# Online Appendix

Monetary Policy, Endogenous Bill Supply, and the Liquidity Premium

Cristián Cuevas

## A.1 Derivation of the pricing equation (4) and the linearization (6)

The investor chooses consumption $C_t$, deposits $D_t$, bills $B_t$, and illiquid lending $L_t$ to maximize $u(C_t, Q_t)$ subject to the flow budget constraint, where

$$Q_t = \left[(1-\lambda_t)(D_t/P_t)^{\rho} + \lambda_t (B_t/P_t)^{\rho}\right]^{1/\rho},$$

where $P_t$ is the nominal price of the consumption good (the price level). The first-order conditions for bills and deposits are

$$\frac{u_Q}{u_C}\,\chi_t^{\rho} Q_t^{1-\rho} \lambda_t (B_t/P_t)^{\rho-1} = \frac{i_t - i_t^b}{1+i_t}, \qquad \frac{u_Q}{u_C}\,\chi_t^{\rho} Q_t^{1-\rho} (1-\lambda_t)(D_t/P_t)^{\rho-1} = \frac{i_t - i_t^d}{1+i_t}.$$

Dividing the first by the second eliminates $u_Q/u_C$, $Q_t^{1-\rho}$, the price level, and the $(1+i_t)$ terms:

$$\frac{\lambda_t}{1-\lambda_t}\Big(\frac{B_t}{D_t}\Big)^{\rho-1} = \frac{i_t - i_t^b}{i_t - i_t^d}.$$

With $i_t^d = \eta i_t$, $i_t - i_t^d = (1-\eta) i_t$, and therefore

$$LP_t \equiv i_t - i_t^b = i_t\,(1-\eta)\frac{\lambda_t}{1-\lambda_t}\Big(\frac{B_t}{D_t}\Big)^{\rho-1} \equiv i_t\,\kappa(B_t/D_t;\lambda_t),$$

which is equation (4) of the paper. The *relative* liquidity shock $\lambda_t$ enters in $u_t$ of equation (3).

**Linearization: from (4) to equation (6).** Write $LP_t = F(i_t, b_t, d_t, \lambda_t) \equiv i_t\,\kappa\big(e^{\,b_t - d_t};\lambda_t\big)$ with $b_t \equiv \log B_t$ and $d_t \equiv \log D_t$. A first-order expansion around sample means $(\bar\imath, \bar b, \bar d, \bar\lambda)$, with $\bar x \equiv e^{\,\bar b - \bar d}$, $x_t \equiv e^{\,b_t - d_t}$, and $\bar\kappa \equiv \kappa(\bar x;\bar\lambda)$, gives

$$LP_t \;\approx\; \bar\imath\bar\kappa \;+\; \bar\kappa\,(i_t - \bar\imath) \;+\; \bar\imath\,\kappa_x(\bar x;\bar\lambda)\,(x_t - \bar x) \;+\; \bar\imath\,\kappa_\lambda(\bar x;\bar\lambda)\,(\lambda_t - \bar\lambda).$$

Two substitutions complete the derivation. First, $x_t - \bar x \approx \bar x\big[(b_t - \bar b) - (d_t - \bar d)\big]$ converts the quantity deviation to logs. Second, the CES power form has *constant* elasticity, $x\,\kappa_x(x) = (\rho-1)\,\kappa(x)$, so

$$\bar\imath\,\kappa_x(\bar x;\bar\lambda)\,\bar x \;=\; \bar\imath\,\bar\kappa\,(\rho-1) \;\equiv\; \gamma_0 \;=\; (\rho-1)\,\overline{LP} \;<\; 0,$$

where the semi-elasticity of the premium with respect to log bill supply, holding $i$ and $d$ fixed, $\gamma_0$, scales with the *level* of the premium. Collecting terms (the $\bar\imath\bar\kappa$ terms cancel),

$$LP_t = \bar\kappa\,i_t + \gamma_0\,b_t - \gamma_0\,d_t + c_0 + u_t, \qquad c_0 = -\gamma_0(\bar b - \bar d), \quad u_t = \bar\imath\,\kappa_\lambda(\bar x;\bar\lambda)\,(\lambda_t - \bar\lambda) + O(\|\cdot\|^2),$$

which is equation (3); the constant $c_0$ is absorbed by the intercept and drops out entirely in the first-differenced specifications. Since $\kappa_\lambda = \kappa/[\lambda(1-\lambda)] > 0$, the error is, to first order, the relative-liquidity shock scaled by $\overline{LP}/[\bar{\lambda}(1-\bar{\lambda})]$: a positive $u_t$ raises the premium and, through the Treasury condition (2) and the deposit first-order condition, moves $b_t$ and $d_t$—the simultaneity that motivates the instruments.

Three remarks. (i) The exact log counterpart of (4) is $\log LP_t = \log i_t + \log(1-\eta) + \log\frac{\lambda_t}{1-\lambda_t} + (\rho-1)(b_t - d_t)$, in which the elasticity $\rho-1$ is exact. The level form is used because it keeps the estimands in the basis-points-per-percentage-point units of the literature (Nagel 2016; Krishnamurthy and Vissing-Jorgensen 2012) and accommodates months where the measured spread is near zero; the decomposition identity of A.3 applies to any linear-in-parameters dependent variable, so nothing in Section 4 hinges on this choice. (ii) The expansion drops the bilinear term $(\rho-1)\bar{\kappa}\,(i_t - \bar{\imath})\big[(b_t - \bar{b}) - (d_t - \bar{d})\big]$: in the exact model the supply semi-elasticity is $(\rho-1)LP_t$, which the linearization freezes at $(\rho-1)\overline{LP}$. This is second order in deviations and affects the structural reading of $\gamma$, not the in-sample identity. (iii) $b_t$ and $d_t$ enter only through $b_t - d_t = \log(B_t/D_t)$, hence with equal and opposite coefficients—a symmetry inherited from the ratio in (4). A common deflator therefore cancels: measuring the included regressor as $\log(B_t/\mathrm{GDP}_t)$ and absorbing $\log(D_t/\mathrm{GDP}_t)$ into the omitted money response preserves $\gamma_0$ while keeping both series stationary.

## A.2 The Treasury problem, the role of $\Phi$, and Proposition 1

**First-order condition.** The Treasury minimizes $C(B) = i^b B + i^n(F - B) + \Phi(B)$ with $i^b = i\,[1 - \kappa(B/D)]$. Since $x\kappa'(x) = (\rho-1)\kappa(x)$,

$$\frac{d}{dB}\big[i(1-\kappa)B\big] = i - i\kappa - i(\rho-1)\kappa - i\kappa\cdot 0 = i - \rho\, i\,\kappa,$$

so $C'(B) = i - \rho i\kappa - i^n + \Phi'(B)$ and the interior optimum satisfies equation (5):

$$\Phi'(B) = (i^n - i) + \rho\, i\,\kappa(B/D).$$

**Why a convex $\Phi$ is required (for any $\rho < 1$).** Suppose $\Phi \equiv 0$. Premium revenue is $R(B) \equiv iB\kappa(B/D) \propto B^\rho$.

- If $\rho \in (0,1)$ and $i^n \geq i$: $C'(B) = (i - i^n) - \rho i\kappa < 0$ for all $B$, so cost is strictly decreasing and the Treasury goes to the all-bill corner ($R(B)$ grows without bound in $B$).
- If $\rho < 0$: $R(B) \to \infty$ as $B \to 0$ (the CES Inada property—the premium per bill explodes faster than the quantity shrinks), so $C(B) \to -\infty$ at $B = 0$; moreover any stationary point has $C''(B) = \rho(1-\rho)i\kappa/B < 0$ and is a local cost *maximum*. Again no interior minimum exists.

Hence for every $\rho < 1$ a convex issuance cost is the minimal closure delivering a well-defined interior optimum; we adopt the rollover-risk interpretation of Greenwood, Hanson, and Stein (2015a). The second-order condition with $\Phi$ is

$$C''(B) = \Phi'' + \rho(1-\rho)\, i\kappa/B > 0,$$

automatic for $\rho \in [0,1)$ and, for $\rho < 0$, an assumption that $\Phi'' > -\rho(1-\rho)\bar{\imath}\bar{\kappa}/\bar{B}$—mild in practice because $\bar{\kappa} = \overline{LP}/\bar{\imath}$ is small.

**Proof of Proposition 1.** Differentiate the first-order condition in $i$ holding $D$, using $\partial i^n/\partial i = \theta$ and $\partial(\rho i\kappa)/\partial B = \rho i\kappa'(x)/D = -\rho(1-\rho)i\kappa/B$:

$$\left[\Phi'' + \rho(1-\rho)i\kappa/B\right]\frac{dB}{di} = (\theta - 1) + \rho\kappa \implies \frac{dB}{di} = \frac{\rho\bar{\kappa} - (1-\theta)}{\Phi'' + \rho(1-\rho)\bar{i}\bar{\kappa}/\bar{B}},$$

evaluated at sample means. The denominator is positive by the second-order condition, so $dB/di < 0$ iff $1 - \theta > \rho\bar{\kappa}$. *Reinforcement through money:* if $D$ also responds, the numerator gains the term $\rho i\,\kappa'(x)\,(-B/D^2)\,dD/di$; with $\kappa' < 0$ and $dD/di < 0$ under the corridor, this term is negative, making $dB/di$ more negative. □

**Proof of Corollary 1.** *Hard constraint.* If $B_t = \bar{B}$, then $\delta \equiv db/di = 0$ by construction, and substituting into the decomposition $\Pi = \beta^* + \gamma\delta$ (Section 3; A.3) gives $\Pi = \beta^*$. Because $\beta^* = \bar{\kappa} - \gamma_0(\partial d/\partial i)|_b$ involves only the pricing block—investors, banks, and the central bank—it is unaffected by the Treasury's rule, so the constrained pass-through equals the unconstrained non-issuance partial. Shifts in the constraint move the premium through equation (3): $\partial LP/\partial \log \bar{B} = \gamma_0 < 0$, so forced bill paydowns raise the premium. *Smooth version.* From Proposition 1, $\delta \to 0$ as $\Phi'' \to \infty$, so $\Pi \to \beta^*$ continuously; the hard constraint is the limit point. □

**Remark: maturity compression and the sign of $\delta$.** Under the expectations hypothesis with a mean-reverting short rate with reversion speed $\varphi$, the $n$-maturity yield loads on the current short rate with $\theta(n) = (1 - e^{-\varphi n})/(\varphi n)$, which is decreasing in $n$ with $\theta(n) \to 1$ as $n \to 0$. A Treasury whose feasible alternative maturity is compressed toward the short end therefore operates with $\theta$ near one, and Proposition 1's formula gives

$$\lim_{\theta\to 1}\delta = \frac{\rho\bar{\kappa}}{\Phi'' + \rho(1-\rho)\bar{i}\bar{\kappa}/\bar{B}}, \qquad \operatorname{sign}(\delta) = \operatorname{sign}(\rho).$$

Hence the reversal region $1 - \theta \le \rho\bar{\kappa}$ (where $\delta \ge 0$ and $\Pi \le \beta^*$) is attainable with $\theta \le 1$ only if $\rho > 0$; for $\rho \le 0$ it requires $\theta > 1 + |\rho|\bar{\kappa}$—long yields overreacting to policy, e.g. through rate-sensitive term premia, a channel outside the model. Given the negative implied $\widehat{\rho}$ of Section 4, the empirically relevant reading of maturity compression is attenuation of $\delta$ rather than reversal, absent a term-premium response; the main-text sentence is accordingly phrased as the negation of Proposition 1's condition, without selecting among these routes.

## A.3 Estimands and the exact 2SLS decomposition

Partial all variables on the included exogenous controls $X_t$ (Frisch–Waugh), and let $P_Z$ denote projection on the (partialled) instruments $Z_t$. Write $\tilde{\imath} = P_Z\,\Delta i$ and $\tilde{b} = P_Z\,\Delta b$.

(i) Column (2): $\widehat{\Pi} = (\tilde{\imath}'\tilde{\imath})^{-1}\tilde{\imath}'\Delta LP$.

(ii) Column (3): $(\widehat{\beta}^*, \widehat{\gamma})$ are the coefficients of $\Delta LP$ on $(\tilde{\imath}, \tilde{b})$.

(iii) Column (4): $\widehat{\delta} = (\tilde{\imath}'\tilde{\imath})^{-1}\tilde{\imath}'\Delta b = (\tilde{\imath}'\tilde{\imath})^{-1}\tilde{\imath}'\tilde{b}$.

By the omitted-variable-bias identity applied to (i)–(iii),

$$\widehat{\Pi} = \widehat{\beta}^* + \widehat{\gamma}\,\widehat{\delta} \qquad \text{exactly, in sample,}$$

*provided the instrument set $Z$, the controls $X$, and the estimation sample are identical across columns.* A practical implication: with a single endogenous regressor the first-stage regression is the same in columns (2) and (4), so its $F$ statistic must be numerically identical across those columns; any discrepancy indicates non-harmonized samples.

**Concentrating out deposits.** Equation (6) contains $d_t$; the regressions do not. Let $(\pi_i, \pi_b)$ denote the coefficients of the linear projection of the instrument-induced deposit variation $P_Z\,\Delta d$ on $(P_Z\,\Delta i,\ P_Z\,\Delta b)$. Substituting into (6), 2SLS on $(i, b)$ identifies

$$\beta^* = \bar{\kappa} - \gamma_0\,\pi_i, \qquad \gamma = \gamma_0\,(1 - \pi_b),$$

and the projection residual is orthogonal to $Z$ *by construction*: the estimates, the exact adding-up identity, and the share $s$ are mechanically unaffected by concentrating $d_t$ out. What requires an assumption is the structural reading $\pi_i = (\partial d/\partial i)$, $\pi_b = (\partial d/\partial b)$—i.e., $\beta^*$ as opportunity-cost scaling plus the money response, and $\gamma$ as the supply partial net of central-bank accommodation at a given rate target. That reading requires the instruments to move deposits only through the rate and bill supply. For the futures surprises this holds by design: the deposit response to surprise-driven rate changes *is* the corridor money channel, absorbed into $\beta^*$. For the month dummies it requires calendar seasonality in deposits to be negligible relative to bill seasonality—a magnitude claim, checkable by regressing $\Delta\log(\text{deposits})$ on the month dummies and comparing amplitude and joint $F$ with the same regression for $\Delta b$; an offending month can be moved from $Z$ to the controls. Including $\Delta d$ as a regressor would not improve matters: deposits respond to the same $\lambda_t$ shocks as the premium, so an uninstrumented $\Delta d$ is a bad control that contaminates all coefficients, and conditioning on $d$ would change every estimand's meaning (turning $\beta^*$ into a holding-deposits-fixed partial and severing its money-channel interpretation). Finally, the regressor is $b$ rather than $\log(B/D)$ because the estimand is the Treasury's channel: $\delta$ must be the debt manager's response, with banking-sector quantities in the absorbed background. A variant using instrumented $\Delta\log(B/D)$ would recover $\gamma_0$ free of the accommodation factor $(1 - \pi_b)$, sharpening the implied $\widehat{\rho}$ below.

**Implied substitutability.** As derived in A.1, $\gamma_0 = \bar{\imath}\bar{\kappa}(\rho - 1) = (\rho - 1)\,\overline{LP}$ per log point of bill supply, where $\overline{LP}$ is measured in the same units as the dependent variable (basis points in Table 1). Absent central-bank accommodation ($\partial d/\partial b = 0$), the point estimate therefore implies

$$\widehat{\rho} = 1 + \frac{\widehat{\gamma}}{\overline{LP}}, \qquad \widehat{\sigma} = \frac{1}{1 - \widehat{\rho}},$$

with $\widehat{\sigma}$ the elasticity of substitution between bills and deposits. For $\widehat{\gamma} = -105.9$ and $\overline{LP} \in [20, 30]$ bp, $\widehat{\rho} \in [-4.3, -2.5]$ and $\widehat{\sigma} \in [0.19, 0.28]$: bill demand at the monthly bill–deposit margin is far steeper than long-sample aggregate substitutability estimates. By the delta method $\text{se}(\widehat{\rho}) = \text{se}(\widehat{\gamma})/\overline{LP}$ ($\approx 1.0$ at $\overline{LP} = 25$ bp), so $\widehat{\rho}$ is significantly below one—a test identical to $\widehat{\gamma} \neq 0$. The conclusion is also robust to resampling uncertainty: even the upper endpoint of $\widehat{\gamma}$'s block-bootstrap interval (Appendix A.5) implies $\widehat{\rho} \leq -1$ for $\overline{LP} \leq 30$ bp. Accommodation rescales $\widehat{\gamma}$ by $1 - \partial d/\partial b$, but no plausible magnitude bridges an order of magnitude.

**Levels versus differences.** First-differencing equation (6) alters the error structure and the identifying variation, guarding against stochastic trends. Column (3) is therefore an estimate of the very semi-elasticity a levels regression targets—and the consistent one under the maintained

exclusion restrictions. The OLS levels coefficient of column (1) is not an alternative estimate of $\gamma_0$ for this purpose: the identification argument of Section 4 is precisely that simultaneity renders it uninformative, and it is also not statistically significant, unable to reject $\rho = 1$ (perfect substitutes). Its face value nonetheless reproduces the high substitutability of long-sample studies—evidence that low-frequency, non-instrumented moments understate supply effects, not that substitutability is high. Finally, if substitution is horizon-dependent (steeper at short horizons than in the long run), no single-$\rho$ CES fits both the monthly moment and long-horizon moments; none of the paper's results relies on a particular value of $\rho$ beyond $\rho < 1$.

## A.4 Data notes

All interest rates in the empirical analysis are monthly averages of daily annualized effective yields.

*Yields*: The return on the 3-month Treasury-bill is from FRED, series TB3MS, available from January 1934 onwards.

The GC repo rate comes from Bloomberg, and the federal funds rate corresponds to the effective federal fund rate (in FRED).

The 5- and 10-year Treasury constant-maturity yields used in Appendix A.6 are from FRED, series GS5 and GS10.

*Supply variables*: Data on outstanding T-bills is from the Center for Research in Security Prices (CRSP) monthly Treasury files. The series reports the outstanding face values of Treasury Bills since January 1947. I use this source instead of the Financial Accounts (Table L.210 reports government securities) since the latter only reports quarterly figures. Data on nominal GDP is from the quarterly FRED database, linearly interpolated to monthly values.

The supply of reserves is the total reserves of depository institutions (FRED, series TOTRESNS), available from January 1959 onwards.

The data for long-term bonds and notes in Section 2, and for total marketable debt in Appendix A.6, are from the Monthly Statement of the Public Debt, available at TreasuryDirect.

*Other variables*: VIX index is from FRED, available since 1990. For the period pre-1990, I use Nagel (2016)'s projection, which is based on realized volatility of the S&P 500 index (see that paper's appendix for details).

## A.5 Inference on the issuance response and the share

**Setup.** All results use the common harmonized sample ($T = 211$ months). The moving-block bootstrap resamples the data tuples in blocks of $L$ consecutive months with overlapping starting points, $L \in \{12, 24, 36\}$—multiples of twelve so that every calendar month, and hence every seasonal instrument, appears in every draw—with 1,999 replications and the three 2SLS columns re-estimated per draw; the adding-up identity held to machine precision in every replication. The Anderson–Rubin set inverts the test of $\Delta b_t - \delta_0 \Delta i_t$ on the futures surprise (month dummies and VIX as controls; Newey–West, 12 lags) over a grid of $\delta_0$; it is just-identified, involves no resampling, and is therefore invariant to $L$.

**Results.** The Wald 95% interval for $\widehat{\delta} = -0.0316$ (HAC s.e. 0.015, $p = 0.030$) is $[-0.062, -0.003]$; the Anderson–Rubin 95% set is $[-0.056, -0.001]$. Bootstrap intervals:

| | $L = 12$ | $L = 24$ | $L = 36$ |
|---|---|---|---|
| $\widehat{\delta}$, percentile | $[-0.065,\ 0.016]$ | $[-0.057,\ 0.013]$ | $[-0.047,\ 0.013]$ |
| $\widehat{\delta}$, bias-corrected | $[-0.081,\ 0.000]$ | $[-0.091,\ -0.013]$ | $[-0.076,\ -0.020]$ |
| $s$, percentile | $[-0.30,\ 0.63]$ | $[-0.31,\ 0.64]$ | $[-0.38,\ 0.58]$ |
| $s$, Fieller | $[-0.06,\ 0.59]$ | $[0.00,\ 0.58]$ | $[0.04,\ 0.54]$ |
| bootstrap bias of $\widehat{\delta}$ | $+0.010$ | $+0.017$ | $+0.021$ |
| draws with $\widehat{\Pi} \leq 0$ (%) | 0.35 | 1.60 | 0.80 |

The following intervals exclude zero at every $L$: $\widehat{\gamma}$ within $[-153, -63]$, $\widehat{\Pi}$ within $[1.4, 17.2]$, $\widehat{\beta}^*$ within $[0.4, 14.0]$. Trimming the $\leq 1.6\%$ of draws with $|\widehat{\Pi}| < 1$ bp barely moves the percentile interval for $s$ ($[-0.26, 0.57]$ vs. $[-0.31, 0.64]$ at $L = 24$), so its width is genuine uncertainty rather than denominator degeneracy.

The bootstrap distribution of $\widehat{\delta}$ is shifted toward zero, with the shift and the share of degenerate-$\widehat{\Pi}$ draws both growing in $L$. This suggests that longer blocks may be omitting several high-surprise episodes, such that resamples have weak first stages, and weak-instrument 2SLS collapses toward OLS ($\approx 0$ for $\delta$). Consistently, $\widehat{\gamma}$—identified off calendar seasonality that every block contains by construction—is unaffected, indicating the fragility reflects the concentration of the surprise variation, not the supply channel. The same shift means the percentile intervals are conservative in a known direction; the bias-corrected intervals adjust for it but, with median-bias statistics $|z_0|$ up to 1.3, rest on extreme order statistics and are reported for completeness rather than emphasis (bias-corrected intervals for $s$—$[-0.01, 2.58]$, $[0.08, 78.4]$, and $[0.12, 12.4]$ at $L = 12, 24, 36$—are dominated by degenerate-$\widehat{\Pi}$ draws in the upper tail, the expected failure of the bias correction for a ratio, and carry no weight).

To summarize, the issuance response is significant at the 5% level under conventional HAC and under identification-robust Anderson–Rubin inference, and borderline under block-bootstrap resampling; the share is accordingly reported as a point estimate, $s = 0.27$, with the percentile and Fieller intervals displayed in Table 1.

## A.6 The Treasury's issuance response

**(a) Institutional background.** The Treasury's stated debt-management objective is to meet the government's borrowing need at the least cost to taxpayers over time. Since the late 1970s it has pursued that objective within a "regular and predictable" framework, where auctions follow a regular weekly schedule (Garbade 2007).

Within this framework, discretion is exercised through offering *amounts*, not by changing the calendar. Coupon auction sizes are adjusted infrequently and in small increments, with changes communicated around the quarterly refundings (February, May, August, November). Bill offering amounts are reset at each weekly announcement (13- and 26-week bills are auctioned every week, 52-week bills at four-week intervals for most of the sample) and cash-management bills absorb residual cash-flow shocks (Garbade 2015; Garbade, Partlan, and Santoro 2004). At the monthly frequency of Table 1, the operative margin of the maturity composition is therefore the bill margin.

Garbade's account is explicit that regular and predictable issuance was embraced *because it lowered borrowing costs*, by reducing marketing uncertainty. In the model, Equation (5) is a reduced form of this problem: $\Phi$ prices the rollover risk and smoothing constraints under which the Treasury operates.

However, the particular content of the rule can suggest different particular mechanisms. First, a rule that conditions on relative financing costs (directly, or through variables the policy rate moves)

is consistent with equation (5): the rule is the form cost minimization takes, and $\widehat{\delta}$ is a causal response under the exclusion restriction of the futures-surprise instrument in either case, since any systematic mapping from the rate to issuance generates a well-defined response to the instrument. In fact, the analytical framework of the Treasury Borrowing Advisory Committee (Belton et al. 2018) is a cost–risk optimization over issuance shares, and its central practical recommendation is a reaction function that adjusts issuance to a small set of economic and financial variables, including the yield curve. What the model supplies is the *interpretation* (composition driven by relative cost) together with the comparative statics of Proposition 1. Second, a fully rigid rule as in Corollary 1 ($\Phi'' \to \infty$ limit, in which $\delta \to 0$ and $\Pi \to \beta^*$), but this is rejected by the data, since the Anderson–Rubin set for $\widehat{\delta}$ (A.5) excludes zero. The third possible rule, a rule under which the maturity composition does not respond to relative costs at monthly frequency—the coupon path is smooth and bills move only with the financing need, is both observationally distinct from the model and contradicts its interpretation. I test this alternative in Part (b).

**(b) Composition versus residual financing.** The residual-financing alternative is operational: coupon issuance follows a predetermined smooth path $\bar{C}_t$ and bills absorb the financing need residually, $B_t = F_t - \bar{C}_t$, so that the maturity composition does not respond to prices at monthly frequency. Because deficits are countercyclical and easing is concentrated in downturns, such a rule would generate a negative $\widehat{\delta}$ with no role for relative costs at the monthly frequency. In this case, the null is operational: a residual rule could itself be cost-rational at the level of framework design (bills would be the absorbing margin precisely because they are the cheapest to adjust) so the hypothesis under test is not that the Treasury is indifferent to borrowing costs, but that the composition margin does not respond to them in the frequency assumed in the model. I test this by conditioning on the scale of borrowing: under the residual rule, bills respond to policy *only* through $F_t$, so controlling for the change in total borrowing should eliminate the rate coefficient; under cost management, the composition responds conditional on scale.

I augment the column (4) specification with $\Delta f_t \equiv \Delta \log(\text{total marketable debt})$ (from the Monthly Statement of the Public Debt),

$$\Delta b_t = \delta_F \, \Delta i_t + \varphi \, \Delta f_t + \Gamma' X_t + e_t,$$

estimated by 2SLS with the identical instrument set, controls, and sample as Table 1. Under the residual-financing null, $\delta_F = 0$ and $\varphi \approx \bar{F}/\bar{B}$; under the model, $\delta_F < 0$. The estimates are $\widehat{\delta}_F = -0.0312$ (s.e. 0.0149) and $\widehat{\varphi} = 0.13$ (0.23): the rate coefficient survives conditioning on the scale of borrowing, rejecting pure residual financing.

Two clarifications. First, $\Delta f_t$ is a mediator of the total response (cyclical deficits are themselves one channel through which policy can move bill supply) so $\delta_F$ is a diagnostic for the residual-financing null, but it does not replace $\widehat{\delta}$ in the decomposition of Section 3, which requires the *total* equilibrium response $db/di$ regardless of margin. Second, the composition tilt can also be read directly from the share: 2SLS of $\Delta[b_t - f_t]$ on the instrumented rate gives $-0.0245$ (s.e. 0.0136)—the bill share of marketable debt falls when policy tightens. As noted, the share regression is consistent with both readings on its own and is reported for completeness. The bill share also comoves with the term spread in 12-month changes (correlation 0.72), the reduced-form counterpart of Proposition 1's relative-cost mechanism. In addition, a placebo shows no response of long-term debt to the rate ($-0.007$, s.e. 0.005), which suggests that the Treasury tilts maturity rather than merely scaling issuance with output.

**(c) The mechanism.** *Step 1: policy rate → relative cost.* The relative cost of short- versus long-term financing the debt manager faces (the empirical counterpart of the relative-cost margin

priced by equation (5)) is measured as: $s_t^n \equiv i_t^n - i_t^b$, for $i^n$ the 5- and 10-year constant-maturity yields and $i^b$ the 3-month bill yield. 2SLS of $\Delta s_t^n$ on $\Delta i_t$, with the identical instrument, controls, and sample as Table 1, gives

$$\widehat{c}_5 = -0.552 \text{ (s.e. } 0.086), \qquad \widehat{c}_{10} = -0.679 \text{ (s.e. } 0.086) :$$

an instrumented 100 bp cut widens the financing spread by 55–68 bp, cheapening bill finance on impact.

The sign is a test of the no-term-structure-channel benchmark. This can be seen by two in-sample identities. First, each spread coefficient equals the difference of the underlying loadings, $\widehat{c}_n = \widehat{\theta}_n - \widehat{\beta}_b$, where $\widehat{\theta}_n$ and $\widehat{\beta}_b$ are the 2SLS loadings of $\Delta i_t^n$ and $\Delta i_t^b$ on the instrumented rate—an identity that holds by linearity under common instruments, controls, and sample. Second, because $LP_t \equiv \text{repo}_t - i_t^b$ by construction, $\widehat{\beta}_b = \widehat{\beta}_{\text{repo}} - \widehat{\Pi}$, so $\widehat{c}_n = \widehat{\theta}_n - \widehat{\beta}_{\text{repo}} + \widehat{\Pi}$, tying the new estimates to Table 1's pass-through ($\widehat{\Pi} = 0.126$ per percentage point). If long yields co-moved one-for-one with the short rate ($\theta_n = \beta_{\text{repo}}$, the model's $\theta = 1$), the spread response would therefore be *positive*—$\widehat{c}_n = +\widehat{\Pi} \approx +0.13$, since a tightening then raises long yields fully while the bill yield lags by the widening premium—whereas the estimated $\widehat{c}_n$ are strongly negative.

For transparency, Table A.1 reports the loadings themselves alongside the spreads: $\widehat{\beta}_{\text{repo}} = 0.813$ (0.044), $\widehat{\beta}_b = 0.688$ (0.048), $\widehat{\theta}_5 = 0.136$ (0.096), and $\widehat{\theta}_{10} = 0.009$ (0.090). The long-yield loadings are positive and well below one in point estimate but individually imprecise, and the contrast with the tightly estimated spreads is mechanical: monthly Treasury-yield innovations are dominated by a common level factor (Litterman and Scheinkman 1991), which is removed by differencing two points on the curve, so the composite $\theta_n - \beta_b$ is estimated far more precisely than either loading alone. Two implications. (i) At the implied $\widehat{\rho}$ of Section 4, Proposition 1's condition $1 - \theta > \rho\bar{\kappa}$ is non-binding for any $\theta \leq 1 + |\rho|\bar{\kappa} \approx 1.1$; and the boundary case itself (one-for-one co-movement) is the sign-rejected benchmark above. The sample economy is far from the region, discussed in Section 5 and the Remark in A.2, where $\delta$ could change sign. (ii) The spread response is Link 1 of the chain: policy moves the Treasury's relative financing cost on impact, in the model's direction.

*Step 2: relative cost → announced composition.* Bill offering amounts are announced weekly and settle within days; coupon amounts are revised only around quarterly refundings and by small increments. A change in relative cost can therefore begin to shift the *announced* composition on the bill margin within weeks, while the coupon margin adjusts across refunding cycles. Thus, the policy rate moves the relative cost on impact (Step 1); announced bill offerings adjust at the weekly profile; and outstanding stocks (the variable in Table 1) cumulate those announcements, since at the monthly frequency the bill stock is, up to predetermined redemptions, the integral of recently announced offerings.

*Step 3: dynamics.* Local projections show the cumulation:

$$b_{t+h} - b_{t-1} = \alpha_h + \delta_h \, \Delta i_t + \Gamma_h' X_t + u_{t+h}, \qquad h = 0, \ldots, 24,$$

with $\Delta i_t$ instrumented by the futures surprise, $X_t$ the month dummies and the VIX, and Newey–West standard errors with $12 + h$ lags; by construction $\delta_0$ coincides with column (4)'s $\widehat{\delta}$. Figure A.1 reports the path: the response is $\approx -0.03$ log points on impact, cumulates to $-0.365$ log points by month 11. The gradual build matches the announcement calendar (weekly bill-size adjustments compounding across months and refunding cycles) rather than a one-off portfolio jump, and its persistence indicates a change in the financing mix and not a transitory cash-management fluctuation.

Taken together: the institutional record identifies the bill margin as the locus of discretionary composition choice and its objective as cost minimization under smoothing; the mediation test

Table A.1: The issuance margin: relative cost and composition

| | (A1) | (A2) | (A3) | (A4) |
|---|---|---|---|---|
| Dep. var. | $\Delta(i_t^{5y} - i_t^b)$ | $\Delta(i_t^{10y} - i_t^b)$ | $\Delta b_t$ | $\Delta[b_t - f_t]$ |
| Estimator | 2SLS | 2SLS | 2SLS | 2SLS |
| $\Delta$Policy rate | −0.552 | −0.679 | −0.0312 | −0.0245 |
| | (0.086) | (0.086) | (0.0149) | (0.0136) |
| $\Delta$ log(total debt) | | | 0.135 | |
| | | | (0.227) | |
| *Memo: loadings on the instrumented rate* | | | | |
| long yield $\widehat{\theta}_n$ | 0.136 | 0.009 | | |
| | (0.096) | (0.090) | | |
| bill yield $\widehat{\beta}_b$ | 0.688 | (0.048) | | |
| repo $\widehat{\beta}_{\text{repo}}$ | 0.813 | (0.044) | | |
| Estimand | $\theta_5 - \beta_b$ | $\theta_{10} - \beta_b$ | $\delta_F$ | tilt |
| First-stage $F$ | 30 | 30 | 30 | 30 |
| Observations | 211 | 211 | 211 | 211 |

Monthly, 1991m5–2008m11; $\Delta \equiv$ 1-month change; identical instruments, controls (VIX, month dummies as applicable), and sample as Table 1. Newey–West SEs (12 lags). $i^b$ is the 3-month bill yield; $f_t \equiv$ log total marketable debt (MSPD). Each spread coefficient equals $\widehat{\theta}_n - \widehat{\beta}_b$ exactly, and $\widehat{\beta}_b = \widehat{\beta}_{\text{repo}} - \widehat{\Pi}$ by the construction of $LP$, tying (A1)–(A2) to Table 1 (Appendix A.6(c)).

of part (b) rejects pure residual financing; and the estimated chain in part (c) characterizes the sequence of the mechanism's events.

## References for the online appendix


Belton, T., Dawsey, K., Greenlaw, D., Li, H., Ramaswamy, S., Sack, B., 2018. Optimizing the maturity structure of U.S. Treasury debt: A model-based framework. Hutchins Center Working Paper 46, Brookings Institution.

Garbade, K., 2007. The emergence of "regular and predictable" as a Treasury debt management strategy. Federal Reserve Bank of New York Economic Policy Review 13(1), 53–71.

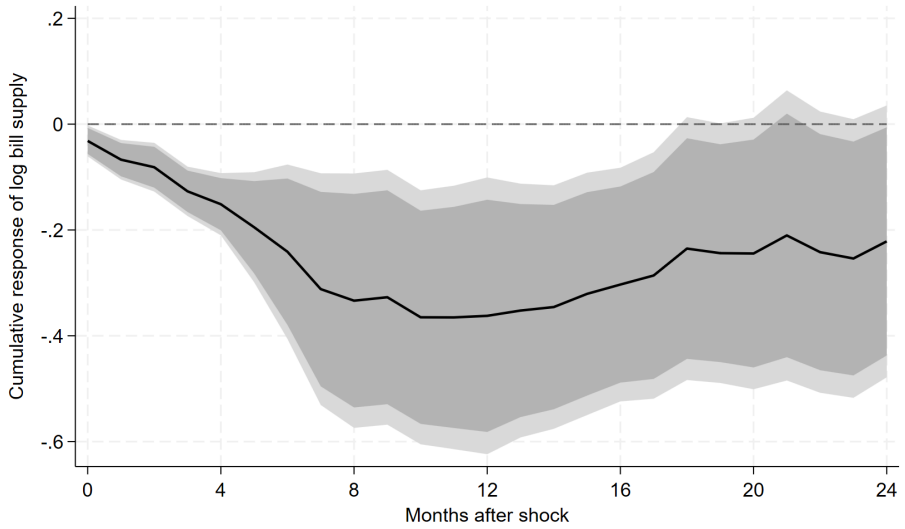
Garbade, K., 2015. Treasury Debt Management under the Rubric of Regular and Predictable Issuance: 1983–2012. Federal Reserve Bank of New York.

Garbade, K., Partlan, J., Santoro, P., 2004. Recent innovations in Treasury cash management. Federal Reserve Bank of New York Current Issues in Economics and Finance 10(11).

Kuttner, K.N., 2001. Monetary policy surprises and interest rates: Evidence from the Fed funds futures market. Journal of Monetary Economics 47, 523–544.

Litterman, R.B., Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income 1(1), 54–61.

Figure A.1: Dynamic response of bill supply to the policy rate (local projections)




Notes: cumulative response $b_{t+h} - b_{t-1}$ to a 1 percentage point instrumented increase in the policy rate, $h = 0, \ldots, 24$ months; futures-surprise instrument; controls: month dummies and VIX; shaded areas are 90% and 95% Newey–West $(12 + h$ lags) confidence bands. The impact response coincides with $\widehat{\delta}$ of Table 1, column (4).