Relative Predictors and Logistic Model for Future Power

1 Construct relative predictors (US vs. UK)

Let t index years. For any variable X, define the *log ratio*

$$x_t \equiv \ln \frac{X_t^{\mathrm{US}}}{X_t^{\mathrm{UK}}},$$

and for shares in $(0,1)$ use *log-odds*.

Affordability (wheat). Given $A_t^{\mathrm{US}} = p_t^{\mathrm{US}} / y_t^{pc,\mathrm{US}}$ and A_t^{UK},

$$a_t \equiv \ln \frac{A_t^{\mathrm{US}}}{A_t^{\mathrm{UK}}} \quad (\text{higher} \Rightarrow \text{wheat less affordable in US vs UK}).$$

Often flip the sign so "cheaper in US" is positive: $\tilde{a}_t \equiv -a_t$.

GDP (level, not per capita).

$$g_t \equiv \ln \frac{\mathrm{GDP}_t^{\mathrm{US}}}{\mathrm{GDP}_t^{\mathrm{UK}}}.$$

Military/industrial capacity (choose one primary: iron–steel, energy, or military expenditure).

$$m_t \equiv \ln \frac{\mathrm{ISPR}_t^{\mathrm{US}}}{\mathrm{ISPR}_t^{\mathrm{UK}}} \quad (\text{replace ISPR with chosen metric}).$$

Education (e.g., avg. years, enrollment, literacy).

$$e_t \equiv \ln \frac{\mathrm{Edu}_t^{\mathrm{US}}}{\mathrm{Edu}_t^{\mathrm{UK}}}.$$

Urbanization share. For $s_t^{\mathrm{US}}, s_t^{\mathrm{UK}} \in (0,1)$ use a log-odds differential:

$$u_t \equiv \underbrace{\ln \frac{s_t^{\mathrm{US}}}{1 - s_t^{\mathrm{US}}}}_{\mathrm{logit}(s_t^{\mathrm{US}})} - \underbrace{\ln \frac{s_t^{\mathrm{UK}}}{1 - s_t^{\mathrm{UK}}}}_{\mathrm{logit}(s_t^{\mathrm{UK}})}.$$

(A plain log ratio $\ln(s_t^{\mathrm{US}} / s_t^{\mathrm{UK}})$ also works, but log-odds behaves better near 0/1.)

Add lags to reduce simultaneity: use $x_{t-\ell}$ with $\ell \in \{1, 2\}$.

2 Define a binary "future power" target

Pick a horizon h (e.g., 5 or 10 years). Let the *relative power score* be R_t:

- **GDP leader (PTT-lite):** $R_t \equiv \ln\big(\mathrm{GDP}_t^{\mathrm{US}}/\mathrm{GDP}_t^{\mathrm{UK}}\big)$.
- **Composite (optional):** $R_t = w_g\, g_t + w_m\, m_t + w_u\, u_t$ (standardize components, set w by PCA or equal weights).

  Two binary targets:

$$\begin{aligned}
\textbf{(A) Leader at } t+h: &\quad Y_t^{(h)} = \mathbf{1}\{R_{t+h} \geq 0\} \quad (= 1 \text{ if US leads at } t+h). \\
\textbf{(B) Overtake in } (t, t+h]: &\quad Y_t^{(h)} = \mathbf{1}\{R_t < 0 \ \wedge \ R_{t+h} \geq 0\}\,.
\end{aligned}$$

(A) predicts who leads at horizon h; (B) predicts overtake risk within the window. With two countries and $\sim$110 years, (A) is simpler and more stable.

3 Logistic specification

Let $X_t = \big[\tilde{a}_{t-\ell},\, g_{t-\ell},\, m_{t-\ell},\, e_{t-\ell},\, u_{t-\ell}\big]$. Include a smooth time trend $f(t)$ (e.g., cubic spline) to absorb global drift:

$$\Pr\big(Y_t^{(h)} = 1 \mid X_t\big) = \Lambda\Big(\beta_0 + \beta_a\, \tilde{a}_{t-\ell} + \beta_g\, g_{t-\ell} + \beta_m\, m_{t-\ell} + \beta_e\, e_{t-\ell} + \beta_u\, u_{t-\ell} + f(t)\Big),$$

where $\Lambda(z) = \frac{1}{1+e^{-z}}$.

Interpretation. $\beta_g > 0$: a higher US/UK GDP ratio today raises the probability the US leads at $t+h$. $\beta_a > 0$: cheaper wheat in the US (relative affordability positive) raises that probability (structural-change channel). Expect $\beta_m > 0$, $\beta_u > 0$, $\beta_e > 0$.

Trend. $f(t)$: natural cubic spline with 3–5 df. With only two countries, do not include country fixed effects (collinear).

Estimation. No weights needed; use robust SEs (HC1).

4 Output: a predictive variable of future relative power

The model returns

$$\hat{q}_t^{(h)} \equiv \widehat{\Pr}\big(Y_t^{(h)} = 1 \mid X_t\big) \in [0, 1],$$

interpretable as the probability the US is the leader at $t+h$ (or the overtake risk in the hazard version).

If a smooth *power score* is preferred, map the same index via a rescaled logistic:

$$S_t^{(h)} \equiv 2\,\hat{q}_t^{(h)} - 1 \in [-1, 1],$$

so > 0 favors US, < 0 favors UK.

5 Practical fitting details (tiny panel)

- **Horizon/lag grid:** try $h \in \{5, 10\}$, $\ell \in \{1, 2\}$.
- **Regularization:** with ∼100 obs, add ridge penalty:

$$\max_{\beta} \sum_t \ell\big(\beta; Y_t, X_t\big) - \lambda \|\beta\|_2^2.$$

- **Validation:** time-based CV (rolling origin) to avoid look-ahead.
- **Robustness:** swap m_t across ISPR/ECR/MER; replace GDP with CINC share if later added.

6 Minimal data pipeline (pseudocode)

1. **Build relative predictors:**

$$\tilde{a}_t = -\ln \frac{A_t^{\mathrm{US}}}{A_t^{\mathrm{UK}}}, \quad g_t = \ln \frac{\mathrm{GDP}_t^{\mathrm{US}}}{\mathrm{GDP}_t^{\mathrm{UK}}}, \quad m_t = \ln \frac{M_t^{\mathrm{US}}}{M_t^{\mathrm{UK}}},$$

$$e_t = \ln \frac{\mathrm{Edu}_t^{\mathrm{US}}}{\mathrm{Edu}_t^{\mathrm{UK}}}, \quad u_t = \operatorname{logit}(s_t^{\mathrm{US}}) - \operatorname{logit}(s_t^{\mathrm{UK}}).$$

2. **Create target $Y_t^{(h)}$:**

$$R_{t+h} = g_{t+h} \quad \text{(or composite)}, \qquad Y_t^{(h)} = \mathbf{1}\{R_{t+h} \geq 0\},$$

   then drop the last h years.

3. **Assemble design** with lag ℓ, add spline $f(t)$.
4. **Fit** logistic (optionally ridge), **predict** $\hat{q}_t^{(h)}$.
5. **Use** $\hat{q}_t^{(h)}$ as the future power predictor; date "inflections" where $\hat{q}_t^{(h)}$ crosses 0.5 or where $\Delta\hat{q}_t^{(h)}$ peaks.