite-difference step 0.05 prior standard deviations and covariance shrinkage 0.10. Information units are relative to the diagnostic’s parameter scaling, not posterior precision.

A higher-precision fresh-seed check of six selected cases does not eliminate weak-direction uncertainty. Some earlier negative directions become positive on fresh seeds; this is not proof of full rank. Substantial marginal wage sensitivity and unresolved joint separation can coexist.

D.4 Information in the frozen representation

A held-out readout compares the nineteen pooled summaries, full paths with summaries, and the actual frozen representation of the eight-coefficient pre-window estimator. Table 21 reports the point-prediction results for one representative member.

Table 21: **Held-out wage-parameter prediction.** Held-out R^2 of point predictions of the three wage coefficients, by input set.

Inputs	β_{IDX}	λ_{RW}	δ_U
Pooled summaries	0.386	0.683	0.766
Full paths and summaries	0.930	0.710	0.997
Actual frozen representation	0.994	0.899	0.999

Notes: Three thousand distinct training groups and 200 matched held-out groups; representative member 42. Development-data diagnostic, not fresh posterior validation.

The representation receives observations, not true parameters; split identities and held-out truths were verified. These readouts show retained predictive signal. They do not test sufficiency, global identification, local posterior coverage at Dutch observations or a valid neural density. The deterministic full-fold continuations retain the initial and all improving checkpoints.

Checkpoint-only evaluation then selected the minimum stored complete-fold validation loss, restored the same architecture and reproduced the loss on all 4,800 validation rows. No training was repeated. Table 22 reports the resulting unchanged SBC screens. All four continuations still fail: the fits from the first training seed reject on aggregate expectations; those from the second reject on quantity response and real-wage adjustment, with unemployment sensitivity additionally rejecting at sixteen threads. The reporting failure and density miscalibration are therefore distinct. These reused development cases do not authorize estimator adoption or replace independent confirmation under the study observation design.

Table 22: **Complete-fold M_1 continuations.** Recovered saved-checkpoint diagnostics for all four saved fits, without retraining: best and terminal epoch, restored validation loss and the coefficients rejected by the frozen screen.

Seed	Threads	Best / terminal epoch	Restored loss	Frozen-screen rejections
42	4	204 / 304	-18.97506	ϕ_{AE}
42	16	145 / 245	-18.94909	ϕ_{AE}
43	4	77 / 177	-18.61452	ϕ_{DPQ}, λ_{RW}
43	16	1 / 101	-18.58612	$\phi_{DPQ}, \lambda_{RW}, \delta_U$

Notes: The complete validation fold has 4,800 rows. Checkpoints were selected by stored validation loss, never by SBC. The report verifies identical restored losses, finite supported draws and exact native SBC replay on the same 200 development cases. Every continuation still fails the unchanged 0.0015625 screen. The four fits reuse the same evaluation cases; they are not 800 independent validation histories. No optimizer update, new model fit, gate change or estimator adoption occurs in this checkpoint-only evaluation.

D.5 Mixture and freshly initialized density heads

One fixed equal-weight joint-density mixture combines the best complete-fold checkpoint from each training seed, selected by validation loss rather than SBC. It draws an independent member label for each posterior sample; ranks and parameter estimates are not averaged. The unchanged screen rejects β_{IDX} and δ_U . This attempt is retained separately from the earlier mixtures and is not used to select new weights or members.

The subsequent treatment rebuilds the density head while freezing the selected checkpoint's complete conditioning pipeline, including its input normalization. Preflight checks verify newly initialized parameter storage, exact conditioning replay, unchanged support and deterministic validation on all 4,800 assigned rows. The two training seeds retain their original group-disjoint splits, prior, architecture, learning rate and acceptance screen.

Table 23 reports both new heads on the same 200 development cases. Only the δ_U Kolmogorov–Smirnov screen rejects; its dispersion screen does not. Neither overall screen passes. Both runs hit the 500-epoch cap rather than stopping through patience, and their minimum validation losses occur at epochs 497 and 500. Best checkpoints are restored and native SBC replay is verified. Fresh initialization is not fresh validation data, and these runs establish neither independent calibration nor training convergence. We report this diagnostic improvement without adopting either posterior.

Table 23: **Fresh M_1 density heads.** Training and stopping diagnostics on the frozen selected conditioning pipeline, and the remaining δ_U calibration discrepancy, by training seed.

	Seed 42	Seed 43
<i>A. Training on the frozen selected conditioning pipeline</i>		
Best / terminal epoch	497 / 500	500 / 500
Restored fold loss	−19.35477	−19.09807
Non-rejecting screens	7 of 8	7 of 8
Overall verdict	Fail: δ_U KS	Fail: δ_U KS
<i>B. Remaining δ_U calibration discrepancy</i>		
KS p	5.06×10^{-6}	0.000639
Dispersion p	0.3892	0.3483
Rank-variance ratio	1.055	0.940
90% coverage	0.865	0.870

Notes: Each run uses 16 CPUs, a 4,800-row complete validation fold and the same 200 reused development cases. There are 1,024 posterior draws per case; native SBC uses the first 256. No cases or nonfinite draws are dropped. The unchanged screen threshold is 0.0015625. KS rejects for δ_U in both runs; dispersion does not. Coverage is a development fraction, not independent validation. The two seeds reuse the same evaluation cases, not 400 independent histories. Fold losses concern each seed's own validation split and are not a common-fold model ranking. Both runs hit the 500-epoch cap; checkpoints are selected by validation loss, never SBC. No gate is relaxed and no estimator is adopted. The separate fixed 50/50 mixture rejects β_{IDX} and δ_U and is not used for these rows.

D.6 Reproducibility boundary

Reports are checked against the raw case records. Diagnostic reports distinguish numerical uncertainty from economic inference; no failed estimator is adopted.

E Initial single-panel experiment

Before the replicated study, each cell of the simulation design was run on one realised panel per phase. With one panel per phase, the experiment provides examples rather than frequencies: it cannot measure detection rates, interval coverage or recovery bias across independently replicated economies. That limitation motivated the replicated design of Section 6, which supersedes these results; the single-panel tables are therefore not reproduced here.

Three observations from the single panel are retained. First, in the baseline-on-baseline break comparison the posterior change in ϕ_{CP} was approximately 0.39, with a 90% interval of $[0.20, 0.56]$, against a generating change of 0.25: the direction of the change was recovered, but its point estimate overshoot. Second, refitting need not make a realised path more typical under its predictive reference. For baseline truth without a shift, the frozen and refitted maximum-discrepancy ranks were approximately 0.674 and 0.116; under the genuine shift they were 0.318 and 0.140. Parameter movement and a single predictive check therefore cannot be equated with useful warning of response errors. Third, after the cost-coefficient shift the generating M_1 -minus- M_0 response difference was about 0.23 percentage points (Monte Carlo standard error 0.07), while the baseline-on-baseline fitted response overshoot its generating response by about 0.62 points (0.05) and the baseline fitted to error-correction data overshoot by approximately 1.23 points (0.20). The error-correction estimator failed calibration on its wage block, so its fitted rows were unvalidated diagnostics throughout.

F Coefficient recovery in the known-truth experiment

Table 24 extends Table 6 to every estimated coefficient. Each entry summarizes the posterior means of 100 independent panels conditioned on by the frozen estimator of the fitted model.

Table 24: **Recovery of every estimated coefficient in the known-truth experiment.** Bias of the posterior mean, its RMSE and coverage of the central 90% interval over 100 panels, by cell and phase.

Coefficient	Truth	Initial phase			Post, no break			Post, break		
		Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.
<i>A. M_0 fitted to M_0 data (correct rule)</i>										
ϕ_{DP}	0.500	+0.010 (0.006)	0.062 (0.004)	0.92 (0.03)	+0.003 (0.006)	0.064 (0.004)	0.93 (0.03)	+0.004 (0.005)	0.047 (0.003)	0.95 (0.02)
ϕ_{CP}	0.250/0.500	−0.010 (0.008)	0.076 (0.005)	0.94 (0.02)	−0.005 (0.008)	0.079 (0.004)	0.89 (0.03)	+0.006 (0.007)	0.070 (0.005)	0.93 (0.03)
ϕ_{AE}	0.500	+0.008 (0.006)	0.059 (0.006)	0.93 (0.03)	+0.003 (0.006)	0.057 (0.004)	0.93 (0.03)	+0.023 (0.008)	0.086 (0.006)	0.87 (0.03)
ϕ_{DP_Q}	0.500	+0.014 (0.008)	0.079 (0.006)	0.89 (0.03)	+0.003 (0.007)	0.068 (0.004)	0.91 (0.03)	+0.015 (0.007)	0.073 (0.005)	0.88 (0.03)

Continued on next page

Table 24, continued

Coefficient	Truth	Initial phase			Post, no break			Post, break		
		Bias	RMSE	Cov.	Bias	RMSE	Cov.	Bias	RMSE	Cov.
ϕ_{G_Q}	0.500	−0.001 (0.005)	0.049 (0.004)	0.90 (0.03)	−0.002 (0.004)	0.043 (0.003)	0.94 (0.02)	+0.001 (0.004)	0.044 (0.003)	0.92 (0.03)
ϕ_{UNION}	0.625	0.000 (0.001)	0.010 (0.001)	0.90 (0.03)	−0.001 (0.001)	0.009 (0.001)	0.95 (0.02)	0.000 (0.001)	0.011 (0.001)	0.87 (0.03)
<i>B. M_0 fitted to M_1 data (omitted error correction)</i>										
ϕ_{DP}	0.500	+0.276 (0.007)	0.286 (0.006)	0.41 (0.05)	+0.273 (0.008)	0.283 (0.007)	0.40 (0.05)	+0.270 (0.007)	0.278 (0.006)	0.10 (0.03)
ϕ_{CP}	0.250/0.500	+0.382 (0.009)	0.391 (0.008)	0.26 (0.04)	+0.387 (0.010)	0.398 (0.010)	0.33 (0.05)	+0.350 (0.006)	0.356 (0.006)	0.15 (0.04)
ϕ_{AE}	0.500	+0.131 (0.011)	0.170 (0.007)	0.97 (0.02)	+0.127 (0.011)	0.167 (0.009)	0.96 (0.02)	+0.126 (0.013)	0.180 (0.009)	0.99 (0.01)
ϕ_{DP_Q}	0.500	+0.184 (0.014)	0.233 (0.012)	0.80 (0.04)	+0.163 (0.016)	0.224 (0.012)	0.77 (0.04)	+0.179 (0.013)	0.221 (0.011)	0.74 (0.04)
ϕ_{G_Q}	0.500	+0.111 (0.007)	0.132 (0.006)	0.36 (0.05)	+0.106 (0.006)	0.120 (0.006)	0.49 (0.05)	+0.094 (0.006)	0.112 (0.005)	0.48 (0.05)
ϕ_{UNION}	−	[0.111] (0.015)	−	−	[0.093] (0.014)	−	−	[0.068] (0.011)	−	−
<i>C. M_1 fitted to M_1 data (correct rule, exploratory)</i>										
ϕ_{DP}	0.500	−0.004 (0.008)	0.084 (0.007)	0.92 (0.03)	+0.003 (0.008)	0.076 (0.005)	0.97 (0.02)	+0.001 (0.008)	0.084 (0.006)	0.89 (0.03)
ϕ_{CP}	0.250/0.500	+0.015 (0.008)	0.076 (0.005)	0.97 (0.02)	+0.014 (0.008)	0.077 (0.005)	0.94 (0.02)	−0.007 (0.007)	0.065 (0.005)	0.97 (0.02)
ϕ_{AE}	0.500	+0.025 (0.008)	0.085 (0.007)	0.93 (0.03)	+0.016 (0.007)	0.069 (0.004)	0.97 (0.02)	+0.031 (0.009)	0.094 (0.008)	0.93 (0.03)
ϕ_{DP_Q}	0.500	+0.025 (0.010)	0.102 (0.007)	0.92 (0.03)	+0.013 (0.009)	0.095 (0.006)	0.96 (0.02)	+0.023 (0.009)	0.092 (0.006)	0.91 (0.03)
ϕ_{G_Q}	0.500	+0.004 (0.007)	0.065 (0.004)	0.89 (0.03)	+0.002 (0.006)	0.056 (0.003)	0.96 (0.02)	+0.001 (0.005)	0.050 (0.004)	0.97 (0.02)
β_{IDX}	0.500	+0.001 (0.000)	0.003 (0.000)	1.00 (0.00)	0.000 (0.000)	0.002 (0.000)	1.00 (0.00)	0.000 (0.000)	0.002 (0.000)	0.99 (0.01)
λ_{RW}	0.100	+0.001 (0.000)	0.003 (0.000)	1.00 (0.00)	+0.002 (0.000)	0.003 (0.000)	1.00 (0.00)	+0.001 (0.000)	0.002 (0.000)	0.99 (0.01)
δ_U	0.200	0.000 (0.000)	0.001 (0.000)	0.99 (0.01)	0.000 (0.000)	0.001 (0.000)	1.00 (0.00)	0.000 (0.000)	0.001 (0.000)	1.00 (0.00)
<i>D. M_1 fitted to M_0 data (reverse omission, exploratory)</i>										
ϕ_{DP}	0.500	−0.006 (0.008)	0.084 (0.006)	0.97 (0.02)	−0.007 (0.009)	0.085 (0.006)	0.97 (0.02)	+0.035 (0.006)	0.072 (0.005)	0.97 (0.02)
ϕ_{CP}	0.250/0.500	+0.078 (0.007)	0.106 (0.006)	0.96 (0.02)	+0.066 (0.009)	0.109 (0.007)	0.98 (0.01)	−0.084 (0.009)	0.125 (0.007)	0.99 (0.01)
ϕ_{AE}	0.500	+0.061 (0.007)	0.092 (0.007)	0.97 (0.02)	+0.048 (0.008)	0.093 (0.006)	0.99 (0.01)	−0.016 (0.008)	0.085 (0.005)	1.00 (0.00)
ϕ_{DP_Q}	0.500	−0.014 (0.010)	0.100 (0.007)	0.89 (0.03)	−0.022 (0.009)	0.088 (0.006)	0.93 (0.03)	−0.052 (0.008)	0.097 (0.006)	0.89 (0.03)
ϕ_{G_Q}	0.500	+0.019 (0.007)	0.072 (0.005)	0.90 (0.03)	+0.021 (0.005)	0.058 (0.004)	0.93 (0.03)	+0.035 (0.006)	0.071 (0.004)	0.83 (0.04)
β_{IDX}	−	[0.636] (0.006)	−	−	[0.639] (0.005)	−	−	[0.622] (0.004)	−	−
λ_{RW}	−	[0.059] (0.006)	−	−	[0.071] (0.008)	−	−	[0.081] (0.008)	−	−
δ_U	−	[0.030] (0.001)	−	−	[0.031] (0.001)	−	−	[0.038] (0.001)	−	−

Notes: Monte Carlo standard errors in parentheses; nominal coverage is 0.90. The true ϕ_{CP} is 0.250, rising to 0.500 in

the break phase. In cross-rule cells the fitted rule’s own wage coefficients have no generating counterpart: for those rows the bracketed entry is the mean posterior mean, and bias, RMSE and coverage are not defined. M_1 -fitted cells use an estimator that fails calibration.

G Evidence status and specification-test implementation

G.1 Results included and results withheld

The baseline and the controlled wage alternative use the markup pricing rule of Section 3.3. Table 25 summarizes what each piece of evidence supports.

Table 25: **Interpretive status of the evidence in this draft.** Each evidence item and the interpretation it supports.

Evidence	Supported interpretation
Baseline empirical SBC	The stated inference gate passes under the simulated observation design; Dutch correctness is not implied.
Baseline Dutch predictive ranks	The observed panel is atypical under the fitted predictive reference; no family-wide rejection is certified.
Initial single-panel experiment	Illustrative panel-level recovery and discrepancies (Appendix E); superseded by the replicated study.
Replicated eight-cell experiment	100 independent panels per cell with frozen estimators; M_0 -fitted cells use an estimator that passes its rank screen, M_1 -fitted cells exploratory. All 1,200 panel-phase conditionings completed; no whole-panel failure.
Error-correction and broader posteriors	Calibration remains failed. New core-model heads leave one rejection and hit the training cap. M_1 results are reported as exploratory, not adopted.
Dutch prior predictive check	Complete audit of 1,200,000 prior paths; descriptive ranks, not a family rejection.
Local mean-information and readouts	Wage-parameter signal exists at tested points; joint/global identification and calibrated densities remain unresolved.
Conditional model-family test	Point-rank and dependence-safe combination implemented; global maximization and agent-based operating characteristics remain unverified.

G.2 The conditional model-family test

Following Dufour (2006), we separately implement a maximized Monte Carlo protocol for the null that one admissible parameter vector gives the correct conditional joint forecast law at every prespecified origin within a regime:

$$H_{0,m,r} : \exists \theta \in \Theta_m : G_o(\cdot \mid \mathcal{I}_o) = P_m(\cdot \mid \theta, \mathcal{I}_o) \quad \text{for all } o \in \mathcal{O}_r. \quad (99)$$

The null includes the initialization and information assumptions. Testing regimes separately differs from requiring one vector across a break.

At each candidate vector and origin, an independent simulation batch fixes means $m_{o,k,h}(\theta)$ and scales $s_{o,k,h}(\theta)$. One discrepancy uses all five variables and twelve horizons:

$$T_o(x; \theta) = \max_{k,h} \frac{|x_{o,k,h} - m_{o,k,h}(\theta)|}{s_{o,k,h}(\theta)}. \quad (100)$$

A separate batch of B complete paths supplies the rank

$$p_o(\theta) = \frac{1 + \sum_{b=1}^B \mathbf{1}\{T_o(x_b; \theta) \geq T_o(x_{\text{obs}}; \theta)\}}{B + 1}. \quad (101)$$

At the true vector, finite-sample protection requires exchangeability of observed and simulated statistics conditional on the stated information and normalization batch. Complete paths preserve within-origin dependence. For K prespecified origins, Bonferroni avoids an independence assumption:

$$p_{\text{combined}}(\theta) = \min\{1, K \min_o p_o(\theta)\}, \quad p_{\text{model}} = \sup_{\theta \in \Theta_m} p_{\text{combined}}(\theta). \quad (102)$$

The search domain is the full declared parameter box, not the SBI posterior. This test works through the simulator and does not require a successful neural posterior.

A searched maximum is a lower bound on p_{model} , not an upper bound. A passing admissible vector prevents rejection by this maximized test. Failure to find one does not certify rejection. The current global upper bound remains one; no optimizer failure or finite grid is reported as an exact model-family rejection. The implementation has passed numerical and Gaussian-control checks, but these do not establish size and power for the full agent-based experiment. No certified Dutch model-family rejection is reported in this draft.

G.3 Numerical checks of the conditional test

The independent normalization batch has 64 paths and the reference batch 999. The statistic uses five variables and twelve quarters, with fixed streams during the parameter search. Synthetic phase tests combine fourteen prespecified origins; a separate cross-phase constancy null combines twenty-eight. The six-dimensional baseline and eight-dimensional wage alternative use their full parameter boxes.

A compute-node implementation check passed seven algebra and integration contracts. Across 2,000 independent Gaussian control histories, the one-origin rank test rejected 5.05% at nominal 5%, with a reported 95% binomial interval of [4.13, 6.10]%. A dependence-safe combination of overlapping Gaussian windows rejected 4.20%, with interval [3.36, 5.17]%. A selected wrong-variance point-null alternative was rejected in all control histories.

These results validate the stated Gaussian-control implementation, not size or power for CANVAS model-family misspecification. In particular, point-null detection against a scale alternative does not establish power after maximizing over the economic parameter family. Bounded simulator searches produce lower bounds and admissible witnesses only. The global upper bound remains one, so small searched ranks are not reported as exact family rejections.

G.4 Reproducibility and outstanding work

All simulations, estimations and scoring runs were executed on the Dutch national supercomputer (Snellius, SURF). Each reported number is generated from an aggregate report whose cryptographic hash, producing job and source and input identities are recorded; raw simulated paths, posterior draws and network checkpoints remain archived with the computations. Failed and superseded runs, including all results under the retired pricing rule, are retained but not

Table 26: **Numerical controls for the conditional Monte Carlo test.** Rejections in independent Gaussian control histories, rejection rate with its 95% binomial interval, and the scope of the model-family conclusions.

Control	Rejections	Histories	Rate	95% interval
<i>A. Independent Gaussian control histories</i>				
Correct point null	101	2000	0.051	[0.041, 0.061]
Correct overlapping-window null	84	2000	0.042	[0.034, 0.052]
Wrong-variance point alternative	2000	2000	1.000	[0.998, 1.000]
<i>B. Scope of model-family conclusions</i>				
Pointwise Monte Carlo rank	Tests the specified conditional distribution at a fixed parameter.			
Largest searched rank	Lower bound on the maximized p -value, not a certified maximum.			
Available global upper bound	One; no model-family rejection is certified.			
Dutch model-family test	No certified empirical result is reported.			

Notes: Nominal alpha is 0.05; normalization and reference batches have 64 and 999 paths. Rates are fractions, not percentages. The outer Gaussian histories are independent; the target windows within a history overlap. Binomial intervals describe outer-experiment Monte Carlo uncertainty. The wrong-variance result is point-null detection, not power after economic-parameter maximization. These controls do not establish ABM family-test size or power.

pooled with current results. Code, run records and aggregate reports are available from the author on request.

The results are preliminary. No validation gate is relaxed and the failed error-correction posterior is not adopted. Open tasks are a calibrated estimator for the wage alternative, operating characteristics beyond the one omitted wage rule studied in Section 6, and a certified conditional specification test if a family rejection is to be claimed.