A Behavioral Explanation for the Puzzling Persistence of the Aggregate Real Exchange Rate[*]

Mario J. Crucini[†], Mototsugu Shintani[‡], and Takayuki Tsuruga[§]

May 13, 2025

Abstract

At the aggregate level, the observation that deviations from purchasing power parity (PPP) are too persistent to be accounted for solely by nominal rigidities has long been a puzzle (Rogoff, 1996). In addition, microeconomic evidence suggests that deviations from the law of one price (LOP) are less persistent than PPP deviations. To reconcile these two empirical anomalies, we incorporate the behavioral inattention approach of Gabaix (2014) into a two-country sticky-price model. Our model shows that firms' behavioral inattention to the aggregate component of real marginal costs generates an endogenous dependence of LOP deviations on PPP deviations. We find strong supporting evidence for this particular formulation of behavioral inattention. Calibrating our model with the estimated degree of attention, we show that our model can fully account for the two empirical anomalies. PPP deviations are more than twice as persistent as those implied by nominal rigidities alone, while the persistence of LOP deviations is about two-thirds that of PPP deviations.

JEL Classification: E31; F31; D40

Keywords: Real exchange rates, Law of one price, Purchasing power parity, Rational inattention, Trade cost

[*]We thank M. Baxter, D. Cook, M. Devereux, I. Fujiwara, Y. Hirose, J. Kim, A. Shibata, K. Shin, and participants at numerous conferences and seminars for their helpful comments and suggestions. We also thank E. Gautier for kindly providing us with the data on the frequency of price changes in the Euro area. Shintani and Tsuruga gratefully acknowledge the financial support of Grants-in-aid for Scientific Research (JP20H05633, JP21H04397, JP24K00241, JP25H00544, and JP25K00622), the Joint Usage/Research Centers at the ISER (The University of Osaka) and the KIER (Kyoto University).

[†]Department of Economics, Daniels School of Business, Purdue University and NBER; e-mail: mcrucini@purdue.edu.

[‡]Faculty of Economics, The University of Tokyo; e-mail: shintani@e.u-tokyo.ac.jp.

[§]Institute of Social and Economic Research, The University of Osaka and CAMA; e-mail: tsuruga@iser.osaka-u.ac.jp.

1 Introduction

It is well established that the aggregate real exchange rate (RER), the deviation from purchasing power parity (PPP), exhibits a high degree of persistence. Rogoff (1996) characterizes this empirical anomaly as the "PPP puzzle," noting that "Consensus estimates for the rate at which PPP deviations damp, however, suggest a half-life of three to five years, seemingly far too long to be explained by nominal rigidities" (p. 648).[1] A closely related feature of RERs is the gap in persistence between PPP deviations and deviations from the law of one price (LOP), the fundamental building block of PPP. Imbs et al. (2005) and Carvalho and Nechio (2011) argue that good-level RERs (i.e., LOP deviations) tend to be significantly less persistent than aggregate RERs (i.e., PPP deviations).[2] These studies emphasize the role of heterogeneity in the speed of price adjustment. As Imbs et al. (2005) assert, "It is this heterogeneity that we find to be an important determinant of the observed real exchange rate persistence since it gives rise to highly persistent aggregate series while relative price persistence is low on average at a disaggregated level" (p. 3).

In this paper, we address two empirical anomalies simultaneously: (1) the gap between the observed persistence of PPP deviations and the persistence predicted by nominal rigidities and (2) the gap between the observed persistence of PPP deviations and that of LOP deviations. We incorporate behavioral inattention into a two-country sticky-price model, following Gabaix (2014). In this framework, firm managers incur costs when paying attention to the real marginal cost of their products. As a result, full attention to the state of the economy is no longer optimal when setting prices.

The key to solving the PPP puzzle lies in strategic complementarity in pricing. After deriving the dynamic equation for good-level RERs, we show that behavioral inattention causes good-level RERs to depend on the aggregate RER through this strategic complementarity. We refer to this relationship as *aggregate RER dependence*. The dynamic equation implies that when firms only pay partial attention to the aggregate component of real marginal costs, good-level RERs respond to changes in the aggregate RER. Consequently, greater persistence in the aggregate RER leads to greater persistence in good-level RERs. Through aggregation, this feedback further reinforces the persistence of the aggregate RER, strengthening the link between the aggregate and good-level RERs. Although our model differs in setup, this mechanism is related to real rigidities in Ball and Romer (1990) and strategic complementarity

[1] Unless stated otherwise, in our exposition, nominal rigidities are those which arise from Calvo time-dependent pricing formulations.

[2] For a comprehensive empirical analysis of the persistence in LOP deviations, see Crucini and Shintani (2008).

emphasized in Woodford (2003) in closed-economy settings.

The dynamic equation yields a directly testable implication. We test the null hypothesis of aggregate RER independence against the alternative hypothesis of aggregate RER dependence, using micro price data from the US, Canada, and European countries. This test is equivalent to assessing whether good-level RERs are uncorrelated with the aggregate RER. Across various specifications, we strongly reject the null in favor of our proposed model of behavioral inattention. We also estimate the degree of attention to be approximately 0.15, substantially lower than the value of 1.0 under full attention.

Two theoretical results emerge under behavioral inattention. First, the model of behavioral inattention explains the gap between the observed persistence of the aggregate RER and the persistence predicted by nominal rigidities alone. In our model, even small nominal frictions can generate a highly persistent aggregate RER. Our estimates suggest that the aggregate RER is more than twice as persistent as it would be under nominal rigidities alone. In terms of half-lives, our model of behavioral inattention successfully replicates the empirically observed three- to five-year half-lives, whereas the sticky-price model under full attention predicts substantially shorter half-lives.

Second, our model accounts for the gap between the highly persistent aggregate RER and the less persistent good-level RERs. This gap arises from the interaction between aggregate RER dependence and idiosyncratic real shocks to individual goods prices. Both aggregate and good-level RERs exhibit greater persistence under behavioral inattention. However, while real shocks at the goods level reduce the persistence of good-level RERs, they do not affect the persistence of the aggregate RER, as aggregation across goods cancels out the effects of idiosyncratic shocks. Consequently, our estimated degree of attention generates a substantial gap in persistence between the aggregate and good-level RERs. Indeed, our model predicts that incorporating inattention into an otherwise standard sticky-price model reduces the persistence of the good-level RER to less than two-thirds that of the aggregate RER. Furthermore, our model of behavioral inattention explains two related findings: (1) good-level RERs are more persistent than the degree of price stickiness suggests (Kehoe and Midrigan, 2007), and (2) good-level RERs exhibit persistence comparable to that of the aggregate RER when focusing only on macroeconomic shocks (Bergin et al., 2013).

We discuss why behavioral inattention is the preferred framework for addressing the PPP puzzle compared to other potential alternatives. It is well known that several alternative economic mechanisms can generate strategic complementarities. For example, Kehoe and Midrigan (2007) incorporate Basu's (1995) framework of roundabout production to explore the role of strategic complementarity in a two-country sticky-price model. Since the behavioral inattention model also gives rise to strategic complementarity, we compare it with the roundabout production model. Our numerical exercises show that the roundabout production model fails to generate the observed persistence of the aggregate RER, even under an extremely high degree of roundabout production. In contrast, we show that behavioral inattention more effectively generates the degree of persistence observed in the aggregate RER.

A standard explanation for why aggregate RERs are more persistent than good-level RERs is heterogeneity in the speed of price adjustment across goods, which introduces an upward bias when prices are aggregated to construct the consumer price index (CPI). Imbs et al. (2005) highlight this aggregation bias in dynamic heterogeneous panels. Carvalho and Nechio (2011) examine its theoretical implications using a multisector sticky-price model in which price stickiness varies across sectors. We note, however, that the aggregation bias in their multisector sticky-price model may play only a limited role in explaining the PPP puzzle under an empirically plausible process of the nominal exchange rate (NER). In contrast, our model of behavioral inattention adequately accounts for the PPP puzzle under a realistic stochastic process for the NER, even when the persistence of good-level RERs is homogeneous across goods.

We also explore an alternative model of inattention by examining rational inattention. In Maćkowiak and Wiederholt (2009), firms set prices based on noisy signals about nominal aggregate demand and idiosyncratic productivity, subject to an information processing capacity constraint. Depending on assumptions, the dynamic equation from our model and theirs can generate observationally equivalent dynamics for good-level RERs. Nevertheless, we argue that behavioral inattention offers greater flexibility in terms of which variables firms pay attention to, making it particularly suited for addressing both the PPP and LOP puzzles.

The observation that the persistence of the aggregate RER exceeds that implied by nominal rigidities connects to a broad literature that has significantly advanced our understanding of persistent aggregate RERs. For instance, Chari et al. (2002) argue that while the sticky-price model with monetary shocks can explain the volatility of the aggregate RER, it substantially underpredicts its persistence. Benigno (2004) highlights the role of monetary policy rules over the degree of price stickiness in explaining the persistence of the aggregate RER. Later, Engel (2019) revisits Benigno (2004) and emphasizes the importance of both monetary policy rules and price stickiness. More recently, Itskhoki and Mukhin (2021) emphasize that financial shocks, rather than monetary shocks, play a dominant role in resolving the PPP puzzle. In our model, a conventional monetary shock remains the main driver of PPP deviations. However, idiosyncratic productivity shocks are necessary to explain why good-level RERs are less persistent than the aggregate RER. Thus, both Itskhoki and Mukhin (2021) and this paper share the view that monetary shocks alone are insufficient to resolve the PPP puzzle.

The remainder of the paper is structured as follows. In Section 2, we present a two-country sticky-price model and introduce behavioral inattention. Section 3 introduces the dynamic equation for the good-level RER and discusses the implications of behavioral inattention in firms' pricing. In Section 4, we implement a test for aggregate RER independence in the context of our model and estimate the degree of behavioral inattention. In Section 5, we assess how much the estimated degree of behavioral inattention can improve model predictions. In Section 6, we discuss alternative explanations for the puzzling persistence of RERs in comparison to our model. Section 7 concludes.

2 The model

The world economy consists of two countries: the United States (home) and Canada (foreign). Following Kehoe and Midrigan (2007) and Crucini et al. (2010b, 2013), there is a continuum of goods and brands of each good. Goods are indexed by $i \in [0, 1]$. Within each good, US brands are indexed by $z \in [0, 1/2]$, while Canadian brands are indexed by $z \in (1/2, 1]$. To conserve space, we primarily present equations for the US economy and avoid repeating them for Canada whenever possible.

We assume that US and Canadian households have identical preferences over brands of a particular good and across goods in the aggregate consumption basket. US household preferences over brands of good i are represented by a constant elasticity of substitution (CES) index. US consumption of good i is given by $c_{it} = \left[\int_{z=0}^{1} c_{it}(z)^{\frac{\varepsilon-1}{\varepsilon}} dz\right]^{\frac{\varepsilon}{\varepsilon-1}}$, and aggregation across goods yields aggregate consumption $c_t = \left[\int_{i=0}^{1} c_{it}^{\frac{\varepsilon-1}{\varepsilon}} di\right]^{\frac{\varepsilon}{\varepsilon-1}}$, where $\varepsilon > 1$. For Canada, we have the analogous expressions for c_{it}^* and c_t^*, where asterisks $(*)$ denote the place of households' consumption.

2.1 Households

The objective of the US households is to maximize $\mathbb{E}_0 \sum_{t=0}^{\infty} \delta^t U(c_t, n_t) = \mathbb{E}_0 \sum_{t=0}^{\infty} \delta^t (\ln c_t - \chi n_t)$, subject to an intertemporal budget constraint,

$$M_t + \mathbb{E}_t(\Delta_{t,t+1} B_{t+1}) = W_t n_t + B_t + M_{t-1} - P_{t-1} c_{t-1} + T_t + \Pi_t, \tag{1}$$

and a cash-in-advance (CIA) constraint $M_t \geq P_t c_t$. Here, $\mathbb{E}_t(\cdot)$ denotes the expectation operator conditional on the information available in period t, $\delta \in (0,1)$, and $\chi > 0$. In addition, we suppress the state contingencies for notational convenience. The left-hand side of (1) represents the total nominal value of household wealth. The household allocates its wealth into money balances M_t for the purchase of consumption goods and into state-contingent nominal bond holdings B_{t+1}, brought into period $t+1$. Here, $\Delta_{t,t+1}$ denotes the nominal stochastic discount factor. On the right-hand side of the budget constraint (1), the household receives a nominal wage W_t per hour of work n_t, carries bonds B_t into period t, as well as any cash that remained in period $t-1$, $M_{t-1} - P_{t-1} c_{t-1}$. The household also receives nominal transfers from the US government, T_t, and nominal profits from US firms, Π_t. In (1), the aggregate price P_t is given by $P_t = \left[\int P_{it}^{1-\varepsilon} di\right]^{\frac{1}{1-\varepsilon}}$, where P_{it} is the price index for good i. This, in turn, is a CES aggregate over US and Canadian brands: $P_{it} = \left[\int P_{it}(z)^{1-\varepsilon} dz\right]^{\frac{1}{1-\varepsilon}}$. The CIA constraint requires nominal money balances for expenditure, which is made at the end of period t. The CIA constraint always binds with equality in equilibrium.

The first-order conditions of the US households are

$$\frac{W_t}{P_t} = \chi c_t, \quad \frac{M_t}{P_t} = c_t, \quad \Delta_{t,t+1} = \delta \left[\left(\frac{c_{t+1}}{c_t} \right)^{-1} \left(\frac{P_t}{P_{t+1}} \right) \right], \tag{2}$$

which are the labor supply condition, the CIA constraint, and the consumption Euler equation, respectively.

Canadian households solve the analogous maximization problem. We assume that complete markets exist for state-contingent financial claims across the US and Canada, and that these financial claims are denominated in US dollars. Thus, we convert US dollar bond holdings into Canadian dollars using the spot nominal exchange rate (NER), S_t. The Canadian households are subject to the budget constraint $M_t^* + \mathbb{E}_t(\Delta_{t,t+1} B_{t+1}^*)/S_t = W_t^* n_t^* + B_t^*/S_t + M_{t-1}^* - P_{t-1}^* c_{t-1}^* + T_t^* + \Pi_t^*$ and an analogous CIA constraint in Canadian dollars. The first-order conditions are similar, except for their consumption Euler equations. Because Canadians buy state-contingent bonds denominated in US dollars, their consumption Euler equation is $\Delta_{t,t+1} = \delta \left\{ \left(c^*_{t+1}/c^*_t\right)^{-1} \left[(S_t P^*_t)/(S_{t+1} P^*_{t+1})\right] \right\}$.

The aggregate RER is defined as $q_t = S_t P^*_t / P_t$. The consumption Euler equations imply $q_{t+1}(c^*_{t+1}/c_{t+1}) = q_t(c^*_t/c_t) = ... = q_0(c^*_0/c_0)$. Normalizing $q_0(c^*_0/c_0)$ to unity yields[3]

$$q_t = \frac{c_t}{c^*_t}. \tag{3}$$

The log of the NER is assumed to follow a random walk, as is often observed in the data. Kehoe and Midrigan (2007) show that (3) and the CIA constraints in both countries imply $S_t = P_t c_t/(P^*_t c^*_t) = M_t/M^*_t$. This equation suggests that $\ln S_t$ follows a random walk if the monetary authority in each country sets the log of the money supply according to,

$$\ln M_t = \ln M_{t-1} + \varepsilon^M_t, \tag{4}$$

$$\ln M^*_t = \ln M^*_{t-1} + \varepsilon^{M^*}_t, \tag{5}$$

where ε^M_t and $\varepsilon^{M^*}_t$ are zero-mean i.i.d. shocks. In this case, $S_t = M_t/M^*_t$ leads to $\ln S_t = \ln S_{t-1} + (\varepsilon^M_t - \varepsilon^{M^*}_t)$.[4]

2.2 Firms

For each good, US firms produce the first half of the continuum, $z \in [0, 1/2]$, of good i and employ $n_{it}(z)$ hours of labor, while Canadian firms produce the second half, $z \in (1/2, 1]$, and employ $n^*_{it}(z)$. The production function of US firms is given by $y_{it}(z) = a_{it} n_{it}(z)$, whereas that of the Canadian firms is given by $y^*_{it}(z) = a^*_{it} n^*_{it}(z)$. Here, a_{it} and a^*_{it} represent labor productivity specific to good i. In both the US and Canada, all firms producing varieties of the same good share the same productivity, but productivity across countries differs by good.

We assume that the log of labor productivity follows a zero-mean i.i.d. process:[5]

$$\ln a_{it} = \varepsilon^a_{it}, \tag{6}$$

$$\ln a^*_{it} = \varepsilon^{a^*}_{it}. \tag{7}$$

Labor productivity is good-specific and uncorrelated with aggregate variables.

[3]This condition relies on our preference assumptions, which we relax in Section 4.

[4]In Section 6, we will discuss the robustness when the NER is determined by the uncovered interest parity and the policy interest rates are determined by the Taylor rule.

[5]In Section 4, we will consider an alternative stochastic process for labor productivity. We will show that the implications for the test for aggregate RER independence remain unchanged.

Goods that are shipped between the US and Canada are subject to iceberg trade costs, τ.[6] Following Kehoe and Midrigan (2007) and Crucini et al. (2010b, 2013), goods are nondurable. Thus, the production of good i undertaken in the US is exhausted between US and Canadian consumption, with Canadian imports of US brands subject to an iceberg trade cost:

$$c_{it}(z) + (1+\tau)c^*_{it}(z) = y_{it}(z), \text{ for } z \in [0, 1/2]\,. \tag{8}$$

Similarly, the production of good i undertaken in Canada is exhausted between Canadian and US consumption, with US imports of Canadian brands subject to an iceberg trade cost:

$$(1+\tau)c_{it}(z) + c^*_{it}(z) = y^*_{it}(z), \text{ for } z \in (1/2, 1]. \tag{9}$$

2.3 Price setting under behavioral inattention

In this subsection, we incorporate behavioral inattention into a standard two-country model with Calvo pricing. Firms set prices in the buyers' currency, known as local currency pricing. We first present the attention-augmented objective function, then describe how firms determine optimal reset prices and their degree of attention. Since Canadian firms face a similar pricing problem, we focus on US firms' decisions.

2.3.1 The attention-augmented objective function

Consider the US firm's period-by-period real profits from selling its brands in the US market. Let $p_{it}(z) = P_{it}(z)/P_t$, $p_{it} = P_{it}/P_t$ and $w_t = W_t/P_t$ denote the real price of brand z of good i, the real price of good i and the real wage, respectively. The firm's real profits are given by $(1/P_t)\,[P_{it}(z) - W_t/a_{it}]\,c_{it}(z) = [p_{it}(z) - w_t/a_{it}]c_{it}(z)$. The demand by US consumers for a particular brand of good i is $c_{it}(z) = [p_{it}(z)/p_{it}]^{-\varepsilon}c_{it}$.

We define the log deviation of a generic variable x_t from the steady-state level as $\hat{x}_t = \ln x_t - \ln \bar{x}$, where $\bar{x}$ denotes the steady-state level of x_t. This allows us to express $x_t = \bar{x}\exp(\hat{x}_t)$. Applying this equation to the US firm's real profits from selling the brand in the US market, we obtain $\{\bar{p}_i(z)\exp[\hat{p}_{it}(z)] - \bar{w}\exp(\hat{w}_t - \hat{a}_{it})\}c_{it}(z)$. In terms of the log deviation, the demand function for $c_{it}(z)$ is written as $c_{it}(z) = (\bar{p}_i(z)/\bar{p}_i)^{-\varepsilon}(-\varepsilon[\exp(\hat{p}_{it}(z) - \hat{p}_{it})])c_{it}$.

We assume that firms cannot change their prices with probability λ, as in Calvo (1983) and Yun (1996). This parameter captures the degree of price stickiness. The objective function

[6] As we elaborate in Section 3, trade costs are important because they lead to a home bias in the expenditure share of consumption goods, which in turn becomes a source of LOP deviations due to variations in labor productivity between the US and Canada.

that a fully attentive US firm maximizes is:

$$v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} \\ \times \frac{P_t}{P_{t+k}} \left\{ \bar{p}_i(z) \exp\left[\hat{p}_{it}(z)\right] - \bar{w} \exp\left(\hat{\mu}_{Ht+k} - \hat{a}_{it+k}\right) \right\} c_{it,t+k}(z), \tag{10}$$

where

$$c_{it,t+k}(z) = \left[\frac{\bar{p}_i(z)}{\bar{p}_i}\right]^{-\varepsilon} \exp\left\{-\varepsilon \left[\hat{p}_{it}(z) - \sum_{l=1}^{k} \pi_{t+l} - \hat{p}_{it+k}\right]\right\} c_{it+k} \tag{11}$$

is the demand for brand z of good i in period $t+k$, conditional on the firm having last reset the price in period t. In (10) and (11), $\hat{\mu}_{Ht+k} = \hat{w}_{t+k} + \sum_{l=1}^{k} \pi_{t+l}$ and $\pi_t = \ln(P_t/P_{t-1})$, both derived under the assumption that steady-state inflation is zero.[7] Here, $v_{it}(z)$ represents the present discounted value of real profits accruing to the firm producing brand z of good i in the US, conditional on the firm having last reset its price in period t. In (10), the second line represents real profits in each period. These profits are discounted by the stochastic discount factor $\delta_{t,t+k} = \delta^k (c_{t+k}/c_t)^{-1}$, which satisfies $\delta_{t,t+k} P_t/P_{t+k} = \Delta_{t,t+k}$, and by the probability λ^k. The real marginal cost consists of the aggregate component $\hat{\mu}_{Ht+k}$ and the idiosyncratic component $\hat{a}_{it+k}$. However, due to price stickiness, the former is adjusted by inflation accumulated from periods t to $t+k$, namely, $\sum_{l=1}^{k} \pi_{t+l}$. In (11), real prices are also adjusted by inflation accumulated from period t to $t+k$. Note that this objective function applies to US firms indexed by $z \in [0, 1/2]$.

Our model assumes that firms are not fully attentive to their real marginal cost, which consists of the aggregate component $\hat{\mu}_{Ht+k}$ and the idiosyncratic component $\hat{a}_{it+k}$. We define $\boldsymbol{m}_H = (m_{1H}, m_{2H})' \in [0,1]^2$, where m_{1H} and m_{2H} represent the degree of attention to $\hat{\mu}_{Ht+k}$ and $\hat{a}_{it+k}$, respectively, for setting prices of their home-produced brands in US markets. Here, the subscript H denotes the place of production.[8] Following Gabaix (2014, 2019, 2020), we assume that inattentive firms replace $v_{it}(z)$ with the *attention-augmented* objective function.

[7] The details of the derivation are provided in Appendix A.1.

[8] Likewise, we define m_{1H}^* and m_{2H}^* as the degree of attention to the aggregate and idiosyncratic components, respectively, for setting prices of US-produced brands in Canadian markets. The degree of attention for pricing foreign-produced goods is represented by m_{1F}^* and m_{2F}^* when selling brands in Canadian markets, and by m_{1F} and m_{2F} when selling brands in US markets.

This function is given by

$$v_{Hi}\left(\hat{p}_{it}(z), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H\right) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k}$$
$$\times \frac{P_t}{P_{t+k}} \left\{\bar{p}_i(z) \exp\left[\hat{p}_{it}(z)\right] - \bar{w} \exp\left(m_{1H}\hat{\mu}_{Ht+k} - m_{2H}\hat{a}_{it+k}\right)\right\} c_{it,t+k}(z), \qquad (12)$$

where $\hat{\boldsymbol{x}}_{Hit}$ is the vector of state variables, defined as $\hat{\boldsymbol{x}}_{Hit} = (\hat{\mu}_{Ht}, \hat{\mu}_{Ht+1}, ..., \hat{a}_{it}, \hat{a}_{it+1}, ...)'$. In the limiting case of $\boldsymbol{m}_H = \boldsymbol{0} = (0, 0)'$, firm managers completely ignore changes in both $\hat{\mu}_{Ht+k}$ and $\hat{a}_{it+k}$. Conversely, in the case of full attention, where $\boldsymbol{m}_H = \boldsymbol{\iota} = (1, 1)'$, the attention-augmented objective function reduces to (10).[9]

The inattentive US firms' real profits from selling their brand in the Canadian market are defined analogously to (12). Let $p^*_{it}(z)$ denote the real price in Canadian markets, given by $p^*_{it}(z) = P^*_{it}(z)/P^*_t$. Their real profits are expressed as $(1/P_t)\left[S_t P^*_{it}(z) - (1+\tau)W_t/a_{it}\right] c^*_{it}(z) = q_t\left[p^*_{it}(z) - (1+\tau)w_t/(q_t a_{it})\right] c^*_{it}(z)$, where $(1+\tau)w_t/(q_t a_{it})$ represents the exporting firms' real marginal cost, measured in Canadian goods.[10] In terms of log deviations, their real profits are given by $q_t\left\{\bar{p}^*_i \exp[\hat{p}^*_{it}(z)] - (1+\tau)(\bar{w}/\bar{q})\exp(\hat{w}_t - \hat{q}_t - \hat{a}_{it})\right\} c^*_{it}(z)$. The attention-augmented objective function that an inattentive US firm maximizes is:[11]

$$v^*_{it}\left(\hat{p}^*_{it}(z), \hat{\boldsymbol{x}}^*_{Hit}, \boldsymbol{m}^*_H\right) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} q_{t+k}$$
$$\times \frac{P^*_t}{P^*_{t+k}} \left[\bar{p}^*_i(z) \exp\left[\hat{p}^*_{it}(z)\right] - (1+\tau)\frac{\bar{w}}{\bar{q}} \exp\left(m^*_{1H}\hat{\mu}^*_{Ht+k} - m^*_{2H}\hat{a}_{it+k}\right)\right] c^*_{it,t+k}(z), \qquad (13)$$

where

$$c^*_{it,t+k}(z) = \left[\frac{\bar{p}^*_i(z)}{\bar{p}^*_i}\right]^{-\varepsilon} \exp\left\{-\varepsilon\left[\hat{p}^*_{it}(z) - \sum_{l=1}^{k}\pi^*_{t+l} - \hat{p}^*_{it+k}\right]\right\} c^*_{it+k}. \qquad (14)$$

In the above equations, the vector of state variables is defined as $\hat{\boldsymbol{x}}^*_{Hit} = (\hat{\mu}^*_{Ht}, \hat{\mu}^*_{Ht+1}, ..., \hat{a}^*_{it}, \hat{a}^*_{it+1}, ...)'$, $\hat{\mu}^*_{Ht+k} = \hat{w}_{t+k} - \hat{q}_{t+k} + \sum_{l=1}^{k}\pi^*_{t+l}$, $\pi^*_t = \ln(P^*_t/P^*_{t-1})$, and $p^*_{it} = P^*_{it}/P^*_t$. In the second line of (13), the cost of supplying a unit of the good to a Canadian consumer is higher due to the iceberg trade cost τ. The aggregate RER in $\hat{\mu}^*_{Ht+k}$ converts this cost into units of real aggre-

[9] In the attention-augmented objective function, we do not explicitly introduce m_{1H} as a coefficient on $\sum_{l=1}^{k}\pi_{t+l}$ and $\hat{p}_{it+k}$ in (11). This is because we examine the log-linearized first-order condition for reset prices. When taking the log-linearization, the presence of the degree of attention in (11) does not affect the first-order terms.

[10] To see this, rewrite the expression in terms of nominal variables: $(1+\tau)w_t/(q_t a_{it}) = (1+\tau)W_t/(S_t P^*_t a_{it})$, where W_t/S_t is the nominal wage denominated Canadian dollars.

[11] The derivation of the objective function that a fully attentive firm maximizes is provided in Appendix A.1.

gate Canadian consumption, allowing it to be compared with the real price $\hat{p}^*_{it}(z)$.[12] When discounting the US firm's real profits in each period, q_{t+k} in the first line of (12) converts these profits into US goods.

2.3.2 The optimal reset prices under behavioral inattention

We now consider the maximization problem when firms are not fully attentive to the state variables that enter their objective function. This problem is referred to as the "sparse max" because Gabaix (2014) originally developed a model in which economic agents respond to only a limited number of variables out of numerous variables.

Consider again an inattentive US firm when selling its brand in the domestic market. The firm maximizes its attention-augmented objective function (12) by choosing $\hat{p}_{it}(z)$:

$$\hat{p}_{Hi}\left(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H\right) = \arg\max_{\hat{p}_{it}(z)} v_{Hi}\left(\hat{p}_{it}(z), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H\right), \tag{15}$$

given $\boldsymbol{m}_H$. We focus on the first-order approximation of $\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)$. By taking the first-order condition with respect to $\hat{p}_{it}(z)$ from (12) and log-linearizing the condition around the steady state, we derive the optimal reset price under behavioral inattention:

$$\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) \;\; = \;\; (1-\lambda\delta)\mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k (m_{1H}\hat{\mu}_{Ht+k} - m_{2H}\hat{a}_{it+k}), \tag{16}$$

which reflects the forward-looking properties in the Calvo pricing. The parameter λ determines the extent to which the firm weights expected marginal costs in its pricing decision. Given assumptions regarding the stochastic processes of money supply and labor productivity, this equation simplifies to:

$$\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) = m_{1H}\hat{w}_t - m_{2H}(1-\lambda\delta)\hat{a}_{it}. \tag{17}$$

The detailed derivation of this expression is provided in Appendix A.2. The appendix also presents the optimal reset price for inattentive US firms operating in the Canadian market, $\hat{p}^*_{Hi}(\hat{\boldsymbol{x}}^*_{Hit}, \boldsymbol{m}^*_H)$ as well as the optimal reset prices for inattentive Canadian firms in both the Canadian and US markets, $\hat{p}^*_{Fi}(\hat{\boldsymbol{x}}^*_{Fit}, \boldsymbol{m}^*_F)$ and $\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F)$, respectively.

In the sparse max problem of Gabaix (2014), agents endogenously determine their degree

[12] For the aggregate component of Canadian firms' real marginal costs, the definitions are $\hat{\mu}^*_{Ft+k} = \hat{w}^*_{t+k} + \sum_{l=1}^{k} \pi^*_{t+l}$ for selling their brands in Canadian markets and $\hat{\mu}_{Ft+k} = \hat{w}^*_{t+k} + \hat{q}_{t+k} + \sum_{l=1}^{k} \pi_{t+l}$ for selling their brands in US markets, respectively.

of attention. More attentiveness increases expected profits, a benefit, but it also incurs a cost. For example, US firms selling their brands in US markets, face the following quadratic cost function in the vector $\boldsymbol{m}_H$,

$$\mathcal{C}\left(\boldsymbol{m}_H\right) = \frac{\kappa_1}{2} m_{1H}^2 + \frac{\kappa_2}{2} m_{2H}^2, \tag{18}$$

where $\kappa_j \geq 0$ for $j = 1, 2$. Given the cost function, US firms determine the optimal degrees of attention by solving,

$$\max_{\boldsymbol{m}_H \in [0,1]^2} \mathbb{E}\left\{ v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right) - \mathcal{C}\left(\boldsymbol{m}_H\right) \right\}, \tag{19}$$

where $\mathbb{E}\left(\cdot\right)$ represents the unconditional expectations. Here, for simplicity, we assume that US firms choose the degrees of attention separately based on the market in which they sell their brands.[13] In (19), we evaluate $v_{Hi}\left(\cdot\right)$ at $\hat{p}_{it}(z) = \hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)$ in the first argument and at $\boldsymbol{m}_H = \boldsymbol{\iota}$ in the third argument. That is, the profit function is the true function under $\boldsymbol{m}_H = \boldsymbol{\iota}$ in the third argument, but the true profit function is evaluated at the inattentive firm's action because $\boldsymbol{m}_H$ in $\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)$ is not equal to $\boldsymbol{\iota}$ in general. The fact that the true profit function is evaluated at the inattentive firm's action implies that more attentiveness in pricing increases expected profits. However, due to the cost function $\mathcal{C}\left(\boldsymbol{m}_H\right)$, firms face a trade-off in choosing the degree of attention.

In this problem, For example,

We define the sparse max problem for US firms selling their brands in US markets as follows. The firms' decision-making process consists of two steps. In the first step, firms determine the degree of attention $\boldsymbol{m}_H$ by solving the simplified problem derived from (19):

$$\boldsymbol{m}_H = \arg \min_{\boldsymbol{m}_H \in [0,1]^2} \frac{1}{2}(\boldsymbol{\iota} - \boldsymbol{m}_H)' \boldsymbol{\Lambda}_H (\boldsymbol{\iota} - \boldsymbol{m}_H) + \frac{1}{2} \boldsymbol{m}_H' \boldsymbol{\kappa} \boldsymbol{m}_H, \tag{20}$$

where

$$\boldsymbol{\Lambda}_H = \begin{bmatrix} \Lambda_{1H} & 0 \\ 0 & \Lambda_{2H} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\kappa} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix}.$$

For further details, see Appendix A.3. The first term on the right-hand side of (20) is derived from the quadratic approximation of $\mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right) - \mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right)$,

[13] We could instead assume that US firms choose the same degree of attention to their real marginal costs regardless of the market. However, we can show that this change in assumption does not substantially affect our main results.

which is the expected profit loss of deviating attention ($\boldsymbol{m}_H$) from full attention in pricing. The second term is $\mathcal{C}(\boldsymbol{m}_H)$. Under the model's assumptions, the diagonal elements of $\boldsymbol{\Lambda}_H$ are given by

$$\Lambda_{1H} = -\left\{\frac{\partial^2 v_{Hi}\left(\hat{p}_{Hi}\left(\mathbf{0},\boldsymbol{\iota}\right),\mathbf{0},\boldsymbol{\iota}\right)}{\partial \hat{p}_{it}(z)^2}\right\}\mathrm{Var}\left(\hat{\mu}_{Ht}\right), \tag{21}$$

$$\Lambda_{2H} = -\left\{\frac{\partial^2 v_{Hi}\left(\hat{p}_{Hi}\left(\mathbf{0},\boldsymbol{\iota}\right),\mathbf{0},\boldsymbol{\iota}\right)}{\partial \hat{p}_{it}(z)^2}\right\}(1-\lambda\delta)^2\mathrm{Var}\left(\hat{a}_{it}\right), \tag{22}$$

respectively.[14] In the second step, firms determine the optimal reset price based on the solution obtained in the first step.

The solution for the degree of attention in the first step is given by $m_{jH} = \Lambda_{jH}/(\Lambda_{jH}+\kappa_j)$ for $j = 1, 2$. In this sparse max, we focus on the case of finite Λ_{jH} to ensure that $m_{jH} < 1$ as long as $\kappa_j > 0$. In the special case of $\kappa_j = 0$ (i.e., when paying attention incurs no cost), $m_{jH} = \Lambda_{jH}/(\Lambda_{jH} + \kappa_j) = 1$ is selected. In addition, we exclude the case of $m_{jH} = 0$ and focus on the range $m_{jH} \in (0, 1]$.[15] As we discuss later, these assumptions are useful for our objective of explaining the PPP puzzle.

There are three remarks on the sparse max of our model. First, the choice variable in the attention-augmented objective functions (12) and (13) is the real price, not the nominal price. In our model, inattention to real marginal costs, rather than to nominal marginal costs, is the key assumption for resolving the PPP puzzles. Gabaix (2014) notes that "a sparse agent will make different predictions in different frames" (p. 1692). He further provides theoretical examples of both a "nominal" frame, in which agents pay attention to nominal variables, and a "real" frame, in which they pay attention to real variables. As he claims, the choice of framing is important for the implications of the model of behavioral inattention. Since real marginal costs have finite variance in our model, firms will become inattentive to real marginal costs in the real frame. The endogenously chosen degree of attention to the aggregate component of real marginal costs, derived as $m_{1H} = \Lambda_{1H}/(\Lambda_{1H} + \kappa_1)$, is less than one. In contrast, the aggregate component of nominal marginal costs is nonstationary under our assumptions. When the variance of the aggregate component of nominal marginal costs diverges to infinity, firms in the nominal frame would be fully attentive because $m_{1H} = \Lambda_{1H}/(\Lambda_{1H} + \kappa_1)$ with $\Lambda_{1H} \to \infty$ implies $m_{1H} \to 1$.

[14] Appendix A.3 also discusses the remaining sparse max problem for US firms selling abroad and Canadian firms operating in both the Canadian and US markets.

[15] Gabaix (2014) shows that under a quadratic cost function, the selected degree of attention is zero if $\Lambda_{jH} = 0$ (e.g., when there is no uncertainty in the variables economic agents pay attention to). He also examines the properties of the selected degree of attention under alternative specifications of the cost function.

Second, the degree of attention is assumed to be common across the macroeconomic variables that constitute real marginal costs. This assumption is made because different representations of real marginal costs make it difficult to interpret behavioral inattention. For example, the aggregate component of real marginal costs for exporting US-produced brands in period t is given by $\hat{\mu}^*_{Ht} = \hat{w}_t - \hat{q}_t$, where the aggregate RER is highly volatile. However, using the equilibrium condition, this expression can also be written as $\hat{\mu}^*_{Ht} = \hat{w}^*_t$. If we allow exporting firms to use variable-specific attention to $\hat{w}_t$, $\hat{q}_t$, and $\hat{w}^*_t$, attention to $\hat{\mu}^*_{Ht}$ as a whole could vary depending on which expression we use for $\hat{\mu}^*_{Ht}$. To avoid such complexity, we assume a common degree of attention in our model.

Third, there are alternative ways to specify the degree of attention. Gabaix (2020) introduces the concept of "cognitive discounting," in which the degree of attention to economic variables k periods ahead weakens as k increases. For instance, firms' real marginal cost is given by $\hat{\mu}_{Ht+k} - \hat{a}_{it+k}$, and the degree of attention may decline exponentially with k, while agents pay full attention to current variables. In contrast, our model assumes that inattention is uniform across time and may apply even to current variables. Nonetheless, within the present model setup with stochastic shocks, we can show that incorporating cognitive discounting does not affect our results.

2.4 Equilibrium

To complete the description of the model, we specify the transfers and the labor market clearing conditions. The profits of US (Canadian) firms accrue exclusively to US (Canadian) households. In other words, $\Pi_t = \int_i \int_{z=0}^{1/2} \Pi_{it}(z) dz di$ and $\Pi^*_t = \int_i \int_{z=1/2}^{1} \Pi^*_{it}(z) dz di$, where $\Pi_{it}(z)$ and $\Pi^*_{it}(z)$ are the total nominal profits of firms producing brand z. Monetary injections are assumed to equal nominal transfers from the government to domestic residents: $T_t = M_t - M_{t-1}$ for the US, and $T^*_t = M^*_t - M^*_{t-1}$ for Canada. The labor market-clearing conditions are $n_t = \int_i \int_{z=0}^{\frac{1}{2}} n_{it}(z) dz di$ and $n^*_t = \int_i \int_{z=1/2}^{1} n^*_{it}(z) dz di$.

An *equilibrium* of the economy is a collection of allocations and prices such that (i) households' allocations are solutions to their maximization problem (namely, $\{c_{it}(z)\}_{i,z}$, n_t, M_t, B_{t+1}, for US households and $\{c^*_{it}(z)\}_{i,z}$, n^*_t, M^*_t, B^*_{t+1}, for Canadian households); (ii) prices and allocations of firms are solutions to their sparse max for $v_{it}(z)$ and $v^*_{it}(z)$ where $z \in [0, 1]$ (namely, $\{P_{it}(z), P^*_{it}(z), n_{it}(z), y_{it}(z)\}_{i,z\in[0,1/2]}$ for US firms and $\{P_{it}(z), P^*_{it}(z), n^*_{it}(z), y^*_{it}(z)\}_{i,z\in(1/2,1]}$ for Canadian firms); (iii) all markets clear; (iv) the productivity, money supply, and transfers satisfy the specifications discussed earlier.

3 The equilibrium good-level RER

This section derives the model's implications for the stochastic process governing the good-level RER, which forms the basic building block of the aggregate RER. The section also discusses the relationship between our model and more restrictive antecedents.

3.1 Derivation of the fundamental dynamic equation governing the good-level RER

To begin, define the good-level RER as $q_{it} = S_t P^*_{it}/P_{it}$. Using the definition of the aggregate RER, $q_t = S_t P^*_t / P_t$, the good-level RER can also be written as $q_{it} = q_t p^*_{it}/p_{it}$. Therefore, we express $\hat{q}_{it}$ as

$$\hat{q}_{it} = \hat{q}_t + \hat{p}^*_{it} - \hat{p}_{it}, \tag{23}$$

where $\hat{p}_{it}$, the price index for good i sold in US markets, is given by

$$\hat{p}_{it} = \lambda(\hat{p}_{it-1} - \pi_t) + (1-\lambda)\hat{p}^{opt}_{it}, \tag{24}$$

and $\hat{p}^*_{it}$ is formulated analogously. Here, $\hat{p}^{opt}_{it}$ represents the weighted average of the optimal reset prices under behavioral inattention. This reset price index is

$$\hat{p}^{opt}_{it} = \omega \hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) + (1-\omega)\,\hat{p}_{Fi}\left(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F\right), \tag{25}$$

where $\omega = (1/2)(\bar{p}_{Hi}/\bar{p}_i)^{1-\varepsilon} = (1+(1+\tau)^{1-\varepsilon})^{-1} \in (1/2, 1]$ represents the degree of home bias in expenditure shares arising from trade costs. The home bias is strictly greater than 1/2 in the presence of iceberg trade costs ($\tau > 0$). Similarly, we can derive $\hat{p}^{opt*}_{it}$, the reset price index in Canada.

Moreover, it is useful to define the weighted average degree of attention given to the aggregate component of real marginal costs in the US and Canada, $m \in (0,1]$, $m = \omega m_{1H} + (1-\omega)m_{1F}$, and the sensitivity of the good-level RER to the idiosyncratic component of real marginal costs as, $\psi = \omega m_{2H} - (1-\omega)m_{2F} \in [-1/2, 1]$, and $\omega = \left(1+(1+\tau)^{1-\varepsilon}\right)^{-1} \in (1/2, 1]$. Finally, for notational convenience, let the (relative) real and nominal shocks be defined as, $\varepsilon^n_t = \varepsilon^M_t - \varepsilon^{M^*}_t \sim i.i.d.\ (0, \sigma^2_n)$ and $\varepsilon^r_{it} = \varepsilon^a_{it} - \varepsilon^{a^*}_{it} \sim i.i.d.\ (0, \sigma^2_r)$, respectively.

Proposition 1 *Under the preferences given by $U(c,n) = \ln c - \chi n$, the CIA constraints, the stochastic processes for money supply* (4) *and* (5)*, the stochastic processes for labor productivity* (6) *and* (7)*, and Calvo pricing with the degree of price stickiness $\lambda \in (0,1)$, the stochastic* *process for the good-level RER is given by*

$$\ln q_{it} = \lambda \ln q_{it-1} + (1-m)(1-\lambda)\ln q_t + \lambda \varepsilon_t^n + (1-\lambda)(1-\lambda\delta)\psi\varepsilon_{it}^r, \tag{26}$$

Proof. See Appendix A.4. ■

3.2 Theoretical discussion

In this sub-section we review models which are nested by our specification in the sense that restrictions on (26) recover these alternative specifications. All models in this class feature Calvo time-dependent price stickiness as their formulation of nominal rigidities.

Kehoe and Midrigan (2007) is an early and important contribution. In their model, the good-level real exchange rate is driven exclusively by nominal shocks, ε_t^n:

$$\ln q_{it} = \lambda \ln q_{it-1} + \lambda \varepsilon_t^n. \tag{27}$$

This equation is a special case of (26) with full attention and no real shocks as represented by the restrictions, $m = 1$ and $\varepsilon_{it}^r = 0$ for all t in our model.

To gain intuition behind (27), recall that $\ln q_{it} = \ln S_t + \ln P_{it}^* - \ln P_{it}$. Suppose that the money supply unexpectedly increases in the US. While this unexpected increase in domestic money supply keeps P_{it}^* constant, it raises S_t and P_{it}. Note that the NER is free to adjust, whereas the adjustment of P_{it} is slow due to price stickiness. As a result, the increase in P_{it} only partially offsets the increase in S_t. The extent of the offset depends on λ. If prices are perfectly flexible such that $\lambda = 0$, the contemporaneous change in P_{it} perfectly offsets the increase in S_t, meaning that the nominal shock is irrelevant for the RER as in the classical dichotomy. At the opposite extreme, if $\lambda = 1$, P_{it} remains unchanged indefinitely, and the good-level RER tracks the NER, and follows a random walk.

If we relax the assumption of $m = 1$ (i.e., if firms are not fully attentive to the aggregate component of real marginal costs), the good-level RER depends on the aggregate RER:

$$\ln q_{it} = \lambda \ln q_{it-1} + (1-m)(1-\lambda)\ln q_t + \lambda \varepsilon_t^n. \tag{28}$$

We refer to this dependence as *aggregate RER dependence*. When $0 < m < 1$, the aggregate RER additionally appears in (28). However, when $m = 1$, the aggregate RER disappears from (28).[16]

[16]Recall that m is the mean degree of attention to the aggregate component of real marginal costs, the

The intuition behind aggregate RER dependence in (28) can be understood in two steps. First, the aggregate prices serve as the default value in firms' pricing decisions. In the behavioral economics literature, the default value is defined as the value on which economic agents rely in their decisions when paying no attention to the state of the economy. In other words, it is "the value that spontaneously comes to mind with no thinking" (Gabaix, 2019, p. 268). Here, firms choose the prior mean of prices over differentiated goods as the default value. To illustrate this, consider the example of US firms selling their brands in US markets. Noting that $\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)$ represents the log deviation of the real price P_{Hit}/P_t from the steady state, we can rewrite (17) as,

$$\ln P_{Hit} = (1 - m_{1H}) \ln P_t + m_{1H} \ln W_t, \tag{29}$$

where we suppress the constant term arising from constant markups and maintain the assumption of no real shocks. If $m_{1H} = 0$, $\ln P_{Hit}$ is fixed at its default value $\ln P_t$ (i.e., the prior mean of all prices in the country). If $0 < m_{1H} < 1$, firms start from the default value and make a partial adjustment of P_{Hit} to W_t. As argued by Gabaix (2019), this is akin to the psychology of "anchoring and adjustment," a concept introduced by Tversky and Kahneman (1974). At the opposite extreme, if $m_{1H} = 1$, the firm's nominal reset price is independent of the default value.

Second, the dependence of good-level prices on aggregate prices leads to the dependence of the good-level RER on the aggregate RER. Appendix A.5 shows that the nominal reset prices for good i are given by:

$$\ln P_{it}^{opt} = (1 - m) \ln P_t + m \ln W_t, \tag{30}$$

$$\ln P_{it}^{opt*} = (1 - m) \ln P_t^* + m \ln W_t^*. \tag{31}$$

Equations (30) and (31) result from the weighted average of the prices of domestic and exported goods (i.e., the weighted average of $\ln P_{Hit}$ and $\ln P_{Fit}$ for (30) and that of $\ln P_{Fit}^*$ and $\ln P_{Hit}^*$ for (31)). However, they can also be interpreted as the weighted average of aggregate prices and nominal wages. Denoting q_{it}^{opt} as the real reset exchange rate for good i, where $q_{it}^{opt} = S_t P_{it}^{opt*}/P_{it}^{opt}$, Appendix A.5 also shows that

$$\ln q_{it}^{opt} = (1 - m) \ln q_t. \tag{32}$$

average of m_{1H} and m_{1F}. Because $1/2 < \omega < 1$ holds for $\tau \in [0, \infty)$, $m = 1$ occurs only if all US and Canadian firms are fully attentive to the aggregate component of their marginal costs.

In (32), when firms are less attentive to the aggregate components of real marginal costs, nominal reset prices become more anchored to the default values given by the aggregate price levels. As the dependence of nominal reset prices on aggregate prices increases, the dependence of the real reset exchange rate on the aggregate RER also strengthens.[17]

As a result, behavioral inattention serves as a source of strategic complementarity in models of pricing. Equations (30) and (31) will also resonate with readers familiar with dynamic pricing in the presence of real rigidities, as discussed by Ball and Romer (1990), or strategic complementarities, as analyzed by Woodford (2003).

Behavioral inattention to the aggregate component of real marginal costs affects the dynamic equation for the good-level RER in significant and subtle ways. In (27), a one-unit increase in ε_t^n raises the good-level RER by λ. However, in (28), it increases the good-level RER by $\lambda \times \left(1 + \frac{(1-m)(1-\lambda)}{1-(1-m)(1-\lambda)}\right)$. Since the good-level RER is also the basic building block of the aggregate RER, it is instructive to aggregate $\ln q_{it}$ over i:[18]

$$\ln q_t = \frac{\lambda}{1-(1-m)(1-\lambda)} \ln q_{t-1} + \frac{\lambda}{1-(1-m)(1-\lambda)} \varepsilon_t^n, \tag{33}$$

which means that a one-unit increase in ε_t^n raises the aggregate RER by $\lambda/[1-(1-m)(1-\lambda)]$. Through strategic complementarities, the increase in the aggregate RER further amplifies the good-level RER, depending on the degree of attention m.

Crucini et al. (2010b, 2013) extended the Kehoe and Midrigan (2007) model to incorporate idiosyncratic real productivity shocks. Their model with full attention implies:

$$\ln q_{it} = \lambda \ln q_{it-1} + \lambda \varepsilon_t^n + (1-\lambda)(1-\lambda\delta)\psi\varepsilon_{it}^r. \tag{34}$$

This equation is a special case of (26) with $m = 1$. Furthermore, $\psi = \omega m_{2H} - (1-\omega)m_{2F} = 2\omega - 1$ holds because $m_{2H} = m_{2F} = 1$. A strictly positive trade cost (i.e., $\tau > 0$) leads to home bias ($\omega > 1/2$) in the price indexes.[19] Since the trade cost ensures that $\psi = 2\omega - 1$ is strictly positive, this friction allows the real shock ε_{it}^r to affect the good-level RER. To understand

[17] Blanco and Cravino (2020) examine RERs using only newly reset prices and find that fluctuations in the aggregate real reset exchange rate are strongly correlated with the aggregate RER. When $m < 1$, (32) is consistent with their empirical finding.

[18] To derive (33), we integrate (26) across good i. In aggregation, $\int_{i=0}^{1} \ln q_{it} di = \ln q_t$ holds from the definition of the good-level RER. From the definition of q_{it}, $\ln q_{it} = \ln q_t + \ln p_{it}^* - \ln p_{it}$. For US real prices, the integral of the real prices over i is zero because $\int_{i=0}^{1} \ln p_{it} di = \int_{i=0}^{1} \ln P_{it} di - \ln P_t = 0$. The same result holds for the Canadian real price so that $\int_{i=0}^{1} \ln p_{it}^* di = 0$. These results lead to $\int_{i=0}^{1} \ln q_{it} di = \ln q_t$. The resulting equation is $\ln q_t = \lambda \ln q_{t-1} + (1-\lambda)(1-m)\ln q_t + \lambda\varepsilon_t^n$. Simplifying the above equation, we obtain (33).

[19] Home bias is reflected in the weighted average of the optimal reset prices. See (25).

the role of real shocks, again recall that $\ln q_{it} = \ln S_t + \ln P^*_{it} - \ln P_{it}$. Positive productivity shocks in US firms producing good i reduce both P^*_{it} and P_{it} because these firms sell their goods in both countries. However, the expenditure home bias generated by trade costs means that productivity gains in the US generate larger reductions in the cost of expenditure for US consumers than their Canadian counterparts, decreasing P_{it} more than P^*_{it}, leading to an increase in q_{it}.

Behavioral inattention to the idiosyncratic component of real marginal costs does not significantly affect the structure of the dynamic equation. In particular, even if m_{2H} and m_{2F} are less than one, no new term appears on the right-hand side of (26). Moreover, (33) continues to hold regardless of the degree of attention to the idiosyncratic component. In the process of aggregating the good-level RERs, all idiosyncratic real shocks are washed out in the integral over i because $\int_{i=0}^{1} \varepsilon^r_{it} di = 0$. In what follows, we focus mainly on a firm's attention to the aggregate component of real marginal costs.

The emphasis on attention to the aggregate component can be justified for at least three reasons. First, idiosyncratic shocks tend to exhibit much larger variations than aggregate shocks. Using the example of US firms selling their brands in US markets, this relationship implies that m_{2H} is much closer to one than m_{1H}.[20] Indeed, $m_{jH} = \Lambda_{jH}/(\Lambda_{jH} + \kappa_j)$, and Λ_{jH} increases with the volatility of shocks (see (21) and (22)). Second, firms may incur lower costs in paying attention to their idiosyncratic productivity than to macroeconomic variables because idiosyncratic productivity is firm-specific internal information. In this case, κ_2 (for idiosyncratic variables) may be much lower than κ_1 (for aggregate variables), again implying that m_{2H} is much closer to one than m_{1H}.[21] Third, focusing on firms' attention to the aggregate component is convenient for examining the validity of our model. In the empirical analysis, we will test the null hypothesis that the good-level RER is independent of the aggregate RER based on (26). Within the context of our model, aggregate RER independence corresponds to full attention to the aggregate component of real marginal costs. However, under our conjecture that firms' attention to the idiosyncratic component of real marginal costs is closer to full attention than to the aggregate component, full attention to the aggregate component automatically implies full attention to the idiosyncratic component. In other words, acceptance of the null hypothesis can be interpreted as evidence of full attention to all variables.

[20] In their rational inattention model, Maćkowiak and Wiederholt (2009) emphasize this result due to difference in variations between aggregate and idiosyncratic shocks. In Section 6.3, we will discuss similarities between the two models of inattention.

[21] The solution for m_{jH} implies that the condition for $m_{2H} \geq m_{1H}$ is $\kappa_2/\Lambda_{2H} \leq \kappa_1/\Lambda_{1H}$. Therefore, if we allow for $\kappa_2 \leq \kappa_1$ in addition to $\Lambda_{2H} \geq \Lambda_{1H}$, then $m_{2H} \geq m_{1H}$ is more likely to hold.

To summarize, behavioral inattention to the aggregate component of real marginal costs generates a new term that affects the good-level RER, namely, the aggregate RER. This new term arises because inattentive firms refer to aggregate prices as default values in their pricing. As such, behavioral inattention of firms serves as a source of strategic complementarity. The model with Calvo pricing for the good-level and aggregate RERs has been theoretically developed and empirically assessed by Kehoe and Midrigan (2007), Crucini et al. (2010b, 2013), and many others. However, these previous studies do not focus on aggregate RER dependence.

4 Empirical Results

This section develops a formal test of aggregate RER independence and provides strong evidence against the null hypothesis of aggregate RER independence. In addition, we estimate the degree of attention using good-level RER data.

4.1 A test for aggregate RER independence

We derive a panel regression model to test for aggregate RER independence. Define $\ln \tilde{q}_{it} = \ln q_{it} - \lambda \ln q_{it-1} - \lambda \Delta \ln S_t = \ln \left[q_{it} / \left(q_{it-1} S_t / S_{t-1} \right)^{\lambda} \right]$ and $\ln \tilde{q}_t = (1-\lambda) \ln q_t = \ln(q_t / q_t^{\lambda})$. In addition, we can replace the nominal shock ε_t^n with $\Delta \ln S_t$ because the NER follows a random walk with an increment ε_t^n. Consequently, we can rewrite (26) as

$$\ln \tilde{q}_{it} = (1-m) \ln \tilde{q}_t + (1-\lambda)(1-\lambda\delta)\psi \varepsilon_{it}^r. \tag{35}$$

Our panel regression is given by

$$\ln \tilde{q}_{it} = \alpha + \beta \ln \tilde{q}_t + \gamma' X_{it} + u_{it}, \tag{36}$$

where α, β, and γ are regression coefficients, X_{it} is a vector of control variables, and u_{it} is the error term. To implement the regression, we rely on empirical estimates of λ to construct $\ln \tilde{q}_{it}$ and $\ln \tilde{q}_t$. The error term $u_{it} = (1-\lambda)(1-\lambda\delta)\psi\varepsilon_{it}^r$ arises from an i.i.d. real shock and is uncorrelated with the regressor $\ln \tilde{q}_t = (1-\lambda) \ln q_t$ because ε_{it}^r does not appear in (33). Therefore, we estimate (36) using ordinary least squares (OLS). The control variables X_{it} include time-invariant fixed effects and other time-varying components.

A test of the null hypothesis of $\beta = 0$ serves as a test for aggregate RER independence. In our model of behavioral inattention, $m = 1$ implies $\beta = 0$, and rejecting the null hypothesis suggests a rejection of full attention. In addition, the degree of attention, m, can be estimated from $\hat{m} = 1 - \hat{\beta}$, where $\hat{\beta}$ is the OLS estimator of β.[22]

Generalizations The benchmark regression analysis can be generalized in two important ways.

First, the stochastic process for labor productivity can be generalized to include an aggregate country-specific shock, η_t, in addition to the idiosyncratic shock of the benchmark model, ε^a_{it}. Because country-specific shocks are not washed out by aggregation, regression (36) requires modification. For example, if η_t and η^*_t each follow an AR(1) process with the same AR coefficient ρ_η, a new control variable emerges:

$$\ln \tilde{q}_{it} = (1-m)\ln \tilde{q}_t + \frac{(1-\lambda)(1-\lambda\delta)}{1-\lambda\delta\rho_\eta}\psi\eta^r_t + (1-\lambda)(1-\lambda\delta)\psi\varepsilon^r_{it}, \tag{37}$$

where η^r_t denotes productivity differentials and is given by $\eta^r_t = \eta_t - \eta^*_t$. A similar equation can be derived even if additional lags are introduced into the stochastic process for the country-specific component.

Second, we relax the assumption of a common λ and allow for heterogeneity in price stickiness when testing the null hypothesis of $\beta = 0$. Specifically, we replace λ with λ_i and apply the following transformations: $\ln \tilde{q}_{it} = \ln q_{it} - \lambda_i \ln q_{it-1} - \lambda_i \Delta \ln S_t$ for the dependent variable in the regression, and $\ln \tilde{q}^i_t = (1-\lambda_i)\ln q_t$ for the explanatory variable.

As a methodological remark, we emphasize that our regression is robust to the presence of behavioral inattention to the idiosyncratic component of real marginal costs because the structure of the dynamic equation remains unchanged. While behavioral inattention to the idiosyncratic component affects $\psi(=\omega m_{2H} - (1-\omega)m_{2F})$, the value of ψ is entirely innocuous for testing aggregate RER independence.

4.2 Data

We use retail price data from the *Worldwide Cost of Living Survey* compiled by the Economist Intelligence Unit (EIU), which conducts an extensive annual survey of international retail prices across a wide range of cities. The survey reports prices of individual goods in local currency terms, collected by a single agency in a consistent manner over time. The coverage of

[22]Note that because the NER is the common driving force of the good-level and aggregate RERs ($\ln q_{it}$ and $\ln q_t$), the two variables are expected to be highly correlated. In our regression, however, both the aggregate and good-level RERs are modified so that the two variables ($\ln \tilde{q}_{it}$ and $\ln \tilde{q}_t$) are correlated only when the degree of attention is less than unity.

goods and services is substantial in breadth and thus overlaps with the typical urban consumption basket tabulated by national statistical agencies.[23] Recent studies that use these data include Engel and Rogers (2004), Crucini and Shintani (2008), Bergin et al. (2013), Crucini and Yilmazkuday (2014), Andrade and Zachariadis (2016), Crucini and Landry (2019), and Crucini and Telmer (2020).

4.2.1 Countries of focus

Since our method requires reliable estimates of the frequency of good-level price changes, the cross-country scope of our empirical work is limited by the available data and studies to the US–Canadian city pairs and UK–Euro area city pairs. These combinations represent two of the most integrated trading regions in the world, as well as comparable monetary institutions.

For the US–Canadian city pairs, the data include prices for 274 goods and services across multiple cities from 1990 to 2015. The dataset covers 16 US cities and 4 Canadian cities.[24] This results in 64 unique cross-border city pairs. However, because some US cities have substantial missing values in the early 1990s, the dataset constitutes an unbalanced panel.[25] Despite these gaps, the total number of observations available for our regressions exceeds 350,000. For the UK–Euro area city pairs, the dataset includes two UK cities and 18 cities from the Euro area.[26] The dataset covers 301 goods and services from 1990 to 2015. As in the case of US–Canadian city pairs, the panel is unbalanced. Nevertheless, the number of observations from the 36 UK–Euro area city pairs exceeds 200,000.

We compute the bilateral good-level RERs q_{ijt} for each year ($t = 1990, ..., 2015$), each good ($i = 1, 2, ...$), and each international city pair ($j = 1, 2, ...$). The prices used to construct the good-level RERs are the prices in a city expressed in the local currency unit. We use the spot NERs from the EIU data to convert prices to common currency units. The EIU records the NER vis-à-vis the US dollar at the end of the week of the price survey. Thus, the NER

[23]See Rogers (2007) for a detailed comparison between the EIU data and CPI data from national statistical agencies.

[24]The US cities are Atlanta, Boston, Chicago, Cleveland, Detroit, Honolulu, Houston, Lexington, Los Angeles, Miami, Minneapolis, New York, Pittsburgh, San Francisco, Seattle, and Washington DC. The Canadian cities are Calgary, Montreal, Toronto, and Vancouver.

[25]In particular, the survey data in 1990 and 1991 do not include price data for Honolulu. Additionally, Lexington and Minneapolis were only included in the city list starting in 1998.

[26]The UK cities are London and Manchester. The Euro area cities are Amsterdam, Barcelona, Berlin, Brussels, Dublin, Dusseldorf, Frankfurt, Hamburg, Helsinki, Lisbon, Luxembourg, Lyon, Madrid, Milan, Munich, Paris, Rome, and Vienna. We exclude Athens due to inflation often exceeding 10 percent in the 1990s prior to Euro adoption, which were substantially higher than in other Euro area countries. Likewise, we exclude Bratislava because the Slovak koruna appreciated significantly against the UK pound before the Euro was adopted in 2009.

may not necessarily be common across cities in the same country if the timing of the price survey differs across cities. We confirm that the timings of the price survey in Calgary differ from those in the remaining Canadian cities from 2003 to 2014.[27] The NERs of cities within other countries are common in the EIU data. For our regressions and empirical tests that follow, we augment q_{ijt} with the aggregate RER computed from the official CPIs, which the EIU also reports.

Figure 1 plots two kernel density estimates of the bilateral good-level RERs pooling all goods and services: one for the first year of the sample (1990) and the other for the last year (2015). The upper panel of the figure shows the distribution of the good-level RERs for the US–Canadian city pairs, while the lower panel displays the distribution for the UK–Euro area city pairs.

When we allow for a country-specific shock in labor productivity, it is necessary to control for the difference in the country-specific components of labor productivity, $\eta_t^r = \eta_t - \eta_t^*$, as it appears in (37). As a proxy for η_t^r, we use the difference in real GDP per hour worked between the two countries, obtained from *OECD.Stat*.

4.2.2 Frequencies of price adjustment

We calibrate λ and construct $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t$ in (36). To compute the value of λ, we transform the monthly frequencies of price changes reported in previous studies into the *in*frequencies of price changes at an annual rate. Let f denote the monthly frequency of price changes. Nakamura and Steinsson (2008) report that the median frequency of price changes in the US consumer prices is 8.7 percent per month. Given this value, we set $f = 0.087$. Under our assumptions of sticky prices, the probability that a price remains unchanged for 12 months is $(1-f)^{12}$. Hence, $\lambda = (1-f)^{12} = (1-0.087)^{12} = 0.34$. We then use $\lambda = 0.34$ to construct $\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda \ln q_{ijt-1} - \lambda \Delta \ln S_t$ and $\ln \tilde{q}_t = (1-\lambda) \ln q_t$ for the US–Canadian city pairs. For the UK–Euro area city pairs, we rely on Gautier et al. (2024), who find that the average frequency of price changes is 8.5 percent for consumer prices across 11 Euro area countries.[28] This implies a value of $\lambda = (1-0.085)^{12} = 0.34$ for the UK–Euro area city pairs as well. For descriptive statistics of $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t$, see Appendix A.6.

When allowing for heterogeneity in the frequency of price changes, we require good-specific frequencies of price changes. For the US–Canadian city pairs, we use the monthly

[27] As we discuss later, we adjust our regressions to account for this difference in timing.

[28] Both Nakamura and Steinsson (2008) and Gautier et al. (2024) exclude the effects of sales when calculating the frequencies of price changes. In addition, Nakamura and Steinsson (2008) also adjust for product substitutions.

frequencies reported in Nakamura and Steinsson (2008), which are based on the Entry Level Item (ELI) classification of the US CPI. We match goods and services in the EIU data to the ELI categories and assign the corresponding monthly frequency of Nakamura and Steinsson (2008) to goods and services in the EIU data. For the UK–Euro area city pairs, we use good-specific monthly frequencies calculated by Gautier et al. (2024). These frequencies are based on the Classification of Individual Consumption by Purpose (COICOP) and are aggregated using the country weights of Euro area consumer prices. We assign the COICOP-level frequencies at the five-digit level to the corresponding goods and services in the EIU data.

Figure 2 shows the distribution of the monthly frequencies of price changes after matching goods and services in the CPI with those in the EIU data. The upper panel presents the histogram and kernel density estimates of the frequencies used for the US–Canadian city pairs, while the lower panel presents those used for the UK–Euro area city pairs. Overall, the distributions are right-skewed, exhibiting substantial heterogeneity. The shape of the distribution of frequencies of price changes for the US–Canadian city pairs closely resembles that reported in Nakamura and Steinsson (2013). The number of available frequencies is 274, with a standard deviation of 13.0 percent, ranging from a minimum of 2.4 percent to a maximum of 88.6 percent. Many goods and services (approximately 23 percent) have a frequency below 5 percent per month. For example, the EIU items "Man's haircut (tips included)" and "Woman's cut & blow dry (tips included)." At the same time, the distribution has a long right tail extending up to approximately 90 percent per month, as seen in the EIU item "Regular unleaded petrol." The distribution of frequencies of price changes for the UK–Euro area city pairs also exhibits considerable heterogeneity. The number of available frequencies is 236, with a standard deviation of 10.2 percent, ranging from a minimum of 1.5 percent to a maximum of 45.0 percent.[29]

We construct the data for $\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda_i \ln q_{ijt-1} - \lambda_i \Delta \ln S_t$ and $\ln \tilde{q}_t^i = (1-\lambda_i) \ln q_t$. In calibrating λ_i, we apply the same formula to the good-specific monthly frequencies of price changes: $\lambda_i = (1 - f_i)^{12}$, where f_i denotes the good-specific monthly frequency of price changes. Appendix A.6 reports the corresponding descriptive statistics of $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t^i$, allowing us to compare these variables with those constructed under a common λ.

[29] Gautier et al. (2024) do not report the frequency of price changes for the item corresponding to the EIU item "Regular unleaded petrol." In the UK–Euro area data, the items with the most flexible prices are vegetables.

4.3 Estimation results

Table 1 provides the estimation results of (36) for testing aggregate RER independence. The left panel shows the results for the US–Canadian city pairs, while the right panel presents those for the UK–Euro area city pairs. The table reports the estimated coefficients on $\ln \tilde{q}_t$ along with their standard errors. By default, we include good-specific fixed effects in the regressions. This is because variations in good-specific fixed effects are substantially larger in LOP deviations than in city-pair-specific fixed effects.[30] For robustness, we additionally allow for the inclusion of city-pair-specific fixed effects and/or control for the country-specific component of labor productivity, η_t^r, as motivated by (37). In regressions for the US–Canadian city pairs, we also control for the difference in the timing of the price survey in Calgary by including dummy variables that take the value of one if a city pair involves Calgary in any year from 2003 to 2014.[31]

Overall, $\hat{\beta}$ is approximately 0.85. The standard error of the coefficient indicates a strong rejection of the null hypothesis $\beta = 0$ in favor of the alternative $\beta > 0$.[32] Interpreted through the lens of our theoretical model, the estimated degree of attention, $\hat{m} = 1 - \hat{\beta}$, is around 0.15, suggesting that firms are not fully attentive to the aggregate components of real marginal costs when making their pricing decisions. A comparison between the left and right panels reveals that the estimated coefficients on $\ln \tilde{q}_t$ for the US–Canadian city pairs are close to those for the UK–Euro area city pairs. Taking specification (1) as an example, the first row of Table 1 shows that the estimated β is 0.84 for the US–Canadian city pairs, whereas it is 0.86 for the UK–Euro area city pairs. In terms of the degrees of attention, $\hat{m} = 0.16$ in the US–Canadian city pairs and $\hat{m} = 0.14$ in the UK–Euro area city pairs (see the bottom of the table). Our results on the test for aggregate RER independence are robust to the inclusion of city-pair-specific fixed effects (see specifications (2) and (4)) and to controlling for the log difference in labor productivity (see specifications (3) and (4)).[33]

[30] Crucini and Telmer (2020) emphasize the importance of good-specific fixed effects using an analysis of variance on the EIU data.

[31] The difference in the timing of the price survey causes the aggregate RER to become city-pair- and year-specific. More precisely, let q_t^k and S_t^k denote the aggregate RER and the NER for a city pair k that includes Calgary in a given year from 2003 to 2014. Here, $\ln q_t^k$ is given by $\ln q_t^k = \ln S_t^k + \ln P_t^* - \ln P_t$, which can be rewritten as $\ln q_t^k = (\ln S_t^k - \ln S_t) + \ln q_t$. Likewise, $\ln q_{ijt}^k = (\ln S_t^k - \ln S_t) + \ln q_{ijt}$, where the variables without the superscript k are variables in the other city pairs. Thus, these dummy variables control for the term $\ln S_t^k - \ln S_t$ that arises due to the difference in survey timing in Calgary.

[32] We report standard errors clustered by goods, but the null hypothesis is also rejected when standard errors are clustered by city pairs or years. Likewise, our main findings are robust even when we replace λ with values reported in previous studies on price dynamics, such as Bils and Klenow (2004) and Klenow and Kryvtsov (2008) for the US–Canadian city pairs, and Álvarez et al. (2006) for the UK–Euro area city pairs.

[33] Although we do not report the results here to conserve space, we also estimate a specification including fixed effects specific to both good i and city pair j, and find that the estimated β remains effectively unchanged.

Table 2 presents the estimation results when we drop the assumption of a common λ across goods and instead allow for heterogeneity in price stickiness.[34] Even when good-specific degrees of price stickiness are considered, the null hypothesis $\beta = 0$ is again significantly rejected. Regarding the estimated degrees of attention, $\hat{m}$ tends to decline when we allow for heterogeneity in price stickiness. For instance, under specification (1), $\hat{m}$ decreases from 0.16 to 0.11 for the US–Canadian city pairs and from 0.14 to 0.13 for the UK–Euro area city pairs.

We confirm that the rejection of the null hypothesis is robust. In Appendix A.7, we allow for a more general constant-relative-risk-aversion (CRRA) form in preferences. In this case, the estimation equation becomes more complex, and hypothesis testing requires the use of an instrumental variables estimator.[35] Appendix A.7 derives the estimation equation and presents the empirical results.

As an alternative robustness check, we also regress $\ln q_{it}$ directly on $\ln q_t$, including $\ln q_{it-1}$ and $\Delta \ln S_t$ as additional regressors. The estimation equation is given by

$$\ln q_{it} = \alpha + \beta \ln q_t + \gamma' X_{it} + u_{it}, \tag{38}$$

where β in (38) corresponds to $(1-m)(1-\lambda)$ in (26). The control variables X_{it} in (38) include $\ln q_{it-1}$ and $\Delta \ln S_t$. Note that $\beta = (1-m)(1-\lambda) = 0$ corresponds to $m = 1$, provided that $\lambda < 1$. Therefore, a test of $\beta = 0$ against $\beta > 0$ in (38) allows us to assess whether the data support full attention. Appendix A.8 provides further details on the empirical results.

Finally, we examine how intercity distance affects the estimation results. Greater distance tends to raise RER volatility, which may influence the estimate of m and hypothesis testing.[36] To address this concern, we divide the sample by distance and test the null $\beta = 0$ within each group. We also exclude city pairs below the 5th or above the 95th distance percentile. Even then, the estimate of m and the test results remain largely unchanged. See Appendix A.9 for details.

[34] See also Crucini et al. (2010a, 2010b, 2013), Hickey and Jacks (2011), and Elberg (2016), who emphasize heterogeneity in price stickiness in research on the LOP.

[35] Under the more general CRRA form, firms form expectations about the entire future path of labor supply from the time of price setting to the infinite future. As a result, the estimation equation includes one-period-ahead good-level and aggregate RERs, which leads to endogeneity due to the correlation between the explanatory variables and the forecast errors in the error term.

[36] See Engel and Rogers (1996) and Crucini et al. (2010b).

5 Explaining the PPP puzzle

We turn to the implications of our finding of aggregate RER dependence for the PPP puzzle.

5.1 Persistence of the aggregate RER

Let ρ_q be the first-order autocorrelation of aggregate RERs. Because the AR coefficient in (33) corresponds to the first-order autocorrelation, let us rewrite (33) as:

$$\ln q_t = \rho_q \ln q_{t-1} + \rho_q \varepsilon_t^n, \tag{39}$$

where $\rho_q = \lambda/[1-(1-m)(1-\lambda)]$. In the following proposition, we now discuss Rogoff's (1996) PPP puzzle.

Proposition 2 *Under the same assumptions as in Proposition 1,*

$$\rho_q \geq \lambda, \tag{40}$$

provided $m \in (0,1]$ *and* $\lambda \in (0,1)$. *The equality holds if and only if* $m = 1$.

Proof. It follows from the fact that $(1-m)(1-\lambda) \leq 1$, where (40) holds with the equality if and only if $m = 1$. ■

Proposition 2 suggests that behavioral inattention helps resolve Rogoff's (1996) PPP puzzle. Specifically, the aggregate RER exhibits greater persistence than implied by the degree of price stickiness. Without behavioral inattention (i.e., $m = 1$), ρ_q is equal to λ. When firms are inattentive (i.e., $m < 1$), ρ_q becomes strictly greater than λ. Therefore, even when nominal frictions are small, the model with a small m can account for a highly persistent aggregate RER.

We rule out the possibility that firms are completely inattentive to the aggregate component of real marginal costs (i.e., $m = 0$). In the limiting case where m approaches 0, ρ_q converges to 1, making the aggregate RER identical to the NER, which follows a random walk. However, such equivalence is never observed in the data. The same equivalence arises under fixed prices (i.e., $\lambda = 1$), which is why we exclude $\lambda = 1$ as well as $m = 0$ in Propositions 1 and 2.

We also rule out the case of flexible prices (i.e., $\lambda = 0$) because (39) suggests that $\lambda = 0$ leads to no PPP deviations, even in the short run (i.e., $\ln q_t = 0$ for all t). Thus, our model requires nominal rigidities as an external source of aggregate RER persistence. This feature of our model aligns with the concept of real rigidities in Ball and Romer (1990) and strategic complementarity in Woodford (2003). Using a closed-economy model, Ball and Romer (1990) show that real rigidities alone are insufficient to generate real effects of nominal shocks. They argue that a combination of real rigidities and a small friction in nominal price adjustment is crucial for nominal shocks to have real effects. Similarly, in our model, the combination of behavioral inattention and nominal price adjustment frictions is essential for generating a persistent aggregate RER.

Figure 3 shows how the persistence of the aggregate RER varies with changes in m. The left panel plots ρ_q against $m \in (0,1]$, with λ set to 0.34. For reference, the figure also includes the line representing the lower bound of ρ_q, given by $\lambda = 0.34$. Starting from $\rho_q = \lambda$ when $m = 1$, ρ_q increases monotonically as m decreases. Persistence approaches unity as m approaches zero. The right panel illustrates the ρ_q to λ ratio, defined as:

$$\frac{\rho_q}{\lambda} = \frac{1}{1-(1-m)(1-\lambda)}. \tag{41}$$

This ratio captures the extent to which inattention amplifies the persistence of the aggregate RER that would be explained solely by nominal rigidities under full attention. The figure shows that the ρ_q to λ ratio can be quite large, depending on the value of m.

The estimated degrees of attention suggest that behavioral inattention makes PPP deviations more than twice as persistent as predicted by price stickiness alone. In the left panel of Figure 3, $\rho_q = 0.34$ when $m = 1$. However, the same panel shows that $\rho_q = 0.76$ when we use $m = 0.16$ from specification (1) in Table 1 as the calibrated value for US–Canadian city pairs. The right panel of the figure also indicates that this calibrated value yields a ρ_q to λ ratio exceeding two. In particular, the ratio is 2.24 when $m = 0.16$. When we instead use $m = 0.14$ from specification (1) in Table 1 for the UK–Euro area city pairs, ρ_q rises to 0.79, and the corresponding ρ_q to λ ratio is 2.31.

Let us evaluate the persistence of the aggregate RER in terms of the half-life. The upper panel of Table 3 compares the predicted half-lives of the aggregate RER with those observed in our data, computed using the standard formula $-\ln(2)/\ln\rho_q$. Using the aggregate RER employed in our regressions, we estimate half-lives from an AR(1) model for $\ln q_t$. As shown in the third column (headed "Data"), the observed half-life for the US–Canadian city pairs is 4.92 years. The aggregate RER for the UK–Euro area city pairs exhibits a shorter half-life of 2.40 years.[37]

[37]Note that we have multiple aggregate RERs for the UK–Euro area city pairs because CPIs vary across Euro area countries. The half-life of 2.40 years reported in Table 3 represents the mean of the estimated

How much can the estimated degree of attention explain the observed half-life of the aggregate RER? For US–Canadian city pairs, the predicted half-life is 2.62 years when using $m = 0.16$ (see the first row of the upper panel of Table 3). To account for estimation uncertainty, we also report the 95 percent confidence interval based on the standard errors of $\hat{m}$. Using the lower and upper bounds of this confidence interval, the predicted half-life ranges from 1.99 to 4.01 years. Thus, under $m = 0.16$, the predicted half-life is slightly shorter than the observed half-life of 4.92 years for US–Canadian city pairs. However, recall that when allowing for heterogeneity in price stickiness, the estimated m decreases from 0.16 to 0.11. In this case, the predicted half-life extends to 3.70 years, with a range of 2.52 to 7.61 years, which includes the observed half-life of 4.92 years (see the second row of the upper panel).

For the UK–Euro area city pairs, the predicted half-life is 2.81 years, ranging from 1.90 to 6.13 years, when we use $m = 0.14$ (see the third row of the upper panel of Table 3). The predicted half-life exceeds the observed half-life of 2.40 years. However, the 95 percent confidence interval includes the observed value. Therefore, the model successfully accounts for the observed half-life for the UK–Euro area city pairs. Since $\hat{m}$ decreases only slightly from 0.14 to 0.13 when incorporating the good-specific degree of price stickiness, the model continues to account for the observed half-life for the UK–Euro area city pairs.

We emphasize that the model of behavioral inattention outperforms the model with full attention. When $m = 1$, the first-order autocorrelation of the aggregate RER is only 0.34, as $\rho_q = \lambda = 0.34$. This low persistence of the aggregate RER results in a significantly shorter half-life of 0.64 years compared to the case where $m < 1$.

Using the half-lives obtained here, we can calculate the relative contribution of heterogeneity in price stickiness to the total increase in half-lives. Based on the estimate using a common λ ($\hat{m} = 0.16$), the half-life is extended by 1.98 ($= 2.62 - 0.64$) years for the US–Canadian city pairs. By using the estimate that allows for heterogeneity in price stickiness ($\hat{m} = 0.11$), the half-life is further extended by 1.08 ($= 3.70 - 2.62$) years. Therefore, the contribution of incorporating heterogeneity in price stickiness is approximately 35.5 ($\approx 100 \times 1.08/(1.98 + 1.08)$) percent. A similar calculation for UK–Euro area city pairs results in a contribution of approximately 7.9 percent.

Figure 4 reconfirms the improvement in predicted half-lives based on the impulse response function. Here, the impulse response functions represent the response of the aggregate RER to a nominal shock. The sign of the nominal shock is chosen such that the aggregate RER depreciates on impact. For comparison, initial responses are normalized to unity.[38] In the

half-lives for each country pair to which the UK–Euro area city pairs belong.

[38]Note that the impulse response function of the aggregate RER can be derived from the AR(1) process

figure, the point of intersection between the horizontal line at 0.5 and the impulse response function for each case allows us to identify the half-life graphically. Compared to the model with full attention ($m = 1$, represented by the line with asterisks), the point of intersection shifts further to the right in the cases of $m = 0.11$ (an estimate for the US–Canadian city pairs, represented by the line with pluses) and $m = 0.13$ (an estimate for the UK–Euro area city pairs, represented by the line with circles), indicating that the model of behavioral inattention predicts significantly longer half-lives.

5.2 Persistence of the good-level RER

We next turn to the good-level RER. Let ρ_{qi} denote the first-order autocorrelation of the good-level RER. The following proposition characterizes the relationship between the persistence of good-level RERs and that of the aggregate RER, as predicted by the model.

Proposition 3 *Under the same assumptions as in Proposition 1*

$$\rho_q \geq \rho_{qi}, \tag{42}$$

provided $m \in (0, 1]$, $\lambda \in (0, 1)$, *and* $\sigma_r/\sigma_n \in (0, \infty)$. *The equality holds if* $m = 1$ *or* $\psi = 0$.

Proof. See Appendix A.10. ■

Proposition 3 shows that in the standard model under full attention (i.e., $m = 1$), the predicted persistence of the aggregate RER equals that of good-level RERs, which contradicts the data. However, the proposition also shows that the model of behavioral inattention (i.e., $m < 1$) can account for the puzzling fact that the aggregate RER is significantly more persistent than good-level RERs. Notably, we obtain this result without relying on the "aggregation bias" identified by Imbs et al. (2005). They emphasize that heterogeneity in the persistence of good-level RERs creates an upward bias in the persistence of the aggregate RER. Using multisector sticky-price models with heterogeneity in price stickiness, Carvalho and Nechio (2011) successfully account for this upward bias. In contrast, our model deliberately assumes homogeneity in persistence across goods. Nevertheless, it can qualitatively explain the gap in persistence between the aggregate and good-level RERs.

In our model, two parameters, m and ψ, play a crucial role in explaining this gap. Their

(39).

role can be further examined through the ρ_q to ρ_{qi} ratio, defined as:

$$\frac{\rho_q}{\rho_{qi}} = \frac{1}{1-(1-m)(1-\lambda)\frac{A}{1+A}}, \tag{43}$$

where

$$A = \frac{[(1-\lambda)(1-\lambda\delta)]^2}{1-\lambda^2}\left\{\frac{[1-(1-m)(1-\lambda)]^2-\lambda^2}{\lambda^2}\right\}\psi^2\left(\frac{\sigma_r}{\sigma_n}\right)^2 \in [0,\infty) \tag{44}$$

The derivation is provided in Appendix A.10. From these equations, it follows that the ρ_q to ρ_{qi} ratio is strictly greater than one as long as firms are inattentive to the aggregate component of real marginal costs and real shocks are present in (26). In (43), behavioral inattention, namely $m < 1$, is necessary to increase the ρ_q to ρ_{qi} ratio. However, behavioral inattention alone is not sufficient to generate the gap between ρ_q and ρ_{qi}. Recall that, if ψ is nonzero, real shocks matters for the good-level RER in (26). If $\psi^2 > 0$, A is always positive in (44), reducing the denominator of (43) to a value less than one. Intuitively, a nonzero ψ ensures that i.i.d. real shocks in (26) reduce the persistence of good-level RERs, leading to $\rho_q/\rho_{qi} > 1$.

To assess the effect of m and ψ on the gap between ρ_q and ρ_{qi}, we calibrate the parameters in (43) and (44). We set τ to 74 percent and ε to 4, based on Anderson and van Wincoop (2004) and Broda and Weinstein (2006), respectively.[39,40] Using these values, we obtain a home bias parameter ω of 0.84, which is broadly consistent with values reported in the literature.[41] As discussed in Section 3, the degree of attention to the idiosyncratic component is likely close to unity. Thus, we simply assume full attention to the idiosyncratic component of real marginal costs (i.e., $m_{2H} = m_{2F} = 1$), resulting in $\psi = 0.68$. We later consider an alternative value of ψ to evaluate its effect on the persistence of good-level RERs. We set $\sigma_r/\sigma_n = 5$, which Crucini et al. (2013, p. 64) suggest is a plausible value based on sectoral RER data. The household discount factor δ is set to 0.98, and the degree of price stickiness λ is again set to 0.34.

Figure 5 illustrates the extent to which the good-level RER becomes less persistent than

[39] Using US data, Anderson and van Wincoop (2004) estimate transportation costs at 21 percent and border-related trade barriers at 44 percent. Based on these values, they calculate total international trade costs as $0.74 (= 1.21 \times 1.44 - 1)$.

[40] Broda and Weinstein (2006) report that the median elasticities of substitution during 1990 - 2001 are 3.1 at the seven-digit level of the Standard International Trade Classification (SITC) and 2.7 at the five-digit level.

[41] For example, Chari et al. (2002, p. 546) set the home bias parameter at 0.76, while Steinsson (2008, p. 525) sets it at 0.94.

the aggregate RER as a function of m. The left panel plots ρ_q and ρ_{qi} against m with a solid line and a dashed line, respectively. The former is the same as that in Figure 3. Without behavioral inattention (i.e., $m = 1$), $\rho_q = \rho_{qi} = \lambda(= 0.34)$, meaning that the two curves intersect at their lower bound, λ. When m deviates from unity, the curve for ρ_{qi} falls below that of ρ_q, provided that ψ is nonzero. Furthermore, as previously discussed, ρ_q converges to 1 in the limiting case of $m \to 0$. Equation (44) indicates that A converges to zero as $m \to 0$. Since (43) and $A \to 0$ imply $\rho_q/\rho_{qi} \to 1$, ρ_{qi} also converges to 1. The right panel plots the ρ_q to ρ_{qi} ratio against m with a solid line. In the same panel, a dashed-dotted line represents the ρ_q to ρ_{qi} ratio when ψ is set to 0.34, half of its baseline value of 0.68. The panel indicates that both curves are hump shaped, with the ratios reaching a value of one at both ends of $m \in (0, 1]$. However, the dashed-dotted line lies below the solid line because a smaller ψ weakens the effect of real shocks in reducing the persistence of the good-level RER.

The estimated degrees of attention suggest that inattention reduces the persistence of the good-level RER to less than two-thirds of that of the aggregate RER. Suppose that $m = 0.16$, the estimated degree of attention for US–Canadian city pairs. The left panel of Figure 5 shows that ρ_{qi} is 0.49, whereas ρ_q is 0.76. The solid line in the right panel indicates that the ρ_q to ρ_{qi} ratio is 1.55. Equivalently, ρ_{qi} is less than two-thirds of ρ_q (i.e., $0.49/0.76 < 2/3$). Similar results hold for the estimated degree of attention in UK–Euro area city pairs. When $m = 0.14$, $\rho_{qi} = 0.51$ and $\rho_q = 0.79$. Thus, our model predicts that ρ_{qi} is also less than two-thirds of ρ_q ($0.51/0.79 < 2/3$) for UK–Euro area city pairs.

The lower panel of Table 3 presents the predicted half-lives of the good-level RER under the estimated degrees of attention. The second column of the table also provides the ranges of the predicted half-lives, allowing for estimation uncertainty. For comparison, the rightmost column reports the median half-lives of the good-level RERs estimated from our dataset.[42] In the data from 1990 to 2015, we find that the half-life of the median good is 1.61 years for the US–Canadian city pairs and 1.18 years for the UK–Euro area city pairs, both of which are much shorter than the half-lives of the aggregate RER shown in the corresponding column of the upper panel. The estimated half-lives are also consistent with previous studies using EIU data. For example, Crucini and Shintani (2008) find that the half-life of the median good ranges from 1.03 to 1.61 years based on EIU data from 1990 to 2005. Bergin et al. (2013)

[42] We estimate the panel AR(1) model for $\ln q_{ijt}$ for each good i, using the generalized method of moments estimator of Arellano and Bond (1991). The estimated AR(1) coefficient is then transformed into the half-life. Typically, the good-by-good panel consists of more than 1,400 observations for the US–Canadian city pairs and more than 700 observations for the UK–Euro area city pairs. The median half-lives reported in Table 3 are based on half-lives estimated from panels with more than 500 observations for the US–Canadian city pairs and more than 250 observations for the UK–Euro area city pairs.

also use EIU data to construct good-level RERs from 1990 to 2007, and estimate an AR(1) model showing an average half-life of 1.15 years.

How much can the estimated degree of attention explain the observed persistence of good-level RERs? Similar to the case of the aggregate RER, $m = 0.16$ is not low enough to fully account for the observed half-life of the good-level RER in the US–Canadian city pairs. The predicted half-life is 0.98 years, and the 95 percent confidence interval is [0.85, 1.29] in years. Thus, the predicted half-life is slightly shorter than the observed half-life of 1.61 years. However, when allowing for heterogeneity in price stickiness by using $\hat{m} = 0.11$, the model predicts a half-life of 1.22 years. The 95 percent confidence interval of [0.96, 2.11], which includes the observed half-life. For the UK–Euro area city pairs, the predicted half-lives are 1.03 years under $m = 0.14$ and 1.07 years under $m = 0.13$. In both cases, the confidence intervals include the observed half-life of 1.18 years, indicating that the model explains the persistence of good-level RERs reasonably well.

We can again calculate the relative contribution of heterogeneity in price stickiness to the total increase in half-lives. When m decreases from 1.00 to 0.16, behavioral inattention extends the half-life of the good-level RER by 0.34 ($= 0.98 - 0.64$) years for the US–Canadian city pairs. By using the estimate that allows for heterogeneity in price stickiness ($\hat{m} = 0.11$), the half-life of the good-level RER increases by 0.24 ($= 1.22 - 0.98$) years. Thus, the contribution of incorporating heterogeneity in price stickiness is approximately 41.0 ($\approx 100 \times 0.24/(0.34 + 0.24)$) percent. In contrast, the contribution is only 9.4 percent for the UK–Euro area city pairs.

Before closing this section, two remarks are in order. First, Propositions 2 and 3 can be combined to derive the ρ_{qi} to λ ratio, which measures the amplification from λ to ρ_{qi}. In particular, using (41) and (43), we obtain

$$\frac{\rho_{qi}}{\lambda} = \frac{1 - (1-m)(1-\lambda)\frac{A}{1+A}}{1 - (1-m)(1-\lambda)} \geq 1, \tag{45}$$

where equality holds if and only if $m = 1$. In other words, as long as $0 < m < 1$, the persistence of the good-level RER exceeds λ. This theoretical result is also consistent with Kehoe and Midrigan's (2007) empirical finding that the good-level RER is more persistent than what is predicted solely by the degree of price stickiness. Together with the result from (43), our model predicts the inequalities $\rho_q > \rho_{qi} > \lambda$.

Second, our model of behavioral inattention can reproduce the findings of Bergin et al. (2013), who analyze the persistence of the good-level RER *conditional on shocks*. Using a vector error correction model for each good, they find that the good-level RER is as persistent as the aggregate RER, conditional on a macroeconomic shock. We can analyze the good-level RER conditional on a macroeconomic shock by letting $\sigma_r \to 0$. According to (44), $\sigma_r \to 0$ implies $A \to 0$. Therefore, (43) and (45) imply that $\rho_q = \rho_{qi} > \lambda$, which is consistent with the empirical finding of Bergin et al. (2013).

6 Discussion

In this section, we compare the model of behavioral inattention to other potential explanations of the PPP puzzle. We discuss three key related contributions in the literature and also show the robustness of our model to the incorporation of a Taylor rule.

6.1 The model of roundabout production

Aggregate RER dependence arises from strategic complementarity. Naturally, the literature on RERs has extensively incorporated this concept in different structural forms (e.g., Kehoe and Midrigan, 2007, and Burstein and Gopinath, 2014). Basu (1995) develops a model of roundabout production in which firms' pricing is influenced by the aggregate prices due to their impact on the cost of intermediate goods.[43] If a model with strategic complementarity produces a similar dynamic equation for the good-level RER, our regression may face difficulties in identifying our model from alternative ones. We address this concern in this section.

Consider the US firm's production function given by

$$y_{it}(z) = a_{it}[\Gamma_{it}(z)]^r [n_{it}(z)]^{1-r}, \qquad 0 \leq r < 1, \tag{46}$$

where $\Gamma_{it}(z)$ denotes intermediate goods for brand z of good i. The parameter r represents the degree of roundabout production. Since intermediate goods can also be consumed as final goods, their price is the same as that of final goods, P_t. The firm's cost minimization problem implies that the nominal marginal cost is $\tilde{r} P_t^r W_t^{1-r}/a_{it}$, where $\tilde{r} = r^{-r}(1-r)^{-(1-r)}$. Given that the prices of intermediate goods equal the aggregate prices, the resulting real marginal cost is $\tilde{r} w_t^{1-r}/a_{it}$. Based on this real marginal cost, the US firm's real profits from

[43] Firms' pricing can also be influenced by the aggregate prices through alternative frameworks, such as Kimball's (1995) kinked demand curve and Bergin and Feenstra's (2001) translog preferences. See also Burstein and Gopinath (2014) for other specifications of strategic complementarity.

selling brand z of good i in the US market are $[p_{it}(z) - \tilde{r} w_t^{1-r}/a_{it}] c_{it}(z) = \{\bar{p}_{it}(z) \exp[\hat{p}_{it}(z)] - \tilde{r}\bar{w}^{1-r} \exp[(1-r)\hat{w}_t - \hat{a}_{it}]\} c_{it}(z)$.

Appendix A.11 derives the dynamic equation for the good-level RER:

$$\ln q_{it} = \lambda \ln q_{it-1} + r\nu(1-\lambda)\ln q_t + \lambda \varepsilon_t^n + (1-\lambda)(1-\lambda\delta)\psi^{RP}\varepsilon_{it}^r, \quad (47)$$

where ν is a function of model parameters, including r, and $\psi^{RP} = 2\omega - 1$, where a superscript RP denotes roundabout production.[44]

Equation (47) shares a structural similarity with (26). Within the coefficient on the aggregate RER, $1-m$, in (26) is replaced by $r\nu$. Analogous to the effect of a smaller m, a larger r leads to a stronger correlation between the good-level RER and the aggregate RER. Consequently, this results in more persistent RERs at both the aggregate and goods levels. Furthermore, the test for aggregate RER independence can be conducted in the same manner as in (36). Rejecting the null hypothesis of $\beta = 0$ implies a rejection of the absence of roundabout production, as $r = 0$ would imply $\beta = 0$. This result has an important implication for testing aggregate RER independence. Even if firms are fully attentive, the data may still reject the null hypothesis of $\beta = 0$ in the presence of roundabout production.

Yet, we can still evaluate the model of roundabout production using the relationship that $r\nu$ in (47) corresponds to β in (36). Simulating $r\nu$ with the parameter values calibrated in the previous section, we evaluate whether the simulated $r\nu$ matches the estimated β.

While we leave the details of the analysis to Appendix A.11, the simulated $r\nu$ fails to match the estimated β. We employ an extremely large value of $r = 0.99$ that maximizes the persistence of the aggregate RER in the model of roundabout production. Under our parameter values, the simulated value of $r\nu$ is only 0.55, which is much lower than the estimated β. Among the estimates of β in Tables 1 and 2, specification (2) of Table 1 for the US–Canadian city pairs provides the lowest estimated value of 0.80, with a standard error of 0.03. The simulated value $r\nu = 0.55$ falls outside the 95 percent confidence interval, [0.75, 0.86]. This result sharply contrasts with the model of behavioral inattention, which can easily accommodate the estimated β within the range of $0 < m < 1$.

Figure 6 depicts the impulse response functions of the aggregate RER to a nominal shock, comparing the model of roundabout production with that of behavioral inattention. The line

[44] The parameter ν is defined as $\nu = (1-\lambda\delta)\psi^{RP}/(1-\lambda\delta\theta_1)$, where θ_1 is the first-order autocorrelation of $\ln q_t$ in the model of roundabout production. Using the method of undetermined coefficients, we can show that θ_1 is given by $\theta_1 = [1/(2\lambda\delta)]\left\{1 + \lambda^2\delta - (1-\lambda)(1-\lambda\delta)\psi^{RP} r - \sqrt{[1 + \lambda^2\delta - (1-\lambda)(1-\lambda\delta)\psi^{RP} r]^2 - 4\lambda^2\delta}\right\}$. For further details, see Appendix A.11.

with squares represents the impulse response function for the model of roundabout production when $r = 0.99$. The line with pluses reproduces the impulse response function for the model of behavioral inattention, using the estimate from the US–Canadian city pairs ($m = 0.11$). Additionally, the impulse response function with asterisks, replicated from Figure 4, represents the model of full attention (without roundabout production). As before, we simulate the aggregate RER depreciation, normalized to unity on impact, and draw a horizontal line at 0.5.

The model of roundabout production also fails to generate the observed half-life. Even with $r = 0.99$, the half-life of the aggregate RER is only about 1.10 years.[45] While the half-life is slightly longer than that in the model of full attention, it remains considerably shorter than the observed data. This result is consistent with Kehoe and Midrigan (2007) who extended their baseline model with roundabout production. They argued that adding real rigidity moves the theory in the right direction, but it is still far from the data. In contrast, the model of behavioral inattention yields a much longer half-life of 3.74 years, which is more consistent with empirical observations.

What drives the differences between the two models? In Appendix A.12, we derive the optimal reset prices in nominal terms for both models. Suppressing constant terms, the optimal reset prices in nominal terms are given by

$$\ln P_{Hit} \;=\; (1 - m_{1H}) \ln P_t + m_{1H}(1 - \lambda\delta) \sum_{k=0}^{\infty} (\lambda\delta)^k \mathbb{E}_t \ln W_{t+k}, \tag{48}$$

for the model of behavioral inattention, and

$$\ln P_{Hit} \;=\; (1 - \lambda\delta) \sum_{k=0}^{\infty} (\lambda\delta)^k \mathbb{E}_t \left[r \ln P_{t+k} + (1 - r) \ln W_{t+k} \right], \tag{49}$$

for the model of roundabout production.[46]

This difference between the two equations highlights why behavioral inattention is more effective in addressing the PPP puzzle. To facilitate the comparison, we assume $r = 1 - m_{1H}$. In this case, the two equations become identical when $\lambda = 0$ or $\lambda = 1$. However, when $0 < \lambda < 1$, the coefficient on $\ln P_t$ in the right-hand side of (48) is r, whereas that in (49) is $(1 - \lambda\delta)r$. Since the former coefficient is strictly larger than the latter, the dependence of the optimal reset price on the aggregate prices is stronger in (48). Consequently, the de-

[45] Under our model's assumption, we can show that the aggregate RER follows an AR(1) process. The first-order autocorrelation is only 0.53 even under $r = 0.99$.

[46] The appendix also shows that our model's assumptions lead to (48) to (29).

gree of aggregate RER dependence also increases. This fact has been overlooked in previous studies, as comparisons of models with strategic complementarity have typically relied on the form of (49) (see Burstein and Gopinath, 2014). Thus, the framework of behavioral inattention introduces a new dimension to the literature on international pricing in macroeconomic models.

6.2 Heterogeneity in price stickiness

In Section 5, we showed that while behavioral inattention significantly improves the prediction of the half-life, the estimated degree of attention that allows for the heterogeneity in price stickiness further improves the predictions. Especially in the US–Canadian city pairs, the contribution of the heterogeneity in price stickiness to the increase in half-lives is nonnegligible. Out of the total improvement in prediction when considering both behavioral inattention and the heterogeneity in price stickiness, the contribution of the latter is 35.5 percent for the aggregate RER and 41.0 percent for the good-level RER. However, this relative contribution only reflects the effect of heterogeneity in price stickiness when estimating m. We do not consider the aggregation bias in multisector sticky-price models with heterogeneity in price stickiness, as discussed in Carvalho and Nechio (2011).

Thus, a natural question arises: Can aggregation bias in multisector sticky-price models *alone* explain the PPP puzzle? We argue that the answer is no because Carvalho and Nechio's (2011) results in the multisector model rely on the persistence of NER growth. In their model, NER growth follows an AR(1) process:

$$\Delta \ln S_t = \rho_{\Delta s} \Delta \ln S_{t-1} + \varepsilon_t^n, \tag{50}$$

with $\rho_{\Delta s} = 0.80$ at a monthly frequency. However, as they acknowledge, this value of $\rho_{\Delta s}$ is unrealistically high compared to what NER data suggest. They point out that "(f)or $\rho_{\Delta s} \approx 0.35$, the model falls short of generating as much persistence in real exchange rates as in the data, even with heterogeneity in price stickiness" (Carvalho and Nechio, 2011, p. 2418). In contrast, our model of behavioral inattention can fully explain the persistent aggregate RER under the assumption of no persistence in NER growth (i.e., $\rho_{\Delta s} = 0$).

We thus conclude that while the aggregation bias in sticky-price models may partially explain the PPP puzzle, it is insufficient to fully account for it under the realistic stochastic process of the NER. Beyond aggregation bias, an additional framework is needed to fill the gap between the model and the data. Behavioral inattention could be a promising candidate for this role.

6.3 Rational inattention

While our discussion thus far has centered on Gabaix's model of inattention, alternative models of inattention also deserve careful consideration. For example, Maćkowiak and Wiederholt (2009) propose a model of rational inattention, in which firms allocate attention to noisy signals about nominal aggregate demand and idiosyncratic productivity, subject to an information processing capacity constraint.[47] In this section, given the conceptual similarities between rational inattention models and our model, we connect our model to the literature on rational inattention.

We first derive the dynamic equation for the good-level RER from the rational inattention model of Maćkowiak and Wiederholt (2009) in which firms choose nominal prices under flexible prices and pay attention to stationary (or detrended) nominal aggregate demand. To incorporate rational inattention into our two-country model, we follow their assumptions that firms' choice variable is nominal prices and set $\lambda = 0$. Furthermore, while we maintain their assumption that firms observe noisy signals about nominal aggregate demand (or equivalently, nominal money supply, given $M_t = P_t C_t$), we additionally assume that M_{t-1} is fully known at the beginning of period t. Under this assumption, information about the (nonstationary) nominal money supply M_t are equivalent to information about the detrended nominal money supply (i.e., its growth rate M_t/M_{t-1}). As we will show, this information structure facilitates comparison with the model of behavioral inattention.

In this model, each period is divided into two stages. In stage 1, firms form expectations about the nominal money supply and idiosyncratic productivity based on noisy signals. Using the example of US firms selling their brands in US markets, the firm's optimal price under flexible prices is given by

$$\begin{aligned} \ln P_{Hit}(z) &= \mathbb{E}_{izt}(\ln M_t - \ln a_{it}) \\ &= \ln M_{t-1} + \mathbb{E}_{izt}\varepsilon_t^M - \mathbb{E}_{izt}\varepsilon_{it}^a, \end{aligned} \tag{51}$$

where constant terms are suppressed, and $\mathbb{E}_{izt}(\cdot)$ denotes the expectation operator conditional on brand-specific signals received by firms. Expectations may differ across brands so that the optimal prices depend on z. This equation reflects the firm's expected nominal marginal cost,

[47] See also the pioneering work by Sims (2003) and the comprehensive survey by Maćkowiak and Wiederholt (2023).

$\mathbb{E}_{izt}(\ln W_t - \ln a_{it}) = \mathbb{E}_{izt}(\ln M_t - \ln a_{it})$.[48] Firms observe the signals $s_{it}^M(z) = \ln M_t + \xi_{it}^M(z)$ and $s_{it}^a(z) = \ln a_{it} + \xi_{it}^a(z)$. Because M_{t-1} is fully known at the beginning of period t, it can be taken outside the expectation operator, leaving only the shocks inside. At this stage, firms set their prices based on their expectations about fundamentals. In stage 2, the true values of the money supply and productivity are revealed, and all subsequent decisions (e.g., consumption and employment) are made under the predetermined prices.

We assume that shocks to fundamentals are normally distributed (i.e., $\varepsilon_t^M \sim N(0, \sigma_M^2)$ and $\varepsilon_{it}^a \sim N(0, \sigma_a^2)$). For simplicity, we also assume that the noises satisfy $\xi_{it}^M(z) \sim N(0, \sigma_{\xi M}^2)$ and $\xi_{it}^a(z) \sim N(0, \sigma_{\xi a}^2)$.[49] Let m_{1H}^{RI} and m_{2H}^{RI} denote the steady-state Kalman gains for US firms when extracting information about the nominal money supply in the US and idiosyncratic labor productivity, respectively. These Kalman gains depend on the variance of noise, as implied by Kalman filtering. Firms rationally choose m_{1H}^{RI} and m_{2H}^{RI} by minimizing $\sigma_{\xi M}^2$ and $\sigma_{\xi a}^2$, respectively, subject to an information processing capacity constraint.

Appendix A.13 shows that the above information structure yields the following equation:

$$\ln q_{it} = (1 - m^{RI})\varepsilon_t^n + \psi^{RI}\varepsilon_{it}^r, \tag{52}$$

where $m^{RI} = \omega m_{1H}^{RI} + (1-\omega)m_{1F}^{RI}$ and $\psi^{RI} = \omega m_{2H}^{RI} - (1-\omega)m_{2F}^{RI}$. Here, m_{jF}^{RI} for $j = 1, 2$ denotes the steady-state Kalman gain for Canadian firms selling their brands in US markets.

We next derive a comparable dynamic equation from the model of behavioral inattention. To facilitate comparison between the two models of inattention, we adopt assumptions that differ from those used in the main analysis of behavioral inattention. As in the rational inattention model above, firms are assumed to pay attention to the nominal money supply, M_t. However, as noted in Section A.2, the degree of attention becomes one when firms attend to nonstationary nominal variables. Nonetheless, behavioral inattention can still be analyzed in this context if the degrees of attention are chosen *separately* for the nonstationary $\ln M_{t-1}$ and for the stationary ε_t^M, as defined in (4). Under this specification, the degree of attention to $\ln M_{t-1}$ is one, while that to ε_t^M is less than one. Using the example of US firms selling their brands in US markets, the firm's optimal price under flexible prices is given by

$$\ln P_{Hit} = \ln M_{t-1} + m_{1H}^{BI}\varepsilon_t^M - m_{2H}^{BI}\varepsilon_{it}^a, \tag{53}$$

[48]We suppress a constant term since nominal wages are given by $W_t = \chi M_t$, as implied by (2).

[49]In the rational inattention model, firms choose the distribution of signals through their attention decisions. Under the assumptions of Gaussian fundamentals (ε_t^M and ε_{it}^a), a quadratic objective function, and an unbounded choice set for $\ln P_{it+k}(z)$, Gaussian signals are optimal, which justifies the normality assumption of signals. See Maćkowiak et al. (2023, p. 231) for further details.

where m^{BI}_{jH} for $j = 1, 2$ is the degree of attention under behavioral inattention. Here, we use a superscript BI to differentiate it from rational inattention (RI). For $j = 1$, the object of attention is money growth, $\varepsilon^M_t = \Delta \ln M_t$, rather than the aggregate component of real marginal costs. For $j = 2$, as before, the object of attention is idiosyncratic labor productivity. Note that the coefficient on $\ln M_{t-1}$ is unity, as the degree of attention to $\ln M_{t-1}$ is one. This equation resembles (51), in which the expectations of shocks are replaced by shocks attenuated by the degrees of attention in (53).

While we leave the details to Appendix A.13, the resulting dynamic equation for the good-level RER becomes

$$\ln q_{it} = (1 - m^{BI})\varepsilon^n_t + \psi^{BI}\varepsilon^r_{it}, \tag{54}$$

where m^{BI} and ψ^{BI} are similarly defined as m and ψ.

The two models of inattention share some similarities. First, the dynamic equations (52) and (54) have the same structure. Thus, depending on the assumptions, both models can generate observationally equivalent dynamics for the good-level RER. Second, even when prices are flexible (i.e., $\lambda = 0$), nominal shocks still affect RERs, implying monetary non-neutrality. If firms are inattentive (i.e., $m^{RI} < 1$ and $m^{BI} < 1$), nominal prices do not fully adjust to shocks in money growth, allowing nominal shocks to influence the good-level RER. Hence, in both models, inattention serves as a source of nominal rigidity rather than strategic complementarity. Third, both models fail to account for the PPP puzzle. In both behavioral and rational inattention, the aggregate RER is governed by $\ln q_t = (1 - m^l)\varepsilon^n_t$ for $l = RI$ or BI. Given that $\varepsilon^n_t = \varepsilon^M_t - \varepsilon^{M*}_t$ is i.i.d., the implied half-life of the aggregate RER is zero. Similarly, the good-level RER also exhibits a zero half-life, since both ε^n_t and ε^r_t are i.i.d. in (52) and (54).

The above comparison highlights the importance of framing in behavioral inattention and the flexibility of the model of behavioral inattention. As we noted earlier, Gabaix (2014) argues that different theoretical implications arise depending on whether agents adopt a "nominal" frame, in which agents pay attention to nominal variables, or a "real" frame, in which they pay attention to real variables. In our context, the distinction between real and nominal frames is crucial: Inattention to real marginal costs helps explain the PPP puzzle, while inattention to nominal marginal costs generates only nominal rigidities (as in the rational inattention model). Since theoretical results for the real frame are available in studies on behavioral inattention (but not in those on rational inattention), we take advantage of this to address the PPP puzzle.

6.4 The monetary policy rule

We have assumed a random walk process for the money supply in both countries (see (4) and (5)). While a constant money growth rule leads to an empirically plausible NER, the Taylor rule is also a realistic representation of monetary policy. To assess the robustness of our results, we replace the money growth rule with a Taylor rule.

While the details of the model are presented in Appendix A.14, we assume here that the US policy interest rate is determined by

$$\hat{R}_t = \rho_R \hat{R}_{t-1} + (1 - \rho_R)\alpha_\pi \pi_t + \varepsilon_t^R,$$

where R_t is the gross nominal interest rate in the US, and $\hat{R}_t$ denotes its log deviation from the steady state. The monetary policy shock is denoted by ε_t^R. The parameter $\rho_R \in [0, 1)$ captures interest rate inertia, and the inflation coefficient satisfies $\alpha_\pi > 1$. The Canadian policy interest rate $\hat{R}_t^*$ follows a symmetric structure to that of the US. In this model, the NER is determined by the uncovered interest parity.[50]

The model with the Taylor rule yields two main findings. First, under full attention, the persistence of the aggregate RER is even lower than the degree of price stickiness ($\rho_q < \lambda$), and also lower than the persistence of the good level RER ($\rho_q < \rho_{qi}$). Given that ρ_q is bounded below by λ and ρ_{qi} under the baseline model with the constant money growth rule (see Propositions 2 and 3), the model with the Taylor rule under full attention produces substantially worse predictions. The introduction of the persistence of monetary policy shock (ε_t^R) improves the predictions of the model but the predicted ρ_q and ρ_{qi} under the full attention remain inconsistent with the data.

Second, in contrast to full attention, behavioral inattention better accounts for the persistence of both aggregate and good-level RERs, even in the model with the Taylor rule. Under a reasonable value of m, the model successfully predicts $\rho_q > \rho_{qi} > \lambda$ at the levels consistent with the data. Notably, the predicted half-lives of the aggregate RER range from 2.3 to 2.6 years and those of the good-level RER range from 1.5 to 1.8 years, exceeding 0.64 years implied from the baseline model with the constant money growth rule under full attention.[51] Thus, the model of behavioral inattention continues to have a powerful mechanism to generate persistent RERs at both the aggregate and goods levels.

[50] Using a model similar to Crucini et al. (2013), Nakamura (2022) investigates the volatility of LOP deviations under the Taylor rule.

[51] See the first column of Table A.6 discussed in Appendix A.14.

7 Conclusion

In this paper, we addressed two empirical anomalies. First, the observed aggregate RER is much more persistent than theoretical predictions based on the standard model of nominal price rigidities. Second, micro price evidence indicates that good-level RERs are often less persistent than the aggregate RER. To reconcile the PPP and LOP evidence, we adapted the model of behavioral inattention in Gabaix (2014) to a two-country sticky-price model. We showed that firms' pricing under inattention generates a dependence of good-level RERs on the aggregate RER, which is the key to explaining the puzzling behavior of RERs.

Using international price data, we tested whether good-level RERs independent or the aggregate RER. We strongly reject the null hypothesis of aggregate RER independence. In our model of behavioral inattention, aggregate RER dependence leads to an aggregate RER that is more than twice as persistent as predicted by nominal rigidities alone. Our model also predicts that the persistence of good-level RERs is less than two-thirds that of the aggregate RER. It quantitatively replicates the observed half-lives of both aggregate and good-level RERs.

We also explored alternative explanations for the puzzling persistence of RERs in comparison to our model. We showed that although the roundabout production model yields a dynamic equation for the good-level RER similar to that in the model of behavioral inattention, it fails to replicate the observed persistence of the aggregate RER under our calibration. We further discussed that a multisector sticky-price model with heterogeneity in price stickiness is promising, but insufficient to fully account for the PPP puzzle. Rational inattention is another alternative model of inattention and can generate observationally equivalent dynamics for the good-level RER, depending on assumptions. We argued that behavioral inattention offers greater flexibility in modeling which variables economic agents pay attention to.

Nevertheless, further empirical analysis is desirable to distinguish models of inattention from other models with strategic complementarity. A key distinguishing feature is that the degree of attention is endogenous.[52] Models of inattention predict that endogenous attention increases when the variance of shocks increases. For this reason, during the period in which shocks are volatile, such as the COVID-19 pandemic, strategic complementarity is likely to weaken, holding other factors unchanged. In contrast, strategic complementarity arising from the production or the utility functions (e.g., roundabout production, kinked demand curves, or translog preferences) is less likely to be sensitive to time variations in shock uncertainty. Therefore, one testable implication is that RERs should exhibit greater persistence during

[52] Weber et al. (2025) provide evidence for the state dependency of attention.

periods of lower aggregate shock volatility, after controlling for various factors that may also influence RER persistence. Yet, distinguishing behavioral inattention from rational inattention poses other challenges. To overcome this limitation, it would be ideal to examine survey data that directly ask firm managers about their pricing decisions. Such data could provide valuable insights and represent a promising avenue for future research.

References

[1] Álvarez, L. J., E. Dhyne, M. Hoeberichts, C. Kwapil, H. Le Bihan, P. Lünemann, F. Martins, R. Sabbatini, H. Stahl, R. Vermeulen, and J. Vilmunen, (2006). "Sticky prices in the Euro area: A summary of new micro-evidence," *Journal of the European Economic Association*, vol. 4(2/3), pp. 575-584.

[2] Anderson, J. E. and E. Van Wincoop (2004). "Trade costs," *Journal of Economic Literature*, vol. 42(3), pp. 691-751.

[3] Andrade, P. and M. Zachariadis (2016). "Global versus local shocks in micro price dynamics," *Journal of International Economics*, vol. 98, pp. 78-92.

[4] Arellano, M. and S. Bond (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, vol. 58(2), pp. 277-297.

[5] Basu, S. (1995). "Intermediate goods and business cycles: Implications for productivity and welfare," *American Economic Review*, vol. 85(3), pp. 512-531.

[6] Ball, L. and D. Romer (1990). "Real rigidities and the non-neutrality of money," *Review of Economic Studies*, vol. 57(2), pp. 183-203.

[7] Benigno, G. (2004). "Real exchange rate persistence and monetary policy rules," *Journal of Monetary Economics*, vol. 51(3), pp. 473-502.

[8] Bergin, P. R. and R. C. Feenstra (2001). "Pricing-to-market, staggered contracts, and real exchange rate persistence," *Journal of International Economics*, vol. 54(2), pp. 333-359.

[9] Bergin, P. R., R. Glick, and J.-L. Wu (2013). "The micro-macro disconnect of purchasing power parity," *Review of Economics and Statistics*, vol. 95(3), pp. 798-812.

[10] Bils, M. and P. J. Klenow (2004). "Some evidence on the importance of sticky prices," *Journal of Political Economy*, vol. 112(5), pp. 947-985.

[11] Blanco, A. and J. Cravino (2020). "Price rigidities and the relative PPP," *Journal of Monetary Economics*, vol. 116, pp. 104-116.

[12] Broda, C. and D. E. Weinstein (2006). "Globalization and the gains from variety," *Quarterly Journal of Economics*, vol. 121(2), pp. 541-585.

[13] Burstein, A. and G. Gopinath (2014). "International prices and exchange rates," in G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, vol. 4, pp. 391-451, Elsevier.

[14] Calvo, G. A. (1983). "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, vol. 12(3), pp. 383-398.

[15] Carvalho, C. and F. Nechio (2011). "Aggregation and the PPP puzzle in a sticky-price model," *American Economic Review*, vol. 101(6), pp. 2391-2424.

[16] Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2002). "Can sticky price models generate volatile and persistent real exchange rates?" *Review of Economic Studies*, vol. 69(3), pp. 533-563.

[17] Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory*, Hoboken, NJ: Wiley.

[18] Crucini, M. J. and A. Landry (2019). "Accounting for real exchange rates using micro-data," *Journal of International Money and Finance*, vol. 91, pp. 86-100.

[19] Crucini, M. J. and M. Shintani (2008). "Persistence in law of one price deviations: Evidence from micro-data," *Journal of Monetary Economics*, vol. 55(3), pp. 629-644.

[20] Crucini, M. J., M. Shintani, and T. Tsuruga (2010a). "Accounting for persistence and volatility of good-level real exchange rates: The role of sticky information," *Journal of International Economics*, vol. 81(1), pp. 48-60.

[21] Crucini, M. J., M. Shintani, and T. Tsuruga (2010b). "The law of one price without the border: The role of distance versus sticky prices," *Economic Journal*, vol. 120(544), pp. 462-480.

[22] Crucini, M. J., M. Shintani, and T. Tsuruga (2013). "Do sticky prices increase real exchange rate volatility at the sector level?" *European Economic Review*, vol. 62, pp. 58-72.

[23] Crucini, M. J. and C. I. Telmer (2020). "Microeconomic sources of real exchange rate variation," *Review of Economic Dynamics*, vol. 38, pp. 22-40.

[24] Crucini, M. J. and H. Yilmazkuday (2014). "Understanding long-run price dispersion," *Journal of Monetary Economics*, vol. 66, pp. 226-240.

[25] Elberg, A. (2016). "Sticky prices and deviations from the law of one price: Evidence from Mexican micro-price data," *Journal of International Economics*, vol. 98, pp. 191-203.

[26] Engel, C. (2019). "Real exchange rate convergence: The roles of price stickiness and monetary policy," *Journal of Monetary Economics*, vol. 103, pp. 21-32.

[27] Engel, C. and J. H. Rogers (1996). "How wide is the border?" *American Economic Review*, vol. 86(5), pp. 1112-1125.

[28] Engel, C. and J. H. Rogers (2004). "European product market integration after the Euro," *Economic Policy*, vol. 19(39), pp. 348-384.

[29] Gabaix, X. (2014). "A sparsity-based model of bounded rationality," *Quarterly Journal of Economics*, vol. 129(4), pp. 1661-1710.

[30] Gabaix, X. (2019). "Behavioral inattention," *Handbook of Behavioral Economics: Applications and Foundations* 1. Vol. 2. North-Holland, pp. 261-343.

[31] Gabaix, X. (2020). "A behavioral New Keynesian model," *American Economic Review*, vol. 110(8), pp. 2271-2327.

[32] Gautier E., C. Conflitti, R. P. Faber, B. Fabo, L. Fadejeva, V. Jouvanceau, J. O. Menz, T. Messner, P. Petroulas, P. Roldan-Blanco, F. Rumler, S. Santoro, E. Wieland, and H. Zimmer (2024). "New facts on consumer price rigidity in the Euro area," forthcoming *American Economic Journal: Macroeconomics.*

[33] Hickey, R. D. and D. S. Jacks (2011). "Nominal rigidities and retail price dispersion in Canada over the twentieth century," *Canadian Journal of Economics*, vol. 44(3), pp. 749-780.

[34] Imbs, J., H. Mumtaz, M. O. Ravn, and H. Rey (2005). "PPP strikes back: Aggregation and the real exchange rate," *Quarterly Journal of Economics*, vol. 120(1), pp. 1-43.

[35] Itskhoki, O. and D. Mukhin (2021). "Exchange rate disconnect in general equilibrium," *Journal of Political Economy*, vol. 129(8), pp. 2183-2232.

[36] Kehoe, P. J. and V. Midrigan (2007). "Sticky prices and sectoral real exchange rates," Federal Reserve Bank of Minneapolis, Research Department.

[37] Kimball, M. S. (1995). "The quantitative analytics of the basic neomonetarist model," *Journal of Money, Credit and Banking*, vol. 27(4), pp. 1241-1277.

[38] Klenow, P. J. and O. Kryvtsov (2008). "State-dependent or time-dependent pricing: Does it matter for recent US inflation?" *Quarterly Journal of Economics*, vol. 123(3), pp. 863-904.

[39] Maćkowiak, B. and M. Wiederholt (2009). "Optimal sticky prices under rational inattention," *American Economic Review*, vol. 99(3), pp. 769-803.

[40] Maćkowiak, B., F. Matějka, and M. Wiederholt (2023). "Rational inattention: A review," *Journal of Economic Literature*, vol. 61(1), pp. 226-273.

[41] Nakamura, E. and J. Steinsson (2008). "Five facts about prices: A reevaluation of menu cost models," *Quarterly Journal of Economics*, vol. 123(4), pp. 1415-1464.

[42] Nakamura, E. and J. Steinsson (2013). "Price rigidity: Microeconomic evidence and macroeconomic implications," *Annual Review of Economics*, vol. 5(1), pp. 133-163.

[43] Nakamura, F. (2022). "The origin of the law of one price deviations: Insights from the good-level real exchange rate volatility," *Journal of International Money and Finance*, vol. 128(102717).

[44] Rogers, J. H. (2007). "Monetary union, price level convergence, and inflation: How close is Europe to the USA?" *Journal of Monetary Economics*, vol. 54(3), pp. 785-796.

[45] Rogoff, K. (1996). "The purchasing power parity puzzle," *Journal of Economic Literature*, vol. 34(2), pp. 647-668.

[46] Sims, C. A. (2003). "Implications of rational inattention," *Journal of Monetary Economics*, vol. 50(3), pp. 665-690.

[47] Steinsson, J. (2008). "The dynamic behavior of the real exchange rate in sticky price models," *American Economic Review*, vol. 98(1), pp. 519-533.

[48] Tversky, A. and D. Kahneman (1974). "Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty," *Science*, vol. 185(4157), pp. 1124-1131.

[49] Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*: Princeton University Press.

[50] Yun, T. (1996). "Nominal price rigidity, money supply endogeneity, and business cycles," *Journal of Monetary Economics*, vol. 37(2), pp. 345-370.

[51] Weber, M., B. Candia, H. Afrouzi, T. Ropele, R. Lluberas, S. Frache, B. Meyer, S. Kumar, Y. Gorodnichenko, D. Georgarakos, O. Coibion, K. Geoff, and J. Ponce (2025). "Tell me something I don't already know: Learning in low-and high-inflation settings," *Econometrica*, vol. 93(1), pp. 229-264.

A Appendix

A.1 The objective function for the pricing decision

This appendix derives the objective functions that fully attentive firms aim to maximize. We begin with the case of US firms. The objective function for selling their brands in US markets is given by

$$v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} \left(1/P_{t+k}\right) \left[P_{it}(z) - \frac{W_{t+k}}{a_{it+k}}\right] c_{it,t+k}(z), \tag{55}$$

subject to the demand function by US consumers for brand z of good i conditional on the US firm having last reset its price in period t:

$$c_{it,t+k}(z) = \left[\frac{P_{it}(z)}{P_{it+k}}\right]^{-\varepsilon} c_{it+k}, \tag{56}$$

where $z \in [0, 1/2]$. Using the definitions of $p_{it}(z)$, w_t, and p_{it}, we can also rewrite (55) and (56) as

$$v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} \frac{P_t}{P_{t+k}} \left[p_{it}(z) - \frac{w_{t+k}}{a_{it+k}} \frac{P_{t+k}}{P_t}\right] c_{it,t+k}(z), \tag{57}$$

and

$$c_{it,t+k}(z)=\left\{\left[\frac{p_{it}(z)}{p_{it+k}}\right]\left(\frac{P_t}{P_{t+k}}\right)\right\}^{-\varepsilon}c_{it+k}, \tag{58}$$

respectively.

For a generic variable x_t, we express x_t as $x_t=\bar{x}\exp(\hat{x}_t)$, where $\hat{x}_t=\ln x_t-\ln\bar{x}$ and $\bar{x}$ is the steady-state value of x_t. In addition, the steady-state value of P_{t+k}/P_t is unity since the steady-state inflation is zero. Likewise, the steady-state value of a_{it} is unity from (6). Using these facts, (57) can be rewritten as

$$v_{it}(z)=\mathbb{E}_t\sum_{k=0}^{\infty}\lambda^k\delta_{t,t+k}\frac{P_t}{P_{t+k}}\left\{\bar{p}_i(z)\exp\left[\hat{p}_{it}(z)\right]-\bar{w}\exp\left(\hat{w}_{t+k}+\sum_{l=0}^{k}\pi_{t+l}-\hat{a}_{it+k}\right)\right\}c_{it,t+k}(z),$$

where $P_{t+k}/P_t=\prod_{l=1}^{k}P_{t+l}/P_{t+l-1}=\exp\left[\sum_{l=1}^{k}\ln\left(P_{t+l}/P_{t+l-1}\right)\right]=\exp\left(\sum_{l=1}^{k}\pi_{t+l}\right)$. Using the definition of $\hat{\mu}_{Ht+k}=\hat{w}_{t+k}+\sum_{l=0}^{k}\pi_{t+l}$ yields (10) in the main text. For the demand function, we also rewrite (58) to obtain (11):

$$c_{it,t+k}(z)=\left[\frac{\bar{p}_i(z)}{\bar{p}_i}\right]^{-\varepsilon}\exp\left\{-\varepsilon\left[\hat{p}_{it}(z)-\sum_{l=1}^{k}\pi_{t+l}-\hat{p}_{it+k}\right]\right\}c_{it+k}. \tag{59}$$

Next, we derive the objective function for US firms selling their brands in Canadian markets. When exporting their brands, firms set prices in the local currency. Under this assumption, the objective function is

$$v_{it}^{*}(z)=\mathbb{E}_t\sum_{k=0}^{\infty}\lambda^k\delta_{t,t+k}\left(1/P_{t+k}\right)\left[S_{t+k}P_{it}^{*}(z)-(1+\tau)\frac{W_{t+k}}{a_{it+k}}\right]c_{it+k}^{*}(z), \tag{60}$$

subject to the demand function by Canadian consumers conditional on the US firm having last reset its price in period t:

$$c_{it+k}^{*}(z)=\left[\frac{P_{it}^{*}(z)}{P_{it+k}^{*}}\right]^{-\varepsilon}c_{it+k}^{*}, \tag{61}$$

where $z\in[0,1/2]$.

Using the definitions of $p_{it}^*(z) = P_{it}^*(z)/P_t^*$ and $p_{it}^* = P_{it}^*/P_t^*$, we rewrite (60) as follows:

$$\begin{aligned} v_{it}^*(z) &= \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} \frac{S_{t+k} P_{t+k}^*}{P_{t+k}} \left[\frac{P_{it}^*(z)}{P_t^*} \frac{P_t^*}{P_{t+k}^*} - (1+\tau) \frac{P_{t+k}}{S_{t+k} P_{t+k}^*} \frac{W_{t+k}/P_{t+k}}{a_{it+k}} \right] c_{it,t+k}^*(z) \\ &= \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} q_{t+k} \left[p_{it}^*(z) \frac{P_t^*}{P_{t+k}^*} - (1+\tau) \frac{w_{t+k}}{q_{t+k} a_{it+k}} \right] c_{it,t+k}^*(z) \\ &= \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} q_{t+k} \frac{P_t^*}{P_{t+k}^*} \left[p_{it}^*(z) - (1+\tau) \frac{w_{t+k}}{q_{t+k} a_{it+k}} \frac{P_{t+k}^*}{P_t^*} \right] c_{it,t+k}^*(z). \end{aligned}$$

Again, using the expression of $x_t = \bar{x} \exp(\hat{x}_t)$, the assumption of the zero-inflation steady state and (7), (60) can be rewritten as follows:

$$v_{it}^*(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} q_{t+k} \times \frac{P_t^*}{P_{t+k}^*} \left\{ \bar{p}_i^*(z) \exp\left[\hat{p}_{it}^*(z)\right] - (1+\tau) \frac{\bar{w}}{\bar{q}} \exp\left(\hat{\mu}_{Ht+k}^* - \hat{a}_{it+k} \right) \right\} c_{it,t+k}^*(z),$$

where $\hat{\mu}_{Ht+k}^* = \hat{w}_{t+k} - \hat{q}_{t+k} + \sum_{l=1}^{k} \pi_{t+l}^*$. The demand function by Canadian consumers can be derived from (61) in the same way as (11) is derived from (56).

We can similarly derive the objective function of Canadian firms indexed by $z \in (1/2, 1]$. When Canadian firms sell their brands in Canadian markets, the objective function that fully attentive Canadian firms aim to maximize is

$$\begin{aligned} v_{it}^*(z) &= \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k}^* \left(1/P_{t+k}^*\right) \left[P_{it}^*(z) - \frac{W_{t+k}^*}{a_{it+k}^*} \right] c_{it,t+k}^*(z) \\ &= \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k}^* \frac{P_t^*}{P_{t+k}^*} \left[p_{it}^*(z) - \frac{w_{t+k}^*}{a_{it+k}^*} \left(\frac{P_{t+k}^*}{P_t^*} \right) \right] c_{it,t+k}^*(z) \\ &= \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k}^* \frac{P_t^*}{P_{t+k}^*} \left\{ \bar{p}_i^*(z) \exp\left[\hat{p}_{it}^*(z)\right] - \bar{w}^* \exp\left(\hat{\mu}_{Ft+k}^* - \hat{a}_{it+k}^* \right) \right\} c_{it,t+k}^*(z), \end{aligned}$$

where $\hat{\mu}_{Ft+k}^* = \hat{w}_{t+k}^* + \sum_{l=1}^{k} \pi_{t+l}^*$. Similarly, when Canadian firms sell their brands in US markets, the objective function that fully attentive Canadian firms aim to maximize is

$$
\begin{aligned}
v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta^*_{t,t+k} \left(1/P^*_{t+k}\right) \left[\frac{P_{it}(z)}{S_{t+k}} - (1+\tau)\frac{W^*_{t+k}}{a^*_{it+k}} \right] c_{it,t+k}(z) \\
= \quad \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta^*_{t,t+k} \left(\frac{P_{t+k}}{S_{t+k} P^*_{t+k}} \right) \left[\frac{P_{it}(z)}{P_t} \frac{P_t}{P_{t+k}} - (1+\tau) \frac{W^*_{t+k}/P^*_{t+k}}{a^*_{it+k}} \frac{S_{t+k} P^*_{t+k}}{P_{t+k}} \right] c_{it,t+k}(z) \\
= \quad \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta^*_{t,t+k} q^{-1}_{t+k} \left[p_{it}(z) \frac{P_t}{P_{t+k}} - (1+\tau) \frac{w^*_{t+k} q_{t+k}}{a^*_{it+k}} \right] c_{it,t+k}(z) \\
= \quad \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta^*_{t,t+k} q^{-1}_{t+k} \frac{P_t}{P_{t+k}} \left[p_{it}(z) - (1+\tau) \frac{w^*_{t+k} q_{t+k}}{a^*_{it+k}} \frac{P_{t+k}}{P_t} \right] c_{it,t+k}(z) \\
= \quad \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta^*_{t,t+k} q^{-1}_{t+k} \frac{P_t}{P_{t+k}} \left\{ \bar{p}_i(z) \exp\left[\hat{p}_{it}(z)\right] - (1+\tau) \bar{w}^* \bar{q} \exp\left(\hat{\mu}_{Ft+k} - \hat{a}_{it+k}\right) \right\} c_{it,t+k}(z),
\end{aligned}
$$

where $\hat{\mu}_{Ft+k} = \hat{w}_{t+k} + \hat{q}_{t+k} + \sum_{l=1}^{k} \pi_{t+l}$.

The demand functions conditional on the Canadian firm having last reset its price in period t can easily be derived in the same way as the derivation of (11).

A.2 The optimal reset prices under behavioral inattention

We first derive (17). Using the definition of $\hat{\mu}_{Ht+k} = \hat{w}_{t+k} + \sum_{l=1}^{k} \pi_{t+l}$, we rewrite the log-linearized first-order condition (16) as

$$
\begin{aligned}
& \hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) \\
= \quad & (1-\lambda\delta)\mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k (m_{1H} \hat{w}_{t+k} - m_{2H} \hat{a}_{it+k}) + m_{1H}(1-\lambda\delta)\mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k \sum_{l=1}^{k} \pi_{t+l}. \quad (62)
\end{aligned}
$$

We separately arrange the terms in the right-hand side of (62). First, note that

$$
\begin{aligned}
& (1-\lambda\delta)\mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k (m_{1H}\hat{w}_{t+k} - m_{2H}\hat{a}_{it+k}) \\
= \;& \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k (m_{1H}\hat{w}_{t+k} - m_{2H}\hat{a}_{it+k}) - \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^{k+1} (m_{1H}\hat{w}_{t+k} - m_{2H}\hat{a}_{it+k}) \\
= \;& m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it} \\
& + \mathbb{E}_t[(\lambda\delta)^1 (m_{1H}\hat{w}_{t+1} - m_{2H}\hat{a}_{it+1}) - (\lambda\delta)^1 (m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it})] \\
& + \mathbb{E}_t[(\lambda\delta)^2 (m_{1H}\hat{w}_{t+2} - m_{2H}\hat{a}_{it+2}) - (\lambda\delta)^2 (m_{1H}\hat{w}_{t+1} - m_{2H}\hat{a}_{it+1})] \\
& + ... \\
= \;& m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it} + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k (m_{1H}\Delta\hat{w}_{t+k} - m_{2H}\Delta a_{it+k}).
\end{aligned}
$$

Next, the remaining terms are

$$
\begin{aligned}
& m_{1H}(1-\lambda\delta)\mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k \sum_{l=1}^{k} \pi_{t+l} \\
= \;& m_{1H}(1-\lambda\delta)\mathbb{E}_t \begin{bmatrix} \quad (\lambda\delta)\pi_{t+1} \\ +(\lambda\delta)^2\pi_{t+1} + (\lambda\delta)^2\pi_{t+2} \\ +(\lambda\delta)^3\pi_{t+1} + (\lambda\delta)^3\pi_{t+2} + (\lambda\delta)^3\pi_{t+3} \\ +... \end{bmatrix} \\
= \;& m_{1H}(1-\lambda\delta)\mathbb{E}_t \left\{ (\lambda\delta) \left[\sum_{k=0}^{\infty} (\lambda\delta)^k \right] \pi_{t+1} + (\lambda\delta)^2 \left[\sum_{k=0}^{\infty} (\lambda\delta)^k \right] \pi_{t+2} + (\lambda\delta)^3 \left[\sum_{k=0}^{\infty} (\lambda\delta)^k \right] \pi_{t+3} + ... \right\} \\
= \;& m_{1H}\mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k \pi_{t+k},
\end{aligned}
$$

where the last line uses $\sum_{k=0}^{\infty} (\lambda\delta)^k = (1-\lambda\delta)^{-1}$. Finally, combining the above expressions, (62) becomes

$$
\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) = (m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it}) + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k \{m_{1H}(\Delta\hat{w}_{t+k} + \pi_{t+k}) - m_{2H}\Delta\hat{a}_{it+k}\}. \quad (63)
$$

Under the preferences given by $U(c,n) = \ln c - \chi n$, the first-order conditions of US households (2) imply that $W_t/P_t = \chi c_t$. In terms of log deviations,

$$
\hat{w}_t = \hat{c}_t. \quad (64)
$$

In addition, their CIA constraint ($M_t = P_t c_t$) leads to $\pi_t = \ln M_t/M_{t-1} - \Delta\hat{c}_t$. Thus, using (4), we have $\Delta\hat{w}_t + \pi_t = \ln M_t/M_{t-1} = \varepsilon_t^M$. As a result, (63) becomes

$$\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) = (m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it}) - m_{2H}\mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k \Delta a_{it+k}.$$

If the stochastic process $\hat{a}_{it}$ is given by (6), $\mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k \Delta a_{it+k} = -\lambda\delta\hat{a}_{it}$. Therefore,

$$\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) = m_{1H}\hat{w}_t - m_{2H}(1-\lambda\delta)\hat{a}_{it},$$

which is (17) in the main text.

For the optimal reset price of goods exported by US firms, we have

$$\begin{aligned} \hat{p}^*_{Hi}(\hat{\boldsymbol{x}}^*_{Hit}, \boldsymbol{m}^*_H) &= [m^*_{1H}(\hat{w}_t - \hat{q}_t) - m^*_{2H}\hat{a}_{it}] \\ &+ \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k [m^*_{1H}(\Delta\hat{w}_{t+k} - \Delta\hat{q}_{t+k} + \pi^*_{t+k}) - m^*_{2H}\Delta\hat{a}_{it+k}]. \end{aligned} \tag{65}$$

Here, using the international risk-sharing condition (3), the log deviation of the aggregate RER is

$$\hat{q}_t = \hat{c}_t - \hat{c}^*_t, \tag{66}$$

which also implies

$$\hat{q}_t = \hat{w}_t - \hat{w}^*_t, \tag{67}$$

because of (64) and its foreign analogue (i.e., $\hat{w}^*_t = \hat{c}^*_t$). Thus, (67) implies that (65) can be rewritten as

$$\hat{p}^*_{Hi}(\hat{\boldsymbol{x}}^*_{Hit}, \boldsymbol{m}^*_H) = (m^*_{1H}\hat{w}^*_t - m^*_{2H}\hat{a}_{it}) + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k [m^*_{1H}(\Delta\hat{w}^*_{t+k} + \pi^*_{t+k}) - m^*_{2H}\Delta\hat{a}_{it+k}]. \tag{68}$$

This equation has the same structure as (63). Using the foreign analogue of the CIA constraint, (5), and (6), the above equation can be simplified to

$$\hat{p}^*_{Hi}(\hat{\boldsymbol{x}}^*_{Hit}, \boldsymbol{m}^*_H) = m^*_{1H}\hat{w}^*_t - m^*_{2H}(1-\lambda\delta)\hat{a}_{it}. \tag{69}$$

The remaining optimal reset prices, namely $\hat{p}^*_{Fi}(\hat{\boldsymbol{x}}^*_{Fit}, \boldsymbol{m}^*_F)$ and $\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F)$, are analogously derived. In particular, we obtain

$$\hat{p}^*_{Fi}(\hat{\boldsymbol{x}}^*_{Fit}, \boldsymbol{m}^*_F) = m^*_{1F}\hat{w}^*_t - m^*_{2F}(1-\lambda\delta)\hat{a}^*_{it}, \tag{70}$$

and

$$\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F) = m_{1F}\hat{w}_t - m_{2F}(1-\lambda\delta)\hat{a}^*_{it}. \tag{71}$$

A.3 The sparse max

In this appendix, we describe the sparse max for firms, using the example of US firms for selling their brands in US markets. Specifically, we derive equations (20)–(22). The US firms' objective function for choosing $\boldsymbol{m}_H$ is based on the second-order Taylor expansion of $\mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right) - \mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right)$ around the steady state (i.e., $\hat{\boldsymbol{x}}_{Hit} = 0$), which represents the expected profit loss of deviating attention from full attention in pricing. Here, profit loss arises from choosing a price distorted by behavioral inattention.

To obtain (20)–(22), we take three steps. First, we perform a quadratic approximation of $\mathbb{E}v_{Hi}\left(p_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right)$ around the steady state. Here, the firm's profits are evaluated at $\boldsymbol{m}_H = \boldsymbol{\iota}$ (which appears in the third argument of $v_{Hi}(\cdot)$), while the price is distorted by behavioral inattention due to $\boldsymbol{m}_H \neq \boldsymbol{\iota}$ (which appears in the second argument of $\hat{p}_{Hi}(\cdot)$. The quadratic approximation of $v_{Hi}\left(p_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right)$ around the steady state is

$$\begin{aligned}
& v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right) \\
\simeq\ & v^0_{Hi} + \frac{\partial v^0_{Hi}}{\partial \hat{p}_{it}(z)}\mathbb{E}_t\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) + \frac{\partial v^0_{Hi}}{\partial \hat{\boldsymbol{x}}'_{Hit}}\mathbb{E}_t\hat{\boldsymbol{x}}_{Hit} \\
& + \frac{1}{2}\frac{\partial^2 v^0_{Hi}}{\partial \hat{p}_{it}(z)^2}\mathbb{E}_t[\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)]^2 + \mathbb{E}_t\left[\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v^0_{Hi}}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{p}_{it}(z)}\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)\right] \\
& + \frac{1}{2}\mathbb{E}_t\left(\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v^0_{Hi}}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{\boldsymbol{x}}'_{Hit}}\hat{\boldsymbol{x}}_{Hit}\right),
\end{aligned} \tag{72}$$

where $v^0_{Hi} = v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right)\big|_{\hat{\boldsymbol{x}}_{Hit}=\boldsymbol{0}} = v_{Hi}(\hat{p}_{Hi}(\boldsymbol{0}, \boldsymbol{m}_H), \boldsymbol{0}, \boldsymbol{\iota}) = v_{Hi}(\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{\iota})$. The first and the second derivatives are similarly defined. Taking the unconditional expectations of the above equation yields

$$\begin{aligned}
\mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right) \ \simeq\ & v^0_{Hi} + \frac{1}{2}\frac{\partial^2 v^0_{Hi}}{\partial \hat{p}_{it}(z)^2}\mathbb{E}[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)]^2 \\
& + \mathbb{E}\left[\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v^0_{Hi}}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{p}_{it}(z)}\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)\right] \\
& + \frac{1}{2}\mathbb{E}\left(\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v^0_{Hi}}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{\boldsymbol{x}}'_{Hit}}\hat{\boldsymbol{x}}_{Hit}\right),
\end{aligned} \tag{73}$$

where we used the law of iterated expectations. Note that the linearly approximated terms in (72) are all zero.

We rewrite (73), using the first-order condition for pricing. Based on the attention-augmented objective function, the first-order condition is $\partial v_{Hi}\left(\hat{p}_{it}(z), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H\right)/\partial \hat{p}_{it}(z) = 0$. However, when the attention-augmented objective function is evaluated at $\boldsymbol{m}_H = \boldsymbol{\iota}$, only $\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota})$, namely a price under full attention, satisfies the first-order condition. In equation, it is given by

$$\frac{\partial v_{Hi}(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota})}{\partial \hat{p}_{it}(z)} = 0. \tag{74}$$

Take the total derivatives of the above first-order condition and then evaluate partial derivatives at $\hat{\boldsymbol{x}}_{Hit} = \mathbf{0}$, we have

$$\frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)\partial \hat{\boldsymbol{x}}'_{Hit}}\hat{\boldsymbol{x}}_{Hit} = -\frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2}\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}). \tag{75}$$

Substituting this equation into (73) yields

$$\begin{aligned}
& \mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H), \hat{\boldsymbol{\mu}}_{Ht}, \boldsymbol{\iota}\right) \\
\simeq\ & v_{Hi}^0 + \frac{1}{2}\frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2}\mathbb{E}[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)]^2 - \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2}\mathbb{E}\left[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota})\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)\right] \\
& + \frac{1}{2}\mathbb{E}\left(\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v_{Hi}^0}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{\boldsymbol{x}}'_{Hit}}\hat{\boldsymbol{x}}_{Hit}\right) \\
=\ & v_{Hi}^0 + \frac{1}{2}\frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2}\left\{\mathbb{E}[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)]^2 - 2\mathbb{E}\left[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota})\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)\right]\right\} \\
& + \frac{1}{2}\mathbb{E}\left(\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v_{Hi}^0}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{\boldsymbol{x}}'_{Hit}}\hat{\boldsymbol{x}}_{Hit}\right).
\end{aligned} \tag{76}$$

Second, we take the quadratic approximation of $\mathbb{E}v_{Hi}\left(p_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}), \hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}\right)$, which is the profits from pricing under full attention. Evaluating (76) at $\boldsymbol{m}_H = \boldsymbol{\iota}$ yields

$$\begin{aligned}
& \mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota}), \hat{\boldsymbol{\mu}}_{Ht}, \boldsymbol{\iota}\right) \\
\simeq\ & v_{Hi}^0 + \frac{1}{2}\frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2}\mathbb{E}[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota})]^2 - \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2}\mathbb{E}\left[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota})\right]^2 \\
& + \frac{1}{2}\mathbb{E}\left(\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v_{Hi}^0}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{\boldsymbol{x}}'_{Hit}}\hat{\boldsymbol{x}}_{Hit}\right) \\
=\ & v_{Hi}^0 - \frac{1}{2}\frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2}\mathbb{E}[\hat{p}_{Hit}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{\iota})]^2 + \frac{1}{2}\mathbb{E}\left(\hat{\boldsymbol{x}}'_{Hit}\frac{\partial^2 v_{Hi}^0}{\partial \hat{\boldsymbol{x}}_{Hit}\partial \hat{\boldsymbol{x}}'_{Hit}}\hat{\boldsymbol{x}}_{Hit}\right).
\end{aligned} \tag{77}$$

Third, we obtain the expected profit loss of deviating attention from full attention in pricing. Subtracting (77) from (76) yields

$$\begin{aligned}
& \mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{m}_H),\hat{\boldsymbol{x}}_{Hit},\boldsymbol{\iota}\right)-\mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Ht},\boldsymbol{\iota}),\hat{\boldsymbol{x}}_{Hit},\boldsymbol{\iota}\right)\\
\simeq\ & \frac{1}{2}\frac{\partial^2 v_{Hi}^0}{\partial\hat{p}_{it}(z)^2}\mathbb{E}\left[\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{m}_H)^2-2\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{m}_H)\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{\iota})+\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{\iota})^2\right]\\
=\ & \frac{1}{2}\frac{\partial^2 v_{Hi}^0}{\partial\hat{p}_{it}(z)^2}\mathbb{E}\left[\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{m}_H)-\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{\iota})\right]^2. \qquad (78)
\end{aligned}$$

Using (17), we rewrite the above equation as

$$\begin{aligned}
& \mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit},\boldsymbol{m}_H),\hat{\boldsymbol{x}}_{Hit},\boldsymbol{\iota}\right)-\mathbb{E}v_{Hi}\left(\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Ht},\boldsymbol{\iota}),\hat{\boldsymbol{x}}_{Hit},\boldsymbol{\iota}\right)\\
=\ & \frac{1}{2}\frac{\partial^2 v_{Hi}^0}{\partial\hat{p}_{it}(z)^2}\left[(1-m_{1H})^2\mathbb{E}(\hat{\mu}_{Ht})^2+(1-m_{2H})^2(1-\lambda\delta)^2\mathbb{E}(\hat{a}_{it})^2-2(1-m_{1H})(1-m_{2H})\mathbb{E}(\hat{\mu}_{Ht}\hat{a}_{it})\right]\\
=\ & -\frac{1}{2}\begin{bmatrix}1-m_{1H}\\ 1-m_{2H}\end{bmatrix}'\begin{bmatrix}\Lambda_{1H} & 0\\ 0 & \Lambda_{2H}\end{bmatrix}\begin{bmatrix}1-m_{1H}\\ 1-m_{2H}\end{bmatrix}\\
=\ & -\frac{1}{2}(\boldsymbol{\iota}-\boldsymbol{m}_H)'\boldsymbol{\Lambda}_H(\boldsymbol{\iota}-\boldsymbol{m}_H), \qquad (79)
\end{aligned}$$

where $\hat{\mu}_{Ht}=\hat{w}_t$ and $\boldsymbol{\Lambda}_H$ is the diagonal matrix with the diagonal elements:

$$\Lambda_{1H} = -\frac{\partial^2 v_{Hi}^0}{\partial\hat{p}_{it}(z)^2}\mathrm{Var}(\hat{\mu}_{Ht}), \qquad (80)$$

$$\Lambda_{2H} = -\frac{\partial^2 v_{Hi}^0}{\partial\hat{p}_{it}(z)^2}(1-\lambda\delta)^2\mathrm{Var}(\hat{a}_{it}). \qquad (81)$$

Note that the nondiagonal element is zero because macroeconomic variables are independent of idiosyncratic productivity shock (i.e., $\mathbb{E}(\hat{\mu}_{Ht}\hat{a}_{it})=0$).

Although firms can reduce the profit loss (79) by approaching $\boldsymbol{m}_H$ to $\boldsymbol{\iota}$, they also have to pay costs of increasing attention, which we specify as a quadratic cost function $\mathcal{C}(\boldsymbol{m}_H)=(\kappa_1/2)m_{1H}^2+(\kappa_2/2)m_{2H}^2=(1/2)\boldsymbol{m}_H'\boldsymbol{\kappa}\boldsymbol{m}_H$. Formally, the US firms' choice of attention for selling their brands in US markets is characterized by

$$\min_{\boldsymbol{m}_H\in[0,1]^2}\frac{1}{2}(\boldsymbol{\iota}-\boldsymbol{m}_H)'\boldsymbol{\Lambda}_H(\boldsymbol{\iota}-\boldsymbol{m}_H)+\frac{1}{2}\boldsymbol{m}_H'\boldsymbol{\kappa}\boldsymbol{m}_H.$$

The remaining sparse max can analogously be defined. The sparse max of US firms for selling their brands in Canadian markets is

$$\min_{\boldsymbol{m}_H^*\in[0,1]^2}\frac{1}{2}(\boldsymbol{\iota}-\boldsymbol{m}_H^*)'\boldsymbol{\Lambda}_H^*(\boldsymbol{\iota}-\boldsymbol{m}_H^*)+\frac{1}{2}\boldsymbol{m}_H^{*\prime}\boldsymbol{\kappa}\boldsymbol{m}_H^*,$$

where the diagonal matrix $\boldsymbol{\Lambda}_H^*$ with the diagonal elements:

$$\Lambda_{1H}^* = -\frac{\partial^2 v_{Hi}^{*0}}{\partial \hat{p}_{it}^*(z)^2}\text{Var}(\hat{\mu}_{Ht}^*), \qquad \Lambda_{2H}^* = -\frac{\partial^2 v_{Hi}^{*0}}{\partial \hat{p}_{it}^*(z)^2}(1-\lambda\delta)^2\text{Var}\left(\hat{a}_{it}\right).$$

Next, the sparse max of Canadian firms for selling their brands in Canadian markets is

$$\min_{\boldsymbol{m}_F^*\in[0,1]^2}\frac{1}{2}(\boldsymbol{\iota}-\boldsymbol{m}_F^*)'\boldsymbol{\Lambda}_F^*(\boldsymbol{\iota}-\boldsymbol{m}_F^*)+\frac{1}{2}\boldsymbol{m}_F^{*\prime}\boldsymbol{\kappa}\boldsymbol{m}_F^*,$$

where the diagonal matrix $\boldsymbol{\Lambda}_F^*$ with the diagonal elements:

$$\Lambda_{1F}^* = -\frac{\partial^2 v_{Fi}^{*0}}{\partial \hat{p}_{it}^*(z)^2}\text{Var}(\hat{\mu}_{Ft}^*), \qquad \Lambda_{2F}^* = -\frac{\partial^2 v_{Fi}^{*0}}{\partial \hat{p}_{it}^*(z)^2}(1-\lambda\delta)^2\text{Var}(\hat{a}_{it}^*).$$

By symmetry, we can easily reconfirm that $\boldsymbol{m}_F^* = \boldsymbol{m}_H$ because $\boldsymbol{\Lambda}_F^* = \boldsymbol{\Lambda}_H$. The sparse max of Canadian firms for selling their brands in US markets is

$$\min_{\boldsymbol{m}_F\in[0,1]^2}\frac{1}{2}(\boldsymbol{\iota}-\boldsymbol{m}_F)'\boldsymbol{\Lambda}_F(\boldsymbol{\iota}-\boldsymbol{m}_F)+\frac{1}{2}\boldsymbol{m}_F'\boldsymbol{\kappa}\boldsymbol{m}_F,$$

where the diagonal matrix $\boldsymbol{\Lambda}_F$ with the diagonal elements:

$$\Lambda_{1F} = -\frac{\partial^2 v_{Fi}^{0}}{\partial \hat{p}_{it}(z)^2}\text{Var}(\hat{\mu}_{Ft}), \qquad \Lambda_{2F} = -\frac{\partial^2 v_{Fi}^{0}}{\partial \hat{p}_{it}(z)^2}(1-\lambda\delta)^2\text{Var}(\hat{a}_{it}^*).$$

Again, by symmetry, we have $\boldsymbol{m}_F = \boldsymbol{m}_H^*$.

A.4 Proof of Proposition 1

Combining (23) with (66), we can rewrite $\hat{q}_{it}$ as

$$\hat{q}_{it} = (\hat{p}_{it}^* - \hat{c}_t^*) - (\hat{p}_{it} - \hat{c}_t). \tag{82}$$

Equation (24) implies:

$$\begin{aligned}
\hat{p}_{it} - \hat{c}_t &= \lambda(\hat{p}_{it-1} - \pi_t) + (1-\lambda)\hat{p}_{it}^{opt} - \hat{c}_t \\
&= \lambda(\hat{p}_{it-1} - \hat{c}_{t-1}) - \lambda(\Delta\hat{c}_t + \pi_t) + (1-\lambda)(\hat{p}_{it}^{opt} - \hat{c}_t) \\
&= \lambda(\hat{p}_{it-1} - \hat{c}_{t-1}) - \lambda\varepsilon_t^M + (1-\lambda)(\hat{p}_{it}^{opt} - \hat{c}_t).
\end{aligned} \tag{83}$$

Note that $\Delta\hat{c}_t + \pi_t = \varepsilon_t^M$ holds due to the CIA constraint and the money supply process in (4). Similarly, $\hat{p}_{it}^* - \hat{c}_t^*$ is given by

$$\hat{p}_{it}^* - \hat{c}_t^* = \lambda(\hat{p}_{it-1}^* - \hat{c}_{t-1}^*) - \lambda\varepsilon_t^{M^*} + (1-\lambda)(\hat{p}_{it}^{opt*} - \hat{c}_t^*). \tag{84}$$

Substituting (83) and (84) into (82) yields the following expression for $\hat{q}_{it}$:

$$\hat{q}_{it} = \lambda\hat{q}_{it-1} + \lambda\varepsilon_t^n + (1-\lambda)\,\hat{q}_{it}^{opt}, \tag{85}$$

where $\varepsilon_t^n = \varepsilon_t^M - \varepsilon_t^{M^*}$, and $\hat{q}_{it}^{opt}$, the log deviation of the real reset exchange rate, is given by

$$\hat{q}_{it}^{opt} = (\hat{p}_{it}^{opt*} - \hat{c}_t^*) - (\hat{p}_{it}^{opt} - \hat{c}_t). \tag{86}$$

Note that we can recover $\hat{q}_{it}^{opt}$ from (66):

$$\hat{q}_{it}^{opt} = \hat{q}_t + \hat{p}_{it}^{opt*} - \hat{p}_{it}^{opt}, \tag{87}$$

and

$$q_{it}^{opt} = \frac{S_t P_{it}^{opt*}}{P_{it}^{opt}}. \tag{88}$$

Let us focus on $\hat{p}_{it}^{opt} - \hat{c}_t$ and $\hat{p}_{it}^{opt*} - \hat{c}_t^*$. Starting from (25), we combine (17), (71), and (64) to derive:

$$\begin{aligned}\hat{p}_{it}^{opt} - \hat{c}_t &= \omega\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) + (1-\omega)\,\hat{p}_{Fi}\,(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F) - \hat{c}_t \\ &= -\,(1-m)\,\hat{c}_t - (1-\lambda\delta)\left[\omega m_{2H}\hat{a}_{it} + (1-\omega)\,m_{2F}\hat{a}_{it}^*\right].\end{aligned} \tag{89}$$

Similarly, $\hat{p}_{it}^{opt*} - \hat{c}_t^*$ is

$$\hat{p}_{it}^{opt*} - \hat{c}_t^* = -(1-m)\hat{c}_t^* - (1-\lambda\delta)[\omega m_{2H}\hat{a}_{it}^* + (1-\omega)m_{2F}\hat{a}_{it}], \tag{90}$$

where we used the assumption of a symmetry between the US and Canada: $m = \omega m_{1F}^* + (1-\omega)m_{1H}^* = \omega m_{1H} + (1-\omega)m_{1F}$. Combining (86), (89), and (90), we have

$$\hat{q}_{it}^{opt} = (1-m)(\hat{c}_t - \hat{c}_t^*) + (1-\lambda\delta)(\omega m_{2H} + (\omega-1)m_{2F})(\hat{a}_{it} - \hat{a}_{it}^*).$$

Using (66), (6), and (7), we can show that the real reset exchange rate depends on the aggregate RER:

$$\hat{q}_{it}^{opt} = (1-m)\hat{q}_t + (1-\lambda\delta)\psi\varepsilon_{it}^r, \tag{91}$$

where $\varepsilon_{it}^r = \varepsilon_{it}^a - \varepsilon_t^{a^*}$, and $\psi = \omega m_{2H} - (1-\omega)m_{2F}$.

Substituting (91) into (85) yields

$$\hat{q}_{it} = \lambda\hat{q}_{it-1} + (1-\lambda)(1-m)\hat{q}_t + \lambda\varepsilon_t^n + (1-\lambda)(1-\lambda\delta)\psi\varepsilon_{it}^r. \tag{92}$$

Here, $\hat{q}_{it} = \ln q_{it}$ and $\hat{q}_t = \ln q_t$ because $\ln\bar{q}_i = \ln\bar{q} = 0$ from the symmetry between the two countries. In particular, the symmetry ensures that $\ln\bar{q} = \ln\bar{c} - \ln\bar{c}^* = 0$ and that $\ln\bar{q}_i = \ln\bar{q} + \ln\bar{p}_i^* - \ln\bar{p}_i = 0$. Therefore, (92) is equivalent to (26) in Proposition 1.

A.5 Derivation of (30)–(32)

In deriving (30)–(32), we assume that $\hat{a}_{it} = \hat{a}_{it}^* = 0$ for all t. The log-linearized optimal reset prices derived in Appendix A.2 become $\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) = m_{1H}\hat{w}_t$, $\hat{p}_{Hi}^*(\hat{\boldsymbol{x}}_{Hit}^*, \boldsymbol{m}_H^*) = m_{1H}^*\hat{w}_t^*$, $\hat{p}_{Fi}^*(\hat{\boldsymbol{x}}_{Fit}^*, \boldsymbol{m}_F^*) = m_{1F}^*\hat{w}_t^*$, and $\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F) = m_{1F}\hat{w}_t$, respectively.

Using (25), $\hat{p}_{it}^{opt}$ becomes

$$\hat{p}_{it}^{opt} = \omega m_{1H}\hat{w}_t + (1-\omega)m_{1F}\hat{w}_t = m\hat{w}_t, \tag{93}$$

where we use the definition of $m = \omega m_{1H} + (1-\omega)m_{1F}$. Similarly, $\hat{p}_{it}^{opt*}$ becomes

$$\hat{p}_{it}^{opt*} = \omega m_{1F}^*\hat{w}_t^* + (1-\omega)m_{1H}^*\hat{w}_t^* = m\hat{w}_t^*, \tag{94}$$

where we use the assumption of a symmetry, $m_{1F}^* = m_{1H}$ and $m_{1H}^* = m_{1F}$.

From the definition of the log deviation from the steady state, $\hat{w}_t = \ln w_t - \ln\bar{w} = \ln W_t - \ln P_t - \ln\bar{w}$. Likewise, $\hat{p}_{it}^{opt} = \ln P_{it}^{opt} - \ln P_t - \ln\bar{p}_i$ and $\hat{p}_{it}^{opt*} = \ln P_{it}^{opt*} - \ln P_t^* - \ln\bar{p}_i^*$. Thus, we can rewrite the above equations as

$$\ln P_{it}^{opt} = (1-m)\ln P_t + m\ln W_t + \ln\bar{p}_i - m\ln\bar{w}, \tag{95}$$

$$\ln P_{it}^{opt*} = (1-m)\ln P_t^* + m\ln W_t^* + \ln\bar{p}_i^* - m\ln\bar{w}^*. \tag{96}$$

If we suppress the constant terms, these equations correspond to the nominal reset prices for good i, namely, (30) and (31).

Regarding (32), focus on (86) derived in Appendix A.4. From (93) and (94), we have

$$\begin{aligned}\hat{q}_{it}^{opt} &= (\hat{p}_{it}^{opt*} - \hat{c}_t^*) - (\hat{p}_{it}^{opt} - \hat{c}_t) \\ &= (m\hat{w}_t^* - \hat{c}_t^*) - (m\hat{w}_t - \hat{c}_t) \\ &= (1-m)(\hat{c}_t - \hat{c}_t^*) \\ &= (1-m)\hat{q}_t.\end{aligned}$$

Here we again use (64) and its foreign analogue, as well as (66). Given that $\ln \bar{q} = \ln \bar{q}^{opt} = 0$ by symmetry, we have $\hat{q}_{it}^{opt} = \ln q_{it}^{opt}$ as well as $\hat{q}_t = \ln q_t$. Thus, we derive (32): $\ln q_{it}^{opt} = (1-m)\ln q_t$.

A.6 Descriptive statistics

Table A.1 reports descriptive statistics for the variables used in the regressions. The upper panel presents statistics for $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t$ constructed using the common λ (i.e., $\lambda = 0.34$), while the lower panel shows statistics for $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t^i$ constructed using λ_i. A comparison between the upper and lower panels illustrates how heterogeneity in price stickiness affects the variables used in the regressions. By comparing the left and right panels, we can also contrast the data for the US–Canadian city pairs with those for the UK–Euro area city pairs.

We first confirm that variations in λ_i have only a minor effect on the variability of $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t^i$. Although the lower panel potentially introduces greater variability due to heterogeneity in λ_i, the standard deviations are not substantially different from those in the upper panel. This suggests that the observed variability primarily arises from variations in the RERs themselves.

Next, we observe that the descriptive statistics for the RERs are broadly similar between the US–Canadian and UK–Euro area city pairs. The mean LOP deviations are comparable, typically ranging between 3 and 5 percent in absolute value, and the standard deviations lie between 35 and 38 percent. While the mean PPP deviations are less informative due to their dependence on the choice of base year, the table shows that the standard deviations of the aggregate RERs (7-11 percent) are much smaller than those of the good-level RERs (35-38 percent).

A.7 The model with general CRRA preferences

In the baseline model, we assumed that household preferences are given by $U(c,n) = \ln c - \chi n$. In this appendix, we consider more general CRRA preferences, $U(c,n) = c^{1-\sigma}/(1-\sigma) -$

$\chi n^{1+\varphi}/(1+\varphi)$, where $\sigma > 0$ and $\varphi > 0$. We modify the first-order conditions of households to allow for the degree of relative risk aversion. Under $\sigma > 0$, the first-order conditions imply that $S_t = (M_t/M_t^*)^\sigma (P_t/P_t^*)^{1-\sigma}$.

We maintain the assumption that the NER follows a random walk. To ensure this, we replace (4) and (5) with the following assumption on the money growth rate:

$$\Delta \ln M_t = \frac{\sigma-1}{\sigma}\pi_t + \frac{1}{\sigma}\varepsilon_t^M, \quad (97)$$

$$\Delta \ln M_t^* = \frac{\sigma-1}{\sigma}\pi_t^* + \frac{1}{\sigma}\varepsilon_t^{M^*}. \quad (98)$$

Here, the NER growth is given by: $\Delta \ln S_t = \sigma(\Delta \ln M_t - \Delta \ln M_t^*) + (1-\sigma)(\pi_t - \pi_t^*)$. Substituting (97) and (98) into this equation yields $\Delta \ln S_t = \varepsilon_t^n$.[53]

Using the CIA constraints, we rewrite (97) and (98) as:

$$\sigma \Delta \hat{c}_{t+k} + \pi_{t+k} = \varepsilon_{t+k}^M, \quad (99)$$

$$\sigma \Delta \hat{c}_{t+k}^* + \pi_{t+k}^* = \varepsilon_{t+k}^{M^*}, \quad (100)$$

for $k \geq 0$. We later use (99) and (100) in deriving the estimation equation.

A.7.1 The derivation of the estimation equation

To derive the estimation equation, we follow the procedure used in Appendix A.4. Since σ is no longer equal to one, the international risk-sharing condition (3) is replaced by $q_t = (c_t/c_t^*)^\sigma$ so that $\hat{q}_t = \sigma(\hat{c}_t - \hat{c}_t^*)$. Then, (23) and (87) imply

$$\hat{q}_{it} = (\hat{p}_{it}^* - \sigma \hat{c}_t^*) - (\hat{p}_{it} - \sigma \hat{c}_t), \quad (101)$$

$$\hat{q}_{it}^{opt} = (\hat{p}_{it}^{opt*} - \sigma \hat{c}_t^*) - (\hat{p}_{it}^{opt} - \sigma \hat{c}_t), \quad (102)$$

respectively.

Note that (24) remains valid even under general CRRA preferences. Equation (24) implies:

$$\hat{p}_{it} - \sigma \hat{c}_t = \lambda(\hat{p}_{it-1} - \sigma \hat{c}_{t-1}) - \lambda \varepsilon_t^M + (1-\lambda)\left(\hat{p}_{it}^{opt} - \sigma \hat{c}_t\right), \quad (103)$$

[53] If we instead assume that the money supply follows a random walk, then $\Delta \ln S_t = \sigma \varepsilon_t^n + (1-\sigma)(\pi_t - \pi_t^*)$, implying that $\pi_t - \pi_t^*$ helps forecast $\Delta \ln S_t$. This is inconsistent with the exchange-rate disconnect puzzle.

where we replace $\sigma\Delta\hat{c}_t + \pi_t$ by ε_t^M using (99). Similarly, $\hat{p}_{it}^* - \sigma\hat{c}_t^*$ is given by

$$\hat{p}_{it}^* - \sigma\hat{c}_t^* = \lambda(\hat{p}_{it-1}^* - \sigma\hat{c}_{t-1}^*) - \lambda\varepsilon_t^{M^*} + (1-\lambda)(\hat{p}_{it}^{opt*} - \sigma\hat{c}_t^*). \tag{104}$$

Substituting (103) and (104) into (101) implies that (85) remains valid:

$$\hat{q}_{it} = \lambda\hat{q}_{it-1} + \lambda\varepsilon_t^n + (1-\lambda)\hat{q}_{it}^{opt}. \tag{105}$$

Next, we recalculate the log deviation of the optimal reset prices under general CRRA preferences. Even in this case, (63) continues to hold, but (64) does not. The log deviation of real wages is now given by $\hat{w}_t = \sigma\hat{c}_t + \varphi\hat{n}_t$. Accordingly, we rewrite (63) as

$$\begin{aligned}
\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) &= m_{1H}(\sigma\hat{c}_t + \varphi\hat{n}_t) - m_{2H}\hat{a}_{it} \\
&\quad + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k \left[m_{1H}\left(\sigma\Delta\hat{c}_{t+k} + \pi_{t+k} + \varphi\Delta\hat{n}_{t+k}\right) - m_{2H}\Delta\hat{a}_{it+k} \right] \\
&= m_{1H}\sigma\hat{c}_t - m_{2H}(1-\lambda\delta)\hat{a}_{it} + m_{1H}\varphi\left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}}\right)\hat{n}_t, \qquad (106)
\end{aligned}$$

where L is the lag operator. In the second equality, we used (99) to replace $\sigma\Delta\hat{c}_{t+k} + \pi_{t+k}$ by ε_{t+k}^M. Equation (106) differs from (17) because of the presence of the forward-looking terms for the labor supply. Likewise, equations for the optimal reset prices (69), (70), and (71) must be replaced by

$$\begin{aligned}
\hat{p}_{Hi}^*(\hat{\boldsymbol{x}}_{Hit}^*, \boldsymbol{m}_H^*) &= m_{1H}^*\sigma\hat{c}_t^* - m_{2H}(1-\lambda\delta)\hat{a}_{it} + m_{1H}^*\varphi\left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}}\right)\hat{n}_t, \qquad (107) \\
\hat{p}_{Fi}^*(\hat{\boldsymbol{x}}_{Fit}^*, \boldsymbol{m}_F^*) &= m_{1F}^*\sigma\hat{c}_t^* - m_{2F}^*(1-\lambda\delta)\hat{a}_{it}^* + m_{1F}^*\varphi\left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}}\right)\hat{n}_t^*, \qquad (108) \\
\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F) &= m_{1F}\sigma\hat{c}_t - m_{2F}(1-\lambda\delta)\hat{a}_{it}^* + m_{1F}\varphi\left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}}\right)\hat{n}_t^*, \qquad (109)
\end{aligned}$$

respectively.

We then compute $\hat{p}_{it}^{opt} - \sigma\hat{c}_t$ and $\hat{p}_{it}^{opt*} - \sigma\hat{c}_t^*$ in (102). Using (25), $\hat{p}_{it}^{opt} - \sigma\hat{c}_t$ is given by

$$\begin{aligned}
\hat{p}_{it}^{opt} - \sigma\hat{c}_t &= -(1-m)\sigma\hat{c}_t - (1-\lambda\delta)[\omega m_{2H}\hat{a}_{it} + (1-\omega)m_{2F}\hat{a}_{it}^*] \\
&\quad + \varphi\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}}\left[\omega m_{1H}\hat{n}_t + (1-\omega)m_{1F}\hat{n}_t^*\right], \qquad (110)
\end{aligned}$$

Likewise,

$$\begin{aligned}\hat{p}_{it}^{opt*} - \sigma \hat{c}_t^* &= -(1-m)\sigma \hat{c}_t^* - (1-\lambda\delta)[\omega m_{2H}\hat{a}_{it}^* + (1-\omega)m_{2F}\hat{a}_{it}] \\ &\quad + \varphi \frac{1-\lambda\delta}{1-\lambda\delta L^{-1}} \left[\omega m_{1H}\hat{n}_t^* + (1-\omega)m_{1F}\hat{n}_t\right], \end{aligned} \tag{111}$$

where we used $\boldsymbol{m}_F^* = \boldsymbol{m}_H$ and $\boldsymbol{m}_H^* = \boldsymbol{m}_F$.

Plug (110) and (111) into (102). Then, (105) becomes

$$\begin{aligned}\hat{q}_{it} = \lambda \hat{q}_{it-1} + (1-\lambda)(1-m)\hat{q}_t + \lambda\varepsilon_t^n + (1-\lambda)(1-\lambda\delta)\psi_2\varepsilon_{it}^r \\ - \varphi\psi_1 \frac{(1-\lambda)(1-\lambda\delta)}{1-\lambda\delta L^{-1}}(\hat{n}_t - \hat{n}_t^*), \end{aligned} \tag{112}$$

where $\psi_j = \omega m_{jH} - (1-\omega)m_{jF}$ for $j = 1, 2$. Note that the subscript $j = 2$ is newly applied to $\psi = \omega m_{2H} - (1-\omega)m_{2F}$ to distinguish it from $\psi_1 = \omega m_{1H} - (1-\omega)m_{1F}$. Equation (112) differs from (92) because the former includes forward-looking terms for labor supply. If $\varphi = 0$, these forward-looking terms disappear, and the equation reduces to (92).

As discussed in the proof of Proposition 1, a symmetry between the two countries implies that $\hat{q}_{it} = \ln q_{it}$ and $\hat{q}_t = \ln q_t$. Likewise, $\bar{n} = \bar{n}^*$ holds by symmetry, which leads to $\hat{n}_t - \hat{n}_t^* = \ln n_t - \ln n_t^*$. Consequently, we obtain a dynamic equation generalizing (26):

$$\begin{aligned}\ln q_{it} = \lambda \ln q_{it-1} &+ (1-\lambda)(1-m)\ln q_t + \lambda\varepsilon_t^n + (1-\lambda)(1-\lambda\delta)\psi_2\varepsilon_{it}^r \\ &-\varphi\psi_1 \frac{(1-\lambda)(1-\lambda\delta)}{1-\lambda\delta L^{-1}}(\ln n_t - \ln n_t^*). \end{aligned} \tag{113}$$

To derive the estimation equation under general CRRA preferences, rewrite (113) using the definitions of $\tilde{q}_{it}$ and $\tilde{q}_t$:

$$\ln \tilde{q}_{it} = (1-m)\ln \tilde{q}_t + (1-\lambda)(1-\lambda\delta)\psi_2\varepsilon_{it}^r - \varphi\psi_1 \frac{(1-\lambda)(1-\lambda\delta)}{1-\lambda\delta L^{-1}}(\ln n_t - \ln n_t^*),$$

or equivalently,

$$\begin{aligned}\ln \tilde{q}_{it} - \lambda\delta\mathbb{E}_t \ln \tilde{q}_{it+1} &= (1-m)(\ln \tilde{q}_t - \lambda\delta\mathbb{E}_t \ln \tilde{q}_{t+1}) \\ &\quad -(1-\lambda)(1-\lambda\delta)\varphi\psi_1(\ln n_t - \ln n_t^*) + (1-\lambda)(1-\lambda\delta)\psi_2\varepsilon_{it}^r. \end{aligned} \tag{114}$$

Let $\ln \tilde{\tilde{q}}_{it} = \ln \tilde{q}_{it} - \lambda\delta \ln \tilde{q}_{it+1}$ and $\ln \tilde{\tilde{q}}_t = \ln \tilde{q}_t - \lambda\delta \ln \tilde{q}_{t+1}$. Our estimation equation is

$$\ln \tilde{\tilde{q}}_{it} = \alpha + \beta \ln \tilde{\tilde{q}}_t + \gamma' X_{it} + u_{it}, \tag{115}$$

where X_{it} includes the log difference in labor supply, $\ln n_t - \ln n_t^*$. Note that OLS is no longer valid because u_{it} now includes the forecast errors $\ln \tilde{q}_{it+1} - \mathbb{E}_t \ln \tilde{q}_{it+1}$ and $\ln \tilde{q}_{t+1} - \mathbb{E}_t \ln \tilde{q}_{t+1}$. We therefore use an instrumental variables. For the data source of $\ln n_t - \ln n_t^*$, we use the indices of total hours worked , with the year 2010 as the base year, from *OECD.Stat*.

A.7.2 Estimation results under general CRRA preferences

Table A.2 reports the estimation results of (115). The left panel presents the results for the US–Canadian city pairs, while the right panel shows the results for the UK–Euro area city pairs. In both panels, we assume homogeneity in price stickiness (i.e., a common λ) in specifications (1) and (2) and heterogeneity in price stickiness in specifications (3) and (4). Specifications (2) and (4) include city-pair-specific fixed effects as additional explanatory variables. In all specifications, we instrument $\ln \tilde{\tilde{q}}_t$ with $\ln \tilde{q}_{t-1}$. In all cases, the null hypothesis of aggregate RER independence, namely $\beta = 0$, is strongly rejected. The estimated values of m are well below one, suggesting robustness to changes in the preference assumptions.

A.8 Estimation results based on dynamic panel

Table A.3 reports the estimation results based on the dynamic panel. In (38), we regress $\ln q_{ijt}$ directly on $\ln q_t$, along with $\ln q_{ijt-1}$ and $\Delta \ln S_t$ as additional regressors. Unlike the case of (36), the presence of a lagged dependent variable as a regressor implies a dynamic panel structure. Therefore, dynamic panel estimators, such as the generalized method of moments estimator by Arellano and Bond (1991), must be used instead of OLS (see, e.g., Crucini and Shintani, 2008; Crucini et al., 2010a). The left panel of the table presents the estimation results for the US–Canadian city pairs, whereas the right panel shows those for the UK–Euro area city pairs. In specifications (2) and (4), we impose the restriction that the coefficients on $\ln q_{it-1}$ and $\Delta \ln S_t$ are equal, as these control variables are assumed to have the same coefficient. This restriction follows from (26), which indicates that both $\ln q_{it-1}$ and $\varepsilon_t^n = \Delta \ln S_t$ share the same coefficient. Specifications (3) and (4) differ from (1) and (2) in that the regressions include η_t^r as an additional control variable.

The table shows that, in all regressions, the null hypothesis of aggregate RER independence is strongly rejected. In addition, the estimates of β are consistently positive, supporting our theoretical model.

A.9 The effect of distance on the estimation results

In this appendix, we examine the impact of distance between two cross-border cities on the estimation results. Table A.4 reports the estimated coefficients on $\ln \tilde{q}_t$ under a common λ. The left panel of the table presents the estimation results for the US–Canadian city pairs, whereas the right panel shows those for the UK–Euro area city pairs. All specifications include good-specific fixed effects. We also consider specifications that include city-pair fixed effects and differences in the log of real GDP per hour worked. However, our estimation results are robust to the inclusion of these additional explanatory variables.

Table A.4 presents four specifications. Specification (1) uses city pairs with below-median distances (2,536 kilometers for US–Canadian city pairs, 919 kilometers for UK–Euro area city pairs), while (2) uses those above the median. Specification (3) excludes city pairs below the 5th percentile (361 kilometers and 321 kilometers, respectively) to address potential outliers near borders. Specification (4) excludes city pairs above the 95th percentile (4,516 kilometers and 1,798 kilometers, respectively).

In all specifications, the results are robust to splitting the samples. That is, the data strongly reject the null hypothesis of $\beta = 0$, indicating aggregate RER dependence. The estimated values of m range from 0.123 to 0.184, which are very close to the estimates in specification (1) of Table 1: $\hat{m} = 0.156$ for the US–Canadian city pairs and $\hat{m} = 0.144$ for the UK–Euro area city pairs. Furthermore, as Table A.5 confirms, the estimation results remain robust to these changes in specification even when we allow for heterogeneity in price stickiness.

A.10 Persistence of the good-level RER under behavioral inattention

We first prove Proposition 3 and then derive (43) and (44).

A.10.1 Proof of Proposition 3

As a preliminary step, we rewrite the dynamic equations, (26) and (39), in terms of log deviations:

$$\hat{q}_{it} = \lambda \hat{q}_{it-1} + \theta \hat{q}_t + \lambda \varepsilon_t^n + \tilde{\psi} \varepsilon_{it}^r, \tag{116}$$

$$\hat{q}_t = \rho_q \hat{q}_{t-1} + \rho_q \varepsilon_t^n, \tag{117}$$

where $\theta = (1-m)(1-\lambda)$, $\tilde{\psi} = (1-\lambda)(1-\lambda\delta)\psi$, and $\rho_q = \lambda/(1-\theta)$. The variance of $\hat{q}_t$ is given by $\sigma_q^2 = [\rho_q^2/(1-\rho_q^2)]\sigma_n^2$, so

$$\sigma_n^2 = \frac{1-\rho_q^2}{\rho_q^2}\sigma_q^2. \tag{118}$$

The two covariances $\mathbb{E}(\hat{q}_t\hat{q}_{it})$ and $\mathbb{E}(\hat{q}_t\hat{q}_{it-1})$ can be expressed as

$$\mathbb{E}(\hat{q}_t\hat{q}_{it}) = \lambda\mathbb{E}(\hat{q}_t\hat{q}_{it-1}) + \theta\sigma_q^2 + \lambda\rho_q\sigma_n^2, \tag{119}$$

$$\mathbb{E}(\hat{q}_t\hat{q}_{it-1}) = \rho_q\mathbb{E}(\hat{q}_{t-1}\hat{q}_{it-1}). \tag{120}$$

To derive the above equations, we used (117) and (116), along with $\mathbb{E}[\hat{q}_t\varepsilon_{it}^r] = 0$ and $\mathbb{E}[\hat{q}_{it-1}\varepsilon_t^n] = 0$. Substituting (118) and (120) into (119) yields

$$(1-\lambda\rho_q)\mathbb{E}(\hat{q}_t\hat{q}_{it}) = \left[\theta + \frac{\lambda(1-\rho_q^2)}{\rho_q}\right]\sigma_q^2, \tag{121}$$

where the stationarity of RERs implies $\mathbb{E}(\hat{q}_t\hat{q}_{it}) = \mathbb{E}(\hat{q}_{t-1}\hat{q}_{it-1})$. Note that the expression inside the brackets can be simplified using the definition of ρ_q:[54]

$$\theta + \frac{\lambda(1-\rho_q^2)}{\rho_q} = 1 - \lambda\rho_q. \tag{122}$$

Using (122), (121) and (120) become

$$\mathbb{E}[\hat{q}_t\hat{q}_{it}] = \sigma_q^2, \tag{123}$$

$$\mathbb{E}[\hat{q}_t\hat{q}_{it-1}] = \rho_q\sigma_q^2, \tag{124}$$

respectively.

We next examine the variance $\sigma_{qi}^2 = \mathbb{E}(\hat{q}_{it}^2)$ and the autocovariance $\gamma_1 = \mathbb{E}(\hat{q}_{it}\hat{q}_{it-1})$. Using (116) and (123), we obtain

$$\begin{aligned}
\sigma_{qi}^2 &= \mathbb{E}(\hat{q}_{it}^2) \\
&= \lambda\mathbb{E}(\hat{q}_{it}\hat{q}_{it-1}) + \theta\mathbb{E}(\hat{q}_{it}\hat{q}_t) + \lambda\mathbb{E}(\hat{q}_{it}\varepsilon_t^n) + \tilde{\psi}\mathbb{E}(\hat{q}_{it}\varepsilon_{it}^r) \\
&= \lambda\gamma_1 + \theta\sigma_q^2 + \lambda\mathbb{E}(\hat{q}_{it}\varepsilon_t^n) + \tilde{\psi}\mathbb{E}(\hat{q}_{it}\varepsilon_{it}^r).
\end{aligned}$$

[54]To see this, $\theta + \lambda(1-\rho_q^2)/\rho_q = \theta + \lambda(1-\rho_q^2)/(\lambda/(1-\theta)) = \theta + (1-\rho_q^2)(1-\theta) = 1 - \rho_q^2(1-\theta)$. Applying the definition of ρ_q again yields (122).

Substituting (116) into this equation yields

$$\begin{aligned}
\sigma_{qi}^2 &= \lambda\gamma_1 + \theta\sigma_q^2 + \lambda\mathbb{E}[(\lambda\hat{q}_{it-1} + \theta\hat{q}_t + \lambda\varepsilon_t^n + \tilde{\psi}\varepsilon_{it}^r)\varepsilon_t^n] + \tilde{\psi}\mathbb{E}[(\lambda\hat{q}_{it-1} + \theta\hat{q}_t + \lambda\varepsilon_t^n + \tilde{\psi}\varepsilon_{it}^r)\varepsilon_{it}^r] \\
&= \lambda\gamma_1 + \theta\sigma_q^2 + \lambda\theta\mathbb{E}(\hat{q}_t\varepsilon_t^n) + \lambda^2\sigma_n^2 + \tilde{\psi}^2\sigma_r^2 \\
&= \lambda\gamma_1 + \theta\sigma_q^2 + \lambda(\theta\rho_q + \lambda)\sigma_n^2 + \tilde{\psi}^2\sigma_r^2 \\
&= \lambda\gamma_1 + \theta\sigma_q^2 + \lambda\rho_q\sigma_n^2 + \tilde{\psi}^2\sigma_r^2, \qquad (125)
\end{aligned}$$

where we used $\mathbb{E}(\hat{q}_{it-1}\varepsilon_t^n) = \mathbb{E}(\hat{q}_{it-1}\varepsilon_{it}^r) = 0$ and $\mathbb{E}(\hat{q}_t\varepsilon_{it}^r) = \mathbb{E}(\varepsilon_t^n\varepsilon_{it}^r) = 0$ for the second equality. For the third and the last equalities, we used (117) and the definition $\rho_q = \lambda/(1-\theta)$, respectively.

Turning to $\gamma_1 = \mathbb{E}(\hat{q}_{it}\hat{q}_{it-1})$, we again use (116) to obtain

$$\begin{aligned}
\gamma_1 &= \mathbb{E}\hat{q}_{it}\hat{q}_{it-1} \\
&= \lambda\mathbb{E}(\hat{q}_{it-1}\hat{q}_{it-1}) + \theta\mathbb{E}(\hat{q}_t\hat{q}_{it-1}) + \lambda\mathbb{E}(\varepsilon_t^n\hat{q}_{it-1}) + \tilde{\psi}\mathbb{E}(\varepsilon_{it}^r\hat{q}_{it-1}) \\
&= \lambda\sigma_{qi}^2 + \theta\rho_q\sigma_q^2. \qquad (126)
\end{aligned}$$

Here, we used $\sigma_{qi}^2 = \mathbb{E}(\hat{q}_{it-1}^2)$, (124), and $\mathbb{E}(\hat{q}_{it-1}\varepsilon_t^n) = \mathbb{E}(\hat{q}_{it-1}\varepsilon_{it}^r) = 0$ for the third equality. Using (118), (122), and (126), (125) becomes

$$\begin{aligned}
\sigma_{qi}^2 &= \lambda\gamma_1 + \theta\sigma_q^2 + \lambda\rho_q\sigma_n^2 + \tilde{\psi}^2\sigma_r^2 \\
&= \lambda\gamma_1 + \left[\theta + \frac{\lambda(1-\rho_q^2)}{\rho_q}\right]\sigma_q^2 + \tilde{\psi}^2\sigma_r^2 \\
&= \lambda\left(\lambda\sigma_{qi}^2 + \theta\rho_q\sigma_q^2\right) + \left(1 - \lambda\rho_q\right)\sigma_q^2 + \tilde{\psi}^2\sigma_r^2.
\end{aligned}$$

Arranging terms yields

$$\begin{aligned}
\sigma_{qi}^2 &= \frac{1-\lambda\rho_q(1-\theta)}{1-\lambda^2}\sigma_q^2 + \frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2 \\
&= \sigma_q^2 + \frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2, \qquad (127)
\end{aligned}$$

where we use $1 - \lambda\rho_q(1-\theta) = 1 - \lambda^2$, given the definition of $\rho_q = \lambda/(1-\theta)$. Substituting

(127) into (126), we have

$$\begin{aligned}
\gamma_1 &= \lambda\left[\sigma_q^2 + \frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2\right] + \theta\rho_q\sigma_q^2 \\
&= (\theta\rho_q + \lambda)\sigma_q^2 + \frac{\lambda\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2 \\
&= \rho_q\sigma_q^2 + \frac{\lambda\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2, \qquad (128)
\end{aligned}$$

where we again used $\theta\rho_q + \lambda = \rho_q$ from the definition of ρ_q.

Now, because the first-order autocorrelation of the good-level RER is given by $\rho_{qi} = \gamma_1/\sigma_{qi}^2$:

$$\begin{aligned}
\rho_{qi} &= \left(\frac{\sigma_q^2}{\sigma_q^2 + \frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2}\right)\rho_q + \left(\frac{\frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2}{\sigma_q^2 + \frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2}\right)\lambda \\
&= \left(\frac{1}{1+A}\right)\rho_q + \left(1 - \frac{1}{1+A}\right)\lambda \qquad (129)
\end{aligned}$$

where A is defined as

$$A = \frac{(1-\lambda)^2(1-\lambda\delta)^2\psi^2}{1-\lambda^2}\left(\frac{\sigma_r}{\sigma_q}\right)^2 \geq 0. \qquad (130)$$

Thus, ρ_{qi} can be expressed as the weighted average of ρ_q and λ:

It is now straightforward to show that $\rho_q \geq \rho_{qi}$. From Proposition 2, we know that $\rho_q \geq \lambda$. Combining this result with (129), the fact that ρ_{qi} is the weighted average of ρ_q and λ implies that $\rho_q \geq \rho_{qi} \geq \lambda$. Moreover, Proposition 2 also shows that $\rho_q = \lambda$ if $m = 1$. Therefore, $\rho_{qi} = \rho_q = \lambda$ hold when $m = 1$. Finally, (130) shows that $A = 0$ if $\psi = 0$. It then follows from (129) that $\rho_{qi} = \rho_q$ when $A = 0$.

A.10.2 Derivation of (43) and (44)

Using $\rho_q = \lambda/(1-\theta)$, eliminate λ from (129):

$$\rho_{qi} = \left(\frac{1}{1+A}\right)\rho_q + \left(1 - \frac{1}{1+A}\right)(1-\theta)\rho_q = \rho_q\left(1 - \theta\frac{A}{1+A}\right). \qquad (131)$$

Rewrite this equation using $\theta = (1-m)(1-\lambda)$. We obtain the ρ_q to ρ_{qi} ratio given by (43):

$$\frac{\rho_q}{\rho_{qi}} = \frac{1}{1-(1-m)(1-\lambda)\frac{A}{1+A}}. \qquad (132)$$

To obtain (44), we rewrite (130) as

$$
\begin{aligned}
A &= \frac{[(1-\lambda)(1-\lambda\delta)\psi]^2}{1-\lambda^2}\left(\frac{1-\rho_q^2}{\rho_q^2}\right)\left(\frac{\sigma_r}{\sigma_n}\right)^2 \\
&= \frac{[(1-\lambda)(1-\lambda\delta)]^2}{1-\lambda^2}\left\{\frac{[1-(1-m)(1-\lambda)]^2-\lambda^2}{\lambda^2}\right\}\psi^2\left(\frac{\sigma_r}{\sigma_n}\right)^2,
\end{aligned}
$$

where the first equality results from (118) and the second equality is from the definition ρ_q.

A.11 The model of roundabout production

In this section, we first describe the optimal reset prices in the model of roundabout production and derive the system of structural equations for RERs. This system is useful for deriving the dynamic equation for the good-level RER and for conducting impulse response analysis.

A.11.1 Price setting under the model of roundabout production

We consider the roundabout production with fully attentive firms. More specifically, we replace the production function $y_{it}(z) = a_{it}n_{it}(z)$ with $y_{it}(z) = a_{it}[\Gamma_{it}(z)]^r[n_{it}(z)]^{1-r}$, where $\Gamma_{it}(z)$ denotes the quantity of intermediate goods demanded by US firms. In this model, the intermediate goods purchased by each firm ($\Gamma_{it}(z)$) are composites of all goods and brands, and the price of these intermediate goods is given by P_t. Let $\tilde{\Gamma}_t$ denote the aggregate supply of intermediate goods. Then, the market-clearing condition for intermediate goods is given by $\int_{i=0}^{1}\int_{z=0}^{1}\Gamma_{it}(z)\,dz\,di = \tilde{\Gamma}_t$. The aggregate supply of intermediate goods is defined through a CES index: $\tilde{\Gamma}_t = \left[\int_{i=0}^{1}\tilde{\Gamma}_{it}^{(\varepsilon-1)/\varepsilon}di\right]^{\varepsilon/(\varepsilon-1)}$ and $\tilde{\Gamma}_{it} = \left[\int_{z=0}^{1}\tilde{\Gamma}_{it}(z)^{(\varepsilon-1)/\varepsilon}dz\right]^{\varepsilon/(\varepsilon-1)}$. The foreign analogues of the above variables and functions are defined similarly (e.g., the foreign production function is $y_{it}^*(z) = a_{it}^*[\Gamma_{it}^*(z)]^r[n_{it}^*(z)]^{1-r}$).

The market-clearing condition for each brand of each good must satisfy

$$
\begin{aligned}
y_{it}(z) &= c_{it}(z) + \tilde{\Gamma}_{it}(z) + (1+\tau)\left[c_{it}^*(z) + \tilde{\Gamma}_{it}^*(z)\right] \quad \text{for } z \in [0, 1/2], \\
y_{it}^*(z) &= (1+\tau)\left[c_{it}(z) + \tilde{\Gamma}_{it}(z)\right] + c_{it}^*(z) + \tilde{\Gamma}_{it}^*(z) \quad \text{for } z \in (1/2, 1],
\end{aligned}
$$

instead of (8) and (9).

The objective function of US firms that sell their brands in US markets is similar to (10) but now has real marginal cost given by $\tilde{r}\bar{w}^{1-r}\exp\left[(1-r)\hat{w}_{t+k} + \sum_{l=0}^{k}\pi_{t+l} - \hat{a}_{it+k}\right]$:

$$
\begin{aligned}
v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} \\
\times \frac{P_t}{P_{t+k}} \left\{ \bar{p}_i(z)\exp[\hat{p}_{it}(z)] - \tilde{r}\bar{w}^{1-r}\exp\left[(1-r)\hat{w}_{t+k} + \sum_{l=0}^{k}\pi_{t+l} - \hat{a}_{it+k}\right]\right\} c_{it,t+k}(z),
\end{aligned}
$$

where $c_{it,t+k}(z)$ is given by (11). When we assume $r=0$, the above equation reduces to (10). The log-linearized first-order condition is

$$
\frac{\hat{p}_{Hit}}{1-\lambda\delta} = \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k \left[(1-r)\hat{w}_{t+k} + \sum_{l=1}^{k}\pi_{t+l} - \hat{a}_{it+k}\right]. \tag{133}
$$

The above equation can be rewritten as follows:

$$
\begin{aligned}
\frac{\hat{p}_{Hit}}{1-\lambda\delta} =& (1-r)\hat{w}_t - \hat{a}_{it} + \mathbb{E}_t\left\{\mathbb{E}_{t+1}\sum_{k=0}^{\infty}(\lambda\delta)^{k+1}\left[(1-r)\hat{w}_{t+1+k} + \pi_{t+1} + \sum_{l=1}^{k}\pi_{t+1+l} - \hat{a}_{it+1+k}\right]\right\} \\
=& (1-r)\hat{w}_t - \hat{a}_{it} + \lambda\delta\mathbb{E}_t\left\{\mathbb{E}_{t+1}\sum_{k=0}^{\infty}(\lambda\delta)^{k}\left[(1-r)\hat{w}_{t+1+k} + \sum_{l=1}^{k}\pi_{t+1+l} - \hat{a}_{it+1+k}\right]\right\} \\
& + \frac{\lambda\delta}{1-\lambda\delta}\mathbb{E}_t\pi_{t+1} \\
=& (1-r)\hat{w}_t - \hat{a}_{it} + \frac{\lambda\delta}{1-\lambda\delta}\mathbb{E}_t\left(\hat{p}_{Hit+1} + \pi_{t+1}\right),
\end{aligned}
$$

where the third equality relies on the recursive structure of the equation: $\hat{p}_{Hit+1}/(1-\lambda\delta)= \mathbb{E}_{t+1}\sum_{k=0}^{\infty}(\lambda\delta)^k\left[(1-r)\hat{w}_{t+1+k} + \sum_{l=1}^{k}\pi_{t+1+l}\right]$. Multiply $1-\lambda\delta$ by both sides to get

$$
\hat{p}_{Hit} = (1-\lambda\delta)[(1-r)\hat{w}_t - \hat{a}_{it}] + \lambda\delta\mathbb{E}_t\left(\hat{p}_{Hit+1} + \pi_{t+1}\right). \tag{134}
$$

The objective function of US firms that sell their brands in Canadian markets is

$$
\begin{aligned}
v^*_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} q_{t+k} \\
\times \frac{P^*_t}{P^*_{t+k}} \left\{ \bar{p}^*_i(z)\exp\left[\hat{p}^*_{it}(z)\right] - (1+\tau)\tilde{r}\frac{\bar{w}^{1-r}}{\bar{q}}\exp\left[(1-r)\hat{w}_{t+k} - \hat{q}_{t+k} + \sum_{l=0}^{k}\pi^*_{t+l} - \hat{a}_{it+k}\right]\right\} c^*_{it,t+k}(z),
\end{aligned}
$$

where $c^*_{it,t+k}(z)$ is given by (14). The log-linearized optimal reset price is

$$\hat{p}^*_{Hit} = (1-\lambda\delta)[(1-r)\hat{w}_t - \hat{q}_t - \hat{a}_{it}] + \lambda\delta\mathbb{E}_t\left(\hat{p}^*_{Hit+1} + \pi^*_{t+1}\right). \tag{135}$$

The remaining optimal reset prices, namely $\hat{p}^*_{Fit}$ and $\hat{p}_{Fit}$, are analogously derived:

$$\begin{aligned} \hat{p}^*_{Fit} &= (1-\lambda\delta)[(1-r)\hat{w}^*_t - \hat{a}^*_{it}] + \lambda\delta\mathbb{E}_t(\hat{p}^*_{Fit+1} + \pi^*_{t+1}), && (136) \\ \hat{p}_{Fit} &= (1-\lambda\delta)[(1-r)\hat{w}^*_t + \hat{q}_t - \hat{a}^*_{it}] + \lambda\delta\mathbb{E}_t(\hat{p}_{Fit+1} + \pi_{t+1}). && (137) \end{aligned}$$

A.11.2 The system of structural equations for RERs

We first note that the model of roundabout production differs from the model of behavioral inattention only in firms' pricing. Therefore, many of the log-linearized equations remain valid, provided that they are derived from the definitions of RERs, the CES indexes, or the households' first-order conditions. In particular, equations (23), (24), (64), (66), (67), and (82)–(86) remain valid.

Second, using (134) and (137), we take the weighted average of the optimal reset prices:

$$\begin{aligned} \hat{p}^{opt}_{it} &= \omega\hat{p}_{Hit} + (1-\omega)\hat{p}_{Fit} \\ &= (1-\lambda\delta)\left\{(1-r)\omega\hat{w}_t + (1-r)(1-\omega)\hat{w}^*_t + (1-\omega)\hat{q}_t - [\omega\hat{a}_{it} + (1-\omega)\hat{a}^*_{it}]\right\} \\ &+ \lambda\delta\left\{\mathbb{E}_t\left[\omega\hat{p}_{Hit+1} + (1-\omega)\hat{p}_{Fit+1})\right] + \mathbb{E}_t[\omega\pi_{t+1} + (1-\omega)\pi_{t+1}]\right\} \\ &= (1-\lambda\delta)\left\{(1-r)\omega\hat{w}_t + (1-r)(1-\omega)(\hat{w}^*_t + \hat{q}_t) + (1-\omega)r\hat{q}_t - [\omega\hat{a}_{it} + (1-\omega)\hat{a}^*_{it}]\right\} \\ &+ \lambda\delta\mathbb{E}_t(\hat{p}^{opt}_{it+1} + \pi_{t+1}). \end{aligned}$$

Using the labor supply condition (64), its foreign analogue, and the international risk-sharing condition (66), this expression simplifies to

$$\hat{p}^{opt}_{it} = (1-\lambda\delta)\left\{(1-r)\hat{c}_t + (1-\omega)r\hat{q}_t - [\omega\hat{a}_{it} + (1-\omega)\hat{a}^*_{it}]\right\} + \lambda\delta\mathbb{E}_t(\hat{p}^{opt}_{it+1} + \pi_{t+1}). \tag{138}$$

Third, subtracting $\hat{c}_t$ from both sides of (138), we obtain

$$\begin{aligned} \hat{p}^{opt}_{it} - \hat{c}_t &= (1-\lambda\delta)\left\{(1-r)\hat{c}_t + (1-\omega)r\hat{q}_t - \hat{c}_t - [\omega\hat{a}_{it} + (1-\omega)\hat{a}^*_{it}]\right\} + \lambda\delta\mathbb{E}_t(\hat{p}^{opt}_{it+1} + \pi_{t+1} - \hat{c}_t) \\ &= (1-\lambda\delta)\left\{-r\hat{c}_t + (1-\omega)r\hat{c}_t - (1-\omega)r\hat{c}^*_t - [\omega\hat{a}_{it} + (1-\omega)\hat{a}^*_{it}]\right\} \\ &+ \lambda\delta\left[\mathbb{E}_t(\hat{p}^{opt}_{it+1} - \hat{c}_{t+1}) + \mathbb{E}_t(\Delta\hat{c}_{t+1} + \pi_{t+1})\right] \\ &= -(1-\lambda\delta)\left\{r\left[\omega\hat{c}_t + (1-\omega)\hat{c}^*_t\right] + [\omega\hat{a}_{it} + (1-\omega)\hat{a}^*_{it}]\right\} + \lambda\delta\mathbb{E}_t(\hat{p}^{opt}_{it+1} - \hat{c}_{t+1}), && (139) \end{aligned}$$

where we used (66), and $\mathbb{E}_t(\Delta c_{t+1}+\pi_{t+1}) = \mathbb{E}_t \Delta \ln M_{t+1} = \mathbb{E}_t \varepsilon^M_{t+1} = 0$ from the CIA constraint and (4). For the foreign counterpart $\hat{p}^{opt*}_{it} - \hat{c}^*_t$, symmetry implies:

$$\hat{p}^{opt*}_{it} - \hat{c}^*_t = -(1-\lambda\delta)\left\{r\left[\omega\hat{c}^*_t + (1-\omega)\hat{c}_t\right] + \left[\omega\hat{a}^*_{it} + (1-\omega)\hat{a}_{it}\right]\right\} + \lambda\delta\mathbb{E}_t(\hat{p}^{opt*}_{it+1} - \hat{c}^*_{t+1}). \quad (140)$$

Combining (139) and (140) with (86) yields the real reset exchange rate $\hat{q}^{opt}_{it}$:

$$\begin{aligned}
\hat{q}^{opt}_{it} &= (\hat{p}^{opt*}_{it} - \hat{c}^*_t) - (\hat{p}^{opt}_{it} - \hat{c}_t) \\
&= (1-\lambda\delta)\left\{r\left[\omega\hat{q}_t - (1-\omega)\hat{q}_t\right] + \omega\hat{a}_{it} - (1-\omega)\hat{a}_{it} - \omega\hat{a}^*_{it} + (1-\omega)\hat{a}^*_{it}\right\} + \lambda\delta\mathbb{E}_t(\hat{q}_{it+1}) \\
&= (1-\lambda\delta)\psi^{RP}(r\hat{q}_t + \varepsilon^r_{it}) + \lambda\delta\mathbb{E}_t(\hat{q}^{opt}_{it+1}), \quad (141)
\end{aligned}$$

where $\psi^{RP} = 2\omega - 1$. Note that the real reset exchange rate depends on the aggregate RER, as in the model of behavioral inattention.

Finally, we derive the system of equations for the good-level and aggregate RERs. Define the log deviation of the real reset exchange rate at the aggregate level as $\hat{q}^{opt}_t = \int_{i=0}^1 \hat{q}^{opt}_{it} di = \hat{q}_t + \int_{i=0}^1 \hat{p}^{opt*}_{it} di - \int_{i=0}^1 \hat{p}^{opt}_{it} di$. Noting that $\int_{i=0}^1 \varepsilon^r_{it} di = 0$ and $\int_{i=0}^1 \hat{q}_{it} di = \hat{q}_t$, we obtain the following system of equations:

$$\begin{aligned}
\hat{q}_{it} &= \lambda\hat{q}_{it-1} + \lambda\varepsilon^n_t + (1-\lambda)\hat{q}^{opt}_{it}, \quad (142) \\
\hat{q}^{opt}_{it} &= (1-\lambda\delta)\psi^{RP}(r\hat{q}_t + \varepsilon^r_{it}) + \lambda\delta\mathbb{E}_t(\hat{q}^{opt}_{it+1}), \quad (143) \\
\hat{q}_t &= \lambda\hat{q}_{t-1} + \lambda\varepsilon^n_t + (1-\lambda)\hat{q}^{opt}_t, \quad (144) \\
\hat{q}^{opt}_t &= (1-\lambda\delta)\psi^{RP} r\hat{q}_t + \lambda\delta\mathbb{E}_t(\hat{q}^{opt}_{t+1}), \quad (145)
\end{aligned}$$

where (142) and (143) restate (85) and (141), respectively, and (144) and (145) result from aggregating (142) and (143) over i, respectively.

A.11.3 The dynamic equation for the good-level RER

To derive the dynamic equation for the good-level RER, we apply the method of undetermined coefficients to the system of equations for RERs.

We begin with the good-level RER in the system. Solving (143) forward yields

$$\hat{q}^{opt}_{it} = \frac{1}{1-\lambda\delta L^{-1}}(1-\lambda\delta)\psi^{RP}(r\hat{q}_t + \varepsilon^r_{it}).$$

Plugging this equation into (142) gives

$$\begin{aligned}\hat{q}_{it} &= \lambda\hat{q}_{it-1} + \lambda\varepsilon_t^n + \frac{(1-\lambda)(1-\lambda\delta)\psi^{RP}}{1-\lambda\delta L^{-1}}(r\hat{q}_t + \varepsilon_{it}^r) \\ &= \lambda\hat{q}_{it-1} + \lambda\varepsilon_t^n + (1-\lambda)(1-\lambda\delta)\psi^{RP}\sum_{k=0}^{\infty}(\lambda\delta)^k[r\mathbb{E}_t(\hat{q}_{t+k}) + \mathbb{E}_t(\varepsilon_{it+k})]. \qquad (146)\end{aligned}$$

We apply the method of undetermined coefficients to $\hat{q}_{t+k}$ in (146). We conjecture that the solutions for the aggregate RER and the aggregate real reset exchange rate take the following form:

$$\hat{q}_t = \theta_1\hat{q}_{t-1} + \theta_2\varepsilon_t^n, \qquad (147)$$

$$\hat{q}_t^{opt} = \beta_1\hat{q}_{t-1} + \beta_2\varepsilon_t^n, \qquad (148)$$

respectively. Here, θ_1, θ_2, β_1, and β_2 are the undetermined coefficients. Note that both $\hat{q}_t$ and $\hat{q}_t^{opt}$ depend on $\hat{q}_{t-1}$ and ε_t^n, but not on $\hat{q}_{it-1}$ or ε_{it}^r, since aggregation over i washes out idiosyncratic components.

Using (147), we can rewrite (146) as

$$\begin{aligned}\hat{q}_{it} &= \lambda\hat{q}_{it-1} + \lambda\varepsilon_t^n + r\frac{(1-\lambda\delta)\psi^{RP}}{1-\lambda\delta\theta_1}(1-\lambda)\hat{q}_t + (1-\lambda)(1-\lambda\delta)\psi^{RP}\varepsilon_{it}^r \\ &= \lambda\hat{q}_{it-1} + \lambda\varepsilon_t^n + r\nu(1-\lambda)\hat{q}_t + (1-\lambda)(1-\lambda\delta)\psi^{RP}\varepsilon_{it}^r, \qquad (149)\end{aligned}$$

where $\nu = (1-\lambda\delta)\psi^{RP}/(1-\lambda\delta\theta_1)$. We here used the assumption that $\varepsilon_{it}^r \sim i.i.d.$ Noting that $\hat{q}_t = \ln q_t$, this expression corresponds to (47) in the main text.

Next, we solve for the undetermined coefficients. Substituting (147) and (148) into (145) yields

$$\begin{aligned}\beta_1\hat{q}_{t-1} + \beta_2\varepsilon_t^n &= (1-\lambda\delta)\psi^{RP}r(\theta_1\hat{q}_{t-1} + \theta_2\varepsilon_t^n) + \lambda\delta\mathbb{E}_t(\beta_1\hat{q}_t + \beta_2\varepsilon_{t+1}^n) \\ &= (1-\lambda\delta)\psi^{RP}r(\theta_1\hat{q}_{t-1} + \theta_2\varepsilon_t^n) + \lambda\delta(\beta_1\theta_1\hat{q}_{t-1} + \beta_1\theta_2\varepsilon_t^n).\end{aligned}$$

Matching coefficients gives

$$\beta_1 = (1-\lambda\delta)\psi^{RP}r\theta_1 + \lambda\delta\beta_1\theta_1, \qquad (150)$$

$$\beta_2 = (1-\lambda\delta)\psi^{RP}r\theta_2 + \lambda\delta\beta_1\theta_2. \qquad (151)$$

Substituting (147) and (148) into (144), we obtain

$$\theta_1\hat{q}_{t-1} + \theta_2\varepsilon_t^n = \lambda\hat{q}_{t-1} + \lambda\varepsilon_t^n + (1-\lambda)(\beta_1\hat{q}_{t-1} + \beta_2\varepsilon_t^n).$$

Matching coefficients yields

$$\theta_1 = \lambda + (1-\lambda)\beta_1, \tag{152}$$

$$\theta_2 = \lambda + (1-\lambda)\beta_2. \tag{153}$$

While the system of equations (150)–(153) can be used to solve for all undetermined coefficients, it is sufficient to solve only for θ_1, since θ_1 is the only coefficient required to compute $\nu = (1-\lambda\delta)\psi^{RP}/(1-\lambda\delta\theta_1)$. It is straightforward to show that:

$$\theta_1 = \frac{1}{2\lambda\delta}\left\{1+\lambda^2\delta-(1-\lambda)(1-\lambda\delta)\psi^{RP}r-\sqrt{[1+\lambda^2\delta-(1-\lambda)(1-\lambda\delta)\psi^{RP}r]^2-4\lambda^2\delta}\right\}. \tag{154}$$

A.11.4 Evaluating the model of roundabout production

The purpose of our analysis is to evaluate the model of roundabout production using calibrated parameters. The coefficient $r\nu$ in (47) corresponds to the regression coefficient β in (36). Therefore, we simulate $r\nu$ based on calibrated values and compare it to the estimated regression coefficient. The left panel of Figure A.1 reports the regression coefficient predicted by $r\nu$ in (47) under the model of roundabout production. Here, all parameter values are set to those in the model of behavioral inattention, except for the degree of roundabout production r. In this panel, the solid line shows that the predicted coefficient ranges between 0.00 and 0.55 and increases with r. For comparison, we choose a conservative estimate of $\hat{\beta} = 0.80$ from Tables 1 and 2. A comparison between $r\nu$ and $\hat{\beta}$ reveals that the predicted coefficient $r\nu$ is considerably lower than the estimated coefficient $\hat{\beta}$, even when r is close to unity.

We also note that θ_1 equals the first-order autocorrelation of the aggregate RER. Since ε_t^n is assumed to be i.i.d., (147) represents an AR(1) process, implying that the first-order autocorrelation of $\hat{q}_t$ is given by θ_1. The right panel plots the predicted first-order autocorrelation of $\ln q_t$, represented by θ_1, namely, the persistence of the aggregate RER. This panel confirms that the persistence predicted by the model of roundabout production (solid line) remains low, provided that all parameters (other than r) match those in the model of behavioral inattention. Although persistence increases with r and is larger than λ (dashed line), the magnitude of increase remains modest. It falls short of 0.60 even if we increase r

up to 0.99.

It is straightforward to derive the impulse response functions of the aggregate RER to a nominal shock ε_t^n in the model of roundabout production, since the aggregate RER follows an AR(1) process with persistence θ_1. The simulated impulse response functions are shown in Figure 6. The aggregate RER in the model of roundabout production is clearly less persistent than in the model of behavioral inattention. Moreover, the predicted first-order autocorrelation can be translated into the half-life of the RER using the standard formula. As discussed in the main text, even under a high degree of roundabout production, the implied half-life is only 1.1 years, given the same parameter values used in the model of behavioral inattention.

We thus conclude that, although the regression equation under the model of roundabout production shares the same structure as that under the behavioral inattention model, it fails to replicate both the estimated regression coefficient $\hat{\beta}$ and the observed persistence of the aggregate RER. It is important to emphasize that our evaluation is based on the presumption that all parameters other than r are identical to those in the model of behavioral inattention. Under this presumption, the roundabout production model does not have a sufficiently strong strategic complementarity to generate persistence in RERs, either at the aggregate or at the good level.

A.12 Derivation of (48) and (49)

We first derive (48). We rewrite (16) as

$$\begin{aligned}
\frac{\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)}{1-\lambda\delta} &= \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k \left[m_{1H} \left(\hat{w}_{t+k} + \sum_{l=1}^{k} \pi_{t+l} \right) - m_{2H} \hat{a}_{it+k} \right] \\
&= m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it} \\
&\quad + \mathbb{E}_t \left\{ \mathbb{E}_{t+1} \sum_{k=0}^{\infty} (\lambda\delta)^{k+1} \left[m_{1H} \left(\hat{w}_{t+1+k} + \pi_{t+1} + \sum_{l=1}^{k} \pi_{t+1+k} \right) - m_{2H} \hat{a}_{it+k+1} \right] \right\} \\
&= m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it} \\
&\quad + \lambda\delta \mathbb{E}_t \left\{ \mathbb{E}_{t+1} \sum_{k=0}^{\infty} (\lambda\delta)^{k} \left[m_{1H} \left(\hat{w}_{t+1+k} + \sum_{l=1}^{k} \pi_{t+1+k} \right) - m_{2H} \hat{a}_{it+k+1} \right] \right\} \\
&\quad + \lambda\delta \mathbb{E}_t \left[m_{1H} \sum_{k=0}^{\infty} (\lambda\delta)^k \pi_{t+1} \right] \\
&= m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it} + \lambda\delta \mathbb{E}_t \left[\frac{\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit+1}, \boldsymbol{m}_H)}{1-\lambda\delta} + m_{1H} \frac{\pi_{t+1}}{1-\lambda\delta} \right] \\
&= m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it} + \frac{\lambda\delta}{1-\lambda\delta} \mathbb{E}_t \left[\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit+1}, \boldsymbol{m}_H) + m_{1H}\pi_{t+1} \right].
\end{aligned}$$

Multiplying both sides by $1-\lambda\delta$ yields

$$\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) = (1-\lambda\delta)(m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it}) + \lambda\delta \mathbb{E}_t[\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit+1}, \boldsymbol{m}_H) + m_{1H}\pi_{t+1}]. \quad (155)$$

Suppressing constant terms, this equation is also be written in terms of the logarithm:

$$\begin{aligned}
\ln P_{Hit} - \ln P_t = (1-\lambda\delta)[m_{1H}(\ln W_t - \ln P_t) - m_{2H} \ln a_{it}] \\
+ \lambda\delta \mathbb{E}_t \left[\ln P_{Hit+1} - \ln P_{t+1} + m_{1H} (\ln P_{t+1} - \ln P_t) \right].
\end{aligned}$$

Collecting terms yields

$$\begin{aligned}
\ln P_{Hit} - (1-m_{1H}) \ln P_t &= (1-\lambda\delta)(m_{1H} \ln W_t - m_{2H} \ln a_{it}) + \lambda\delta \mathbb{E}_t \left[\ln P_{Hit+1} - (1-m_{1H}) \ln P_{t+1} \right] \\
&= (1-\lambda\delta) \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k (m_{1H} \ln W_{t+k} - m_{2H} \ln a_{it+k}).
\end{aligned}$$

When we assume that $\hat{a}_{it} = 0$ for all t, we obtain (48) in the main text.

Under the assumptions in our model, it can easily be shown that (48) reduces to (29). From (2) and (4), we have $\mathbb{E}_t \ln W_{t+k} = \ln \chi + \mathbb{E}_t \ln M_{t+k} = \ln \chi + \ln M_t$. The second term of the right-hand side of (48) becomes $m_{1H} \ln W_t$ so that we obtain (29) in the main text.

Turning to the derivation of (49), we use (134). Suppressing constant terms, (134) turns out to be as follows:

$$\begin{aligned}
\ln P_{Hit} &= \ln P_t + (1-\lambda\delta)[(1-r)(\ln W_t - \ln P_t) - \ln a_{it}] \\
&\quad + \lambda\delta \mathbb{E}_t(\ln P_{Hit+1} - \ln P_{t+1}) + \lambda\delta(\mathbb{E}_t \ln P_{t+1} - \ln P_t) \\
&= (1-\lambda\delta)\left[r \ln P_t + (1-r)\ln W_t - \ln a_{it}\right] + \lambda\delta \mathbb{E}_t \ln P_{Hit+1} \\
&= (1-\lambda\delta)\mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k \left[r \ln P_{t+k} + (1-r)\ln W_{t+k} - \ln a_{it+k}\right].
\end{aligned}$$

When we impose $\hat{a}_{it} = 0$ for all t on the above equation, we obtain (49) in the main text.

A.13 A comparison between behavioral and rational inattention

In this section, we compare two models of inattention: behavioral inattention of Gabaix (2014) and rational inattention of Sims (2003) and Maćkowiak and Wiederholt (2009). We first derive the dynamic equation for good-level RERs in each model, namely, (52) and (54). We then define firms' problems of choosing attention in a unified framework. In both models, firms minimize the quadratically approximated profit loss of deviating from the optimal price under rational expectations with full attention, given a cost function associated with attention. For further details of the models of rational inattention and behavioral inattention, see Maćkowiak et al. (2023) and Gabaix (2019).

A.13.1 Deriving the dynamic equations (52) and (54)

Rational inattention Let us consider US firms' pricing decisions when selling their brands in US markets. For simplicity, we assume flexible prices. Given linear technology, firms' nominal marginal cost is W_t/a_{it}. The household's first-order condition (2) implies that nominal marginal costs can be rewritten as $\chi M_t/a_{it}$. Given the nominal marginal cost, the log optimal nominal price when the firms are fully attentive under rational expectations is

$$\ln P_{Hit}^{RE} = \ln M_t - \ln a_{it}, \tag{156}$$

where a constant term is suppressed. Throughout this section, we use superscripts explicitly to indicate the pricing model. Here, the superscript RE denotes rational expectations with full attention.

In the rational inattention model, firms observe brand-specific signals about nominal aggregate demand (i.e., $M_t = P_tC_t$) and idiosyncratic productivity (i.e., a_{it}). They receive only signals $s_{it}^M(z) = \ln M_t + \xi_{it}^M(z)$ and $s_{it}^a(z) = \ln a_{it} + \xi_{it}^a(z)$. The firms choose the distribution of signals by paying attention. Under the assumptions of Gaussian fundamentals ε_t^M and ε_{it}^a, the quadratic objective function, and the unbounded choice set for $\ln P_{it+k}(z)$, Gaussian signals are optimal (see Maćkowiak et al., 2023, p. 231). Therefore, we assume that $\xi_{it}^M(z) \sim N(0, \sigma_{\xi M}^2)$ and $\xi_{it}^a(z) \sim N(0, \sigma_{\xi a}^2)$. The firms reduce the uncertainty of signals given the distribution of ε_t^M and ε_{it}^a.

The firms in the rational inattention model (denoted by the superscript RI) determine their prices based on their expectations of nominal marginal costs:

$$\begin{aligned}
\ln P_{Hit}^{RI}(z) &= \mathbb{E}_{izt}(\ln M_t - \ln a_{it}) \\
&= \ln M_{t-1} + m_{1H}^{RI}(s_{it}^M(z) - \ln M_{t-1}) - m_{2H}^{RI} s_{it}^a(z) \\
&= \ln M_{t-1} + m_{1H}^{RI}[\varepsilon_t^M + \xi_{it}^M(z)] - m_{2H}^{RI}[\varepsilon_{it}^a + \xi_{it}^a(z)], \qquad (157)
\end{aligned}$$

where we use the simplifying assumption that M_{t-1} is fully known at the beginning of period t. The coefficients m_{1H}^{RI} and m_{2H}^{RI} correspond to the steady-state Kalman gains given by $m_{1H}^{RI} = \sigma_M^2/(\sigma_M^2 + \sigma_{\xi M}^2)$ and $m_{2H}^{RI} = \sigma_a^2/(\sigma_a^2 + \sigma_{\xi a}^2)$, respectively. As we will discuss later, firms endogenously choose m_{1H}^{RI} and m_{2H}^{RI} by minimizing $\sigma_{\xi M}$ and $\sigma_{\xi a}$, subject to the cost of information processing.

Turning to US firms' pricing decisions when selling their brands in Canadian markets, their price is based on their expectations of nominal marginal cost given by $(W_t/S_t)/a_{it}$ or $(\chi M_t/S_t)/a_{it}$. However, the international risk-sharing condition (3) and the CIA constraints in both countries leads to $S_t = M_t/M_t^*$ so that the nominal marginal cost can be rewritten as $\chi M_t^*/a_{it}$. Therefore, the US firms choose their prices based on the expectations:

$$\begin{aligned}
\ln P_{Hit}^{RI*}(z) &= \mathbb{E}_{izt}[\ln M_t^* - \ln a_{it}] \\
&= \ln M_{t-1}^* + m_{1H}^{RI*}[\varepsilon_t^{M^*} + \xi_{it}^{M^*}(z)] - m_{2H}^{RI*}[\varepsilon_{it}^a + \xi_{it}^a(z)]. \qquad (158)
\end{aligned}$$

Note that (158) differs from (157) in that the chosen prices are influenced by the nominal aggregate demand in Canada.

By symmetry, the prices of foreign-produced brands for selling Canadian and US market are

$$\begin{aligned}
\ln P_{Fit}^{RI*}(z) &= \ln M_{t-1}^* + m_{1F}^{RI*}[\varepsilon_t^{M^*} + \xi_{it}^{M^*}(z)] - m_{2F}^{RI*}[\varepsilon_{it}^{a^*} + \xi_{it}^{a^*}(z)], \qquad (159) \\
\ln P_{Fit}^{RI}(z) &= \ln M_{t-1} + m_{1F}^{RI}[\varepsilon_t^M + \xi_{it}^M(z)] - m_{2F}^{RI}[\varepsilon_{it}^{a^*} + \xi_{it}^{a^*}(z)], \qquad (160)
\end{aligned}$$

respectively.

We next aggregate these brands' prices to obtain the price index for good i, P_{it} and P_{it}^*. Under flexible prices, all firms reset prices every period. Therefore, aggregating reset prices for good i are identical to obtaining price indexes for good i (e.g., $\ln P_{it}^{opt} = \ln P_{it}$). Log-linearizing $(p_{it})^{1-\varepsilon} = (p_{it}^{opt})^{1-\varepsilon} = \int_z p_{it}(z)^{1-\varepsilon} dz$ from the CES index and using (157) and (160) yield

$$\begin{aligned}
\ln P_{it} &= \left(\frac{\bar{p}_{Hi}}{\bar{p}_i}\right)^{1-\varepsilon} \int_{z=0}^{\frac{1}{2}} \ln P_{Hit}^{RI}(z) dz + \left(\frac{\bar{p}_{Fi}}{\bar{p}_i}\right)^{1-\varepsilon} \int_{z=\frac{1}{2}}^{1} \ln P_{Fit}^{RI}(z) dz \\
&= \left(\frac{\bar{p}_{Hi}}{\bar{p}_i}\right)^{1-\varepsilon} \int_{z=0}^{\frac{1}{2}} \left\{ \ln M_{t-1} + m_{1H}^{RI} \left[\varepsilon_t^M + \xi_{it}^M(z)\right] - m_{2H}^{RI} \left[\varepsilon_t^a + \xi_{it}^a(z)\right] \right\} dz \qquad (161) \\
&\quad + \left(\frac{\bar{p}_{Fi}}{\bar{p}_i}\right)^{1-\varepsilon} \int_{z=\frac{1}{2}}^{1} \left\{ \ln M_{t-1} + m_{1F}^{RI} \left[\varepsilon_t^M + \xi_{it}^M(z)\right] - m_{2F}^{RI} \left[\varepsilon_t^{a^*} + \xi_{it}^{a^*}(z)\right] \right\} dz.
\end{aligned}$$

In addition, aggregation across brands eliminates the noise. Namely, we have $\int_0^{1/2} \xi_{it}^M(z) dz = \int_{1/2}^2 \xi_{it}^M(z) dz = \int_0^{1/2} \xi_{it}^a(z) dz = \int_{1/2}^1 \xi_{it}^{a^*}(z) dz = 0$. Thus, using the definition of $\omega = (1/2)(\bar{p}_{Hi}/\bar{p}_i)^{1-\varepsilon}$, we have the nominal price index for good i in US markets:

$$\begin{aligned}
\ln P_{it} &= \ln M_{t-1} + \omega \left(m_{1H}^{RI} \varepsilon_t^M - m_{2H}^{RI} \varepsilon_{it}^a \right) + (1-\omega) \left(m_{1F}^{RI} \varepsilon_t^M - m_{2F}^{RI} \varepsilon_{it}^{a^*} \right) \\
&= \ln M_{t-1} + m^{RI} \varepsilon_t^M - \left[\omega m_{2H}^{RI} \varepsilon_{it}^a + (1-\omega) m_{2F}^{RI} \varepsilon_{it}^{a^*} \right], \qquad (162)
\end{aligned}$$

where $m^{RI} = \omega m_{1H}^{RI} + (1-\omega) m_{1F}^{RI}$. Analogously, we obtain the nominal price index for good i in Canadian markets:

$$\ln P_{it}^* = \ln M_{t-1}^* + m^{RI} \varepsilon_t^{M^*} - \left[\omega m_{2F}^{RI*} \varepsilon_{it}^{a^*} + (1-\omega) m_{2H}^{RI*} \varepsilon_{it}^a \right], \qquad (163)$$

where $m^{RI} = \omega m_{1F}^{RI*} + (1-\omega) m_{1H}^{RI*}$. Note that $m_{jH}^{RI} = m_{jF}^{RI*}$ and $m_{jF}^{RI} = m_{jH}^{RI*}$ for $j = 1, 2$ hold as in the baseline model.

Combining (162) and (163) with the definition of the good-level RER, we obtain the dynamic equation under rational inattention (52):

$$\begin{aligned}
\ln q_t &= \ln S_t + \ln P_{it}^* - \ln P_{it} \\
&= \left(\ln M_t - \ln M_{t-1} - m^{RI} \varepsilon_t^M \right) - \left(\ln M_t^* - \ln M_{t-1}^* - m^{RI} \varepsilon_t^{M^*} \right) \\
&\quad + [\omega m_{2H}^{RI} - (1-\omega) m_{2F}^{RI}] \varepsilon_{it}^a - [\omega m_{2H}^{RI} - (1-\omega) m_{2F}^{RI}] \varepsilon_{it}^{a^*} \\
&= (1 - m^{RI}) \varepsilon_t^n + \psi^{RI} \varepsilon_{it}^r,
\end{aligned}$$

where $\psi^{RI} = \omega m_{2H}^{RI} - (1-\omega) m_{2F}^{RI}$, $\varepsilon_t^n = \varepsilon_t^M - \varepsilon_t^{M^*}$, and $\varepsilon_{it}^r = \varepsilon_{it}^a - \varepsilon_{it}^{a^*}$.

Behavioral inattention We again begin with considering US firms' pricing for selling their brands in US markets. For comparison, we assume that firms pay attention to nominal marginal costs, $\chi M_t/a_{it}$, rather than to real marginal costs. In terms of the information set, firms have full access to nominal aggregate demand and idiosyncratic labor productivity, but paying attention to these variables is costly. If we apply the degree of attention to the nonstationary variable $\ln M_t$ in (156), however, the degree of attention to $\ln M_t$ becomes one because the variance of $\ln M_t$ diverges to infinity (see Section 2.3.2). To address this, we decompose $\ln M_t (= \ln M_{t-1} + \varepsilon_t^M)$ into $\ln M_{t-1}$ and ε_t^M and assume that firms choose the degrees of attention to nonstationary $\ln M_{t-1}$ and stationary ε_t^M *separately.* In this case, the degree of attention to $\ln M_{t-1}$ becomes one but the degree of attention to ε_t^M is less than one.

In the model of behavioral inattention, the log optimal nominal price (denoted by the superscript BI) is given by

$$\ln P_{Hit}^{BI} = \ln M_{t-1} + m_{1H}^{BI} \varepsilon_t^M - m_{2H}^{BI} \varepsilon_{it}^a. \tag{164}$$

While the degree of attention to $\ln M_{t-1}$ is unity, the degrees of attention m_{1H}^{BI} and m_{2H}^{BI} are not necessarily equal to one because ε_t^M and ε_{it}^a have finite variances. Later, we will redefine the sparse max problem that determines the optimal degrees of attention to ε_t^M and ε_{it}^a, subject to a cost function $C(\boldsymbol{m}_H^{BI})$. In contrast to the rational inattention model, the log optimal nominal prices under behavioral inattention do not depend on z because there is no brand-specific noise to fundamentals.

Turning to the log optimal nominal price of US-produced brands in Canadian markets, it is given by

$$\ln P_{Hit}^{BI*} = \ln M_{t-1}^* + m_{1H}^{BI*} \varepsilon_t^{M*} - m_{2H}^{BI*} \varepsilon_{it}^a, \tag{165}$$

because their nominal marginal costs are given by $\chi M_t^*/a_{it}$.

The log optimal nominal prices chosen by Canadian firms are

$$\ln P_{Fit}^{BI*} = \ln M_{t-1}^* + m_{1F}^{BI*} \varepsilon_t^{M*} - m_{2F}^{BI*} \varepsilon_{it}^{a^*}, \tag{166}$$

$$\ln P_{Fit}^{BI} = \ln M_{t-1} + m_{1F}^{BI} \varepsilon_t^{M} - m_{2F}^{BI} \varepsilon_{it}^{a^*}, \tag{167}$$

for brands sold in Canadian and US markets, respectively.

Aggregating reset prices described above yields the price indexes for good i, P_{it} and P_{it}^*. As in the case of rational inattention, log-linearizing $(p_{it})^{1-\varepsilon} = (p_{it}^{opt})^{1-\varepsilon} = \int_z p_{it}(z)^{1-\varepsilon} dz$ from the CES index and using (164) and (167) yield

$$\begin{aligned}\ln P_{it} &= \left(\frac{\bar{p}_{Hi}}{\bar{p}_i}\right)^{1-\varepsilon} \int_{z=0}^{\frac{1}{2}} \ln P_{Hit}^{BI} dz + \left(\frac{\bar{p}_{Fi}}{\bar{p}_i}\right)^{1-\varepsilon} \int_{z=\frac{1}{2}}^{1} \ln P_{Fit}^{BI} dz \\ &= \frac{1}{2}\left(\frac{\bar{p}_{Hi}}{\bar{p}_i}\right)^{1-\varepsilon} \left(\ln M_{t-1} + m_{1H}^{BI}\varepsilon_t^M - m_{2H}^{BI}\varepsilon_t^a\right) + \frac{1}{2}\left(\frac{\bar{p}_{Fi}}{\bar{p}_i}\right)^{1-\varepsilon} \left(\ln M_{t-1} + m_{1F}^{BI}\varepsilon_t^M - m_{2F}^{BI}\varepsilon_t^{a^*}\right).\end{aligned}$$

Using the definition of ω and arranging terms, we obtain the nominal price index for good i in US markets:

$$\begin{aligned}\ln P_{it} &= \omega\left(\ln M_{t-1} + m_{1H}^{BI}\varepsilon_t^M - m_{2H}^{BI}\varepsilon_t^a\right) + (1-\omega)\left(\ln M_{t-1} + m_{1F}^{BI}\varepsilon_t^M - m_{2F}^{BI}\varepsilon_t^{a^*}\right) \\ &= \ln M_{t-1} + m^{BI}\varepsilon_t^M - [\omega m_{2H}^{RI}\varepsilon_{it}^a + (1-\omega)m_{2F}^{RI}\varepsilon_{it}^{a^*}],\end{aligned} \tag{168}$$

where $m^{BI} = \omega m_{1H}^{BI} + (1-\omega)m_{1F}^{BI}$. Analogously, we have the nominal price index for good i in Canadian markets:

$$\ln P_{it}^* = \ln M_{t-1}^* + m^{BI}\varepsilon_t^{M^*} - [\omega m_{2F}^{RI*}\varepsilon_{it}^{a^*} + (1-\omega)m_{2H}^{RI*}\varepsilon_{it}^a], \tag{169}$$

where $m^{BI} = \omega m_{1F}^{BI*} + (1-\omega)m_{1H}^{BI*} = \omega m_{1H}^{BI} + (1-\omega)m_{1F}^{BI}$.

Combining (168) and (169) with the definition of the good-level RER yields the dynamic equation under behavioral inattention (54):

$$\begin{aligned}\ln q_t &= \ln S_t + \ln P_{it}^* - \ln P_{it}^* \\ &= (1-m^{BI})\varepsilon_t^n + \psi^{BI}\varepsilon_{it}^r,\end{aligned}$$

where $\psi^{BI} = \omega m_{2H}^{BI} - (1-\omega)m_{2F}^{BI}$.

As discussed in the main text, the dynamic equations we derived are observationally equivalent if $m^{RI} = m^{BI}$ and $\psi^{RI} = \psi^{BI}$ hold. In both models, the degrees of attention are chosen endogenously. For example, in the rational inattention model, m_{1H}^{RI} becomes closer to one as firms make more effort to reduce the uncertainty of signals (i.e., reducing $\sigma_{\xi M}$). In the model of behavioral inattention, firms directly choose the degrees of attention. The two models are similar but the endogenously chosen degree of attention differ in terms of underlying structural parameters.

A.13.2 Deriving profit loss

In this section, we introduce the maximization problem for choosing attention for the models of rational and behavioral inattention in a unified framework. As discussed in Section 2.3.2, the objective function is based on a quadratically approximated profit loss of deviating from the optimal price under rational expectations. By default, the optimal price under rational expectations is set by fully attentive firms. The derivation of a quadratically approximated profit loss follows the literature. For further details, see also Maćkowiak and Wiederholt (2009).

We take the expected discounted sum of real profits and approximate it to the second order. Using the example of US firms' profits from selling their brands in US markets, the profit loss of deviating $\hat{p}_{it+k}(z)$ from $\hat{p}^{RE}_{it+k}$ is approximated as

$$
\begin{aligned}
& \mathbb{E}\left[\tilde{v}_{Hi}(\{\hat{p}_{it+k}(z)\}_{k=0}^{\infty})\right] - \mathbb{E}\left[\tilde{v}_{Hi}(\{\hat{p}^{RE}_{Hit+k}\}_{k=0}^{\infty})\right] \\
\simeq\ & \frac{1}{2}\sum_{k=0}^{\infty} \frac{\partial^2 \tilde{v}^0_{Hi}}{\partial[\hat{p}_{it+k}(z)]^2} \mathbb{E}\left\{\left[\hat{p}_{it+k}(z) - \hat{p}^{RE}_{Hit+k}\right]^2\right\} \\
=\ & \frac{1}{2}\sum_{k=0}^{\infty} \frac{\partial^2 \tilde{v}^0_{Hi}}{\partial[\hat{p}_{it+k}(z)]^2} \mathbb{E}\left\{\left[\ln P_{it+k}(z) - \ln P^{RE}_{it+k}\right]^2\right\},
\end{aligned} \tag{170}
$$

for any $\hat{p}_{it+k}(z)$ for $z \in [0, 1/2]$. Here $\tilde{v}_{Hi}(\{\hat{p}_{it+k}(z)\}_{k=0}^{\infty})$ is the expected discounted sum of real profits:

$$
\begin{aligned}
\tilde{v}_{Hi}(\{\hat{p}_{it+k}(z)\}_{k=0}^{\infty}) &= \mathbb{E}_t \sum_{k=0}^{\infty} \delta_{t,t+k}\left[\frac{P_{it+k}(z)}{P_{t+k}} - \frac{W_{t+k}/a_{it}}{P_{t+k}}\right] c_{it+k}(z) \\
&= \mathbb{E}_t \sum_{k=0}^{\infty} \delta_{t,t+k}\left\{\bar{p}_i(z)\exp[\hat{p}_{it+k}(z)] - \bar{w}\exp(\hat{w}_{t+k} - \hat{a}_{it+k})\right\} c_{it+k}(z),
\end{aligned}
$$

where $c_{it+k}(z) = (P_{it+k}(z)/P_{t+k})^{-\varepsilon} c_{it+k} = [\bar{p}_i(z)/\bar{p}_i]^{-\varepsilon}\exp\{-\varepsilon[\hat{p}_{it+k}(z) - \hat{p}_{it+k}]\}c_{it+k}$. Because flexible prices are assumed in both models, the formulation of the profit function differs slightly from the objective function for choosing prices under sticky prices. The quadratically approximated profit loss (170) has the second derivatives of the objective function $\partial^2\tilde{v}^0_{Hi}/\partial[\hat{p}_{it+k}(z)]^2$ evaluated at the steady state, similar to (78). However, we take the derivatives with respect to $\hat{p}_{it+k}(z)$ for $k = 0, 1, 2, ...$ and sum over k. Since $\partial^2\tilde{v}^0_{Hi}/\partial[\hat{p}_{it+k}(z)]^2 = \delta^k(1-\varepsilon)2\omega\bar{c} < 0$ and decays exponentially over k, the summation $\sum_{k=0}^{\infty}\partial^2\tilde{v}^0_{Hi}/\partial[\hat{p}_{it+k}(z)]^2$ converges to a constant. In addition, recall that $\hat{p}_{it+k}(z)$ and $\hat{p}^{RE}_{it+k}$ are defined as $\hat{p}_{it+k}(z) = \ln(P_{it+k}(z)/P_{t+k}) - \ln\bar{p}_i(z)$ and $\hat{p}^{RE}_{it+k} = \ln(P^{RE}_{it+k}/P_{t+k}) - \ln\bar{p}_i(z)$, respectively. Therefore, as shown in the third equality in (170), $\hat{p}_{it+k}(z) - \hat{p}^{RE}_{Hit+k}$ is equal to the deviation from the log optimal nominal price: $\ln P_{it+k}(z) - \ln P^{RE}_{it+k}$.

Firms choose the degrees of attention by solving

$$\boldsymbol{m}_H = \arg\min_{\boldsymbol{m}_H \in [0,1]^2} -\frac{1}{2}\sum_{k=0}^{\infty} \frac{\partial^2 \tilde{v}^0_{Hi}}{\partial[\hat{p}_{it+k}(z)]^2}\mathbb{E}\left\{\left[\ln P_{it+k}(z) - \ln P^{RE}_{it+k}\right]^2\right\} + \tilde{\mathcal{C}}(\boldsymbol{m}_H), \tag{171}$$

where $\boldsymbol{m}_H = (m_{1H}, m_{2H})'$ is $\boldsymbol{m}^{RI}_H$ or $\boldsymbol{m}^{BI}_H$ when $\ln P_{it+k}(z)$ is evaluated at (157) or (164). The cost function $\tilde{\mathcal{C}}(\boldsymbol{m}_H)$ differs between the models of rational and behavioral inattention. Typically, $\tilde{\mathcal{C}}(\boldsymbol{m}_H)$ under rational inattention is a linear function of the Shannon mutual information between signals and fundamentals (see Maćkowiak et al., 2023, p. 230), while $\tilde{\mathcal{C}}(\boldsymbol{m}_H)$ under behavioral inattention is a polynomial function of m_{jH} (see Gabaix, 2019, p. 293), as shown in (18).

A.13.3 Endogenous choice of attention

In this section, we specify the quadratically approximated profit loss and the cost function for each model, based on (171).

Rational inattention We compute the expected value of the squared deviation of $\ln P^{RI}_{it+k}(z)$ from the log optimal nominal price under rational expectations $\ln P^{RE}_{it+k}$. In the case of US firms that sell their brands in US markets, it is given by

$$\begin{aligned}
\mathbb{E}[\ln P^{RI}_{Hit}(z) - \ln P^{RE}_{Hit}]^2 &= \mathbb{E}\left[\mathbb{E}_{izt}(\ln M_t - \ln a_{it}) - (\ln M_t - \ln a_{it})\right]^2 \\
&= \mathbb{E}\{\mathbb{E}_{izt}(\ln M_t - \ln M_{t-1}) - (\ln M_t - \ln M_{t-1}) - [\mathbb{E}_{izt}(\ln a_{it}) - \ln a_{it}]\}^2 \\
&= \mathbb{E}[\varepsilon^M_t - \mathbb{E}_{izt}(\varepsilon^M_t)]^2 + \mathbb{E}[\varepsilon^a_{it} - \mathbb{E}_{izt}(\varepsilon^a_{it})]^2 \\
&= \sigma^2_{M|s} + \sigma^2_{a|s},
\end{aligned} \tag{172}$$

where the second equality follows from the assumption that M_{t-1} is fully known at the beginning of period t and the third equality follows from (4) and (6), along with the independence between ε^M_t and ε^a_{it}. In the last equality, $\sigma^2_{M|s}$ and $\sigma^2_{a|s}$ denote the forecast error variances (posterior variances) after optimally choosing the variances of noises in $s^M_{it}(z)$ and $s^a_{it}(z)$, respectively. Note that the posterior variance $\sigma^2_{M|s}$ is rewritten as

$$\begin{aligned}
\sigma^2_{M|s} &= \mathbb{E}[\varepsilon^M_t - \mathbb{E}_{izt}(\varepsilon^M_t)]^2 \\
&= \mathbb{E}\{\varepsilon^M_t - m^{RI}_{1H}[s^M_{it}(z) - \ln M_{t-1}]\}^2 \\
&= \mathbb{E}[(1 - m^{RI}_{1H})\varepsilon^M_t - m^{RI}_{1H}\xi^M_{it}(z)]^2 \\
&= (1 - m^{RI}_{1H})^2\sigma^2_M + (m^{RI}_{1H})^2\sigma^2_{\xi M} \\
&= \left(\frac{\sigma^2_{\xi M}}{\sigma^2_M + \sigma^2_{\xi M}}\right)^2 \sigma^2_M + \left(\frac{\sigma^2_M}{\sigma^2_M + \sigma^2_{\xi M}}\right)^2 \sigma^2_{\xi M} \\
&= (1 - m^{RI}_{1H})\sigma^2_M, \qquad (173)
\end{aligned}$$

where the second equality results from a Kalman filtering, the third equality is from $s^M_{it}(z) = \ln M_t + \xi^M_{it}(z)$, and the fourth equality follows from the independence between ε^M_t and $\xi^M_{it}(z)$. In the last two equalities, we use $m^{RI}_{1H} = \sigma^2_M/(\sigma^2_M + \sigma^2_{\xi M})$. Similarly, $\sigma^2_{a|s}$ is given by

$$\sigma^2_{a|s} = (1 - m^{RI}_{2H})\sigma^2_a. \qquad (174)$$

Consequently, (172) becomes

$$\mathbb{E}[\ln P^{RI}_{Hit}(z) - \ln P^{RE}_{Hit}]^2 = (1 - m^{RI}_{1H})\sigma^2_M + (1 - m^{RI}_{2H})\sigma^2_a \qquad (175)$$

The intuition behind (175) is straightforward: If firms pay no attention to signals (i.e., $\boldsymbol{m}_H = \mathbf{0}$), they receive no useful information from the signals. In this case, they cannot lower the forecast error variance, which remains equal tothe unconditional variance $\sigma^2_M + \sigma^2_a$. In contrast, by paying full attention to signals, firms fully eliminate noise and extract the information of fundamentals. In this case, they make no forecast errors and minimize the profit loss from deviating from the log optimal nominal price under rational expectations.

We next turn to $\tilde{\mathcal{C}}(\boldsymbol{m}_H)$ in the rational inattention model. The cost of information depends on the Shannon mutual information between the two random variables: fundamentals (e.g., ε^M_t) and signals (e.g., $s^M_{it}(z)$). The Shannon mutual information is the difference between the entropy of fundamentals and the conditional entropy of fundamentals given signals. For example, the entropy of ε^M_t and the conditional entropy of ε^M_t given $s^M_{it}(z)$ are

$$\begin{aligned}
\mathcal{H}(\varepsilon^M_t) &= \frac{1}{2}\ln(2\pi e\sigma^2_M), \\
\mathcal{H}[\varepsilon^M_t | s^M_{it}(z)] &= \frac{1}{2}\ln(2\pi e\sigma^2_{M|s}),
\end{aligned}$$

respectively.[55] Here the unit of information flow is measured in nats. This mutual information quantifies the reduction in uncertainty after observing the signals and is given by

$$\begin{aligned}\mathcal{I}[\varepsilon_t^M|s_{it}^M(z)] &= \mathcal{H}(\varepsilon_t^M) - \mathcal{H}[\varepsilon_t^M|s_{it}^M(z)] \\ &= \frac{1}{2}\ln\left(\frac{\sigma_M^2}{\sigma_{M|s}^2}\right).\end{aligned}$$

Fundamentals $\boldsymbol{\varepsilon}_t \equiv (\varepsilon_t^M, \varepsilon_{it}^a)'$ is independent across the entity, and signals $\boldsymbol{s}_{it}(z) = (s_{it}^M(z), s_{it}^a(z))'$ are uncorrelated across the entity. Moreover, they are serially uncorrelated. Therefore, the mutual information between $\{\boldsymbol{\varepsilon}_{t+k}\}_{k=0}^K$ and $\{\boldsymbol{s}_{it}(z)\}_{k=0}^K$ is

$$\begin{aligned}\mathcal{I}(\{\boldsymbol{\varepsilon}_{t+k}\}_{k=0}^K|\{\boldsymbol{s}_{it}(z)\}_{k=0}^K) &= \frac{K}{2}\ln\left(\frac{\sigma_M^2}{\sigma_{M|s}^2}\right) + \frac{K}{2}\ln\left(\frac{\sigma_a^2}{\sigma_{a|s}^2}\right) \\ &= \frac{K}{2}\ln\left(\frac{1}{1-m_{1H}^{RI}}\right) + \frac{K}{2}\ln\left(\frac{1}{1-m_{2H}^{RI}}\right),\end{aligned}$$

where the second equality results from (173) and (174). As discussed in Maćkowiak et al. (2023, p. 245), a popular specification of the cost function in the infinite horizon model is $\tilde{\mathcal{C}}(\boldsymbol{m}_H^{RI}) = \Theta \times [\lim_{K\to\infty} K^{-1}\mathcal{I}(\{\boldsymbol{\varepsilon}_{t+k}\}_{k=0}^K|\{\boldsymbol{s}_{it}(z)\}_{k=0}^K)]$, where $\Theta > 0$ is a constant information cost parameter. As a result,

$$\tilde{\mathcal{C}}(\boldsymbol{m}_H^{RI}) = \frac{\Theta}{2}\left[\ln\left(\frac{1}{1-m_{1H}^{RI}}\right) + \ln\left(\frac{1}{1-m_{2H}^{RI}}\right)\right]. \tag{176}$$

We now apply (175) and (176) to (171). The minimization problem for choosing attention under rational inattention is given by

$$\min_{\boldsymbol{m}_H^{RI}\in[0,1]^2} \frac{1}{2}\left[(1-m_{1H}^{RI})\tilde{\Lambda}_{1H} + (1-m_{2H}^{RI})\tilde{\Lambda}_{2H}\right] - \frac{\Theta}{2}\left[\ln(1-m_{1H}^{RI}) + \ln(1-m_{2H}^{RI})\right], \tag{177}$$

where $\tilde{\Lambda}_{1H} = -\sigma_M^2\left\{\sum_{k=0}^{\infty}\frac{\partial^2\tilde{v}_{Hi}^0}{\partial[\hat{p}_{it+k}(z)]^2}\right\}$ and $\tilde{\Lambda}_{2H} = -\sigma_a^2\left\{\sum_{k=0}^{\infty}\frac{\partial^2\tilde{v}_{Hi}^0}{\partial[\hat{p}_{it+k}(z)]^2}\right\}$. The solutions for m_{1H}^{RI} and m_{2H}^{RI} are given by

$$m_{1H}^{RI} = \max\left[0, 1-\frac{\Theta}{\tilde{\Lambda}_{1H}}\right], \tag{178}$$

$$m_{2H}^{RI} = \max\left[0, 1-\frac{\Theta}{\tilde{\Lambda}_{2H}}\right], \tag{179}$$

[55] For the derivation, see Cover and Thomas (2006, p. 244).

respectively. Note that analogous minimization problems can be defined for choosing attention $\boldsymbol{m}_H^{RI*}$, $\boldsymbol{m}_F^{RI*}$, and $\boldsymbol{m}_F^{RI}$.

Behavioral inattention As in the previous section, we again compute the expected value of the squared deviation from the log optimal nominal price under rational expectations with full attention. In the case of US firms that sell their brands in US markets, it is given by

$$\begin{aligned}
\mathbb{E}[\ln P_{Hit}^{BI}(z) - \ln P_{Hit}^{RE}]^2 &= \mathbb{E}\left[(\ln M_{t-1} + m_{1H}^{BI}\varepsilon_t^M - m_{2H}^{BI}\varepsilon_{it}^a) - (\ln M_{t-1} + \varepsilon_t^M - \varepsilon_{it}^a)\right]^2 \\
&= \mathbb{E}[(1 - m_{1H}^{RI})\varepsilon_t^M - (1 - m_{2H}^{RI})\varepsilon_{it}^a]^2 \\
&= (1 - m_{1H}^{RI})^2\sigma_M^2 + (1 - m_{2H}^{RI})^2\sigma_a^2. \qquad (180)
\end{aligned}$$

The intuition behind (180) is similar to that of (175). Under behavioral inattention, firms have full access to the nominal aggregate demand and idiosyncratic labor productivity. Despite full access, firms may choose to pay no attention to these variables ($\boldsymbol{m}_{1H} = \mathbf{0}$). In this case, the expected profit loss of deviating from the log optimal nominal price under full attention is equal to the unconditional variance $\sigma_M^2 + \sigma_a^2$. In contrast, if firms pay full attention, firms minimize the expected profit loss of deviating from the log optimal nominal price under full attention.

The cost function is a polynomial function of the degrees of attention. Here, we continue to assume that the cost function is given by (18):

$$\mathcal{C}\left(\boldsymbol{m}_H\right) = \frac{\kappa_1}{2}m_{1H}^2 + \frac{\kappa_2}{2}m_{2H}^2,$$

where $\kappa_j \geq 0$ for $j = 1, 2$.

We now apply (180) and this equation to (170). The US firms' minimization problem of choosing attention under behavioral inattention is:

$$\min_{\boldsymbol{m}_H^{BI} \in [0,1]^2} (1 - m_{1H}^{BI})^2\tilde{\Lambda}_{1H} + (1 - m_{2H}^{BI})^2\tilde{\Lambda}_{2H} - \left[\kappa_1(m_{1H}^{BI})^2 + \kappa_2(m_{2H}^{BI})^2\right]. \qquad (181)$$

The solution is

$$m_{1H}^{BI} = \frac{\tilde{\Lambda}_{1H}}{\tilde{\Lambda}_{1H} + \kappa_1}, \qquad (182)$$

$$m_{2H}^{BI} = \frac{\tilde{\Lambda}_{2H}}{\tilde{\Lambda}_{2H} + \kappa_2}. \qquad (183)$$

We can similarly define the minimization problem of choosing attention $\boldsymbol{m}_H^{BI*}$, $\boldsymbol{m}_F^{BI*}$, and $\boldsymbol{m}_F^{BI}$. While the solution for the degrees of attention has the same structure as in the main text, it differs in that we assume flexible prices and that firms pay attention to nominal marginal costs in the present case.

We note that, as discussed in Gabaix (2014), the cost function is flexible under the model of behavioral inattention. If the cost function is linear, rather than quadratic, in the degrees of attention, the solution for m_{jH} takes the same form as the rational inattention model: $m_{jH}^{BI} = \max[0, 1 - \Theta/\tilde{\Lambda}_{jH}]$ for $j = 1, 2$. In this case, the dynamic equation for the good-level RER under behavioral inattention becomes identical to that under rational inattention because m^{BI} and ψ^{BI} in (54) become identical to m^{RI} and ψ^{RI} in (52).

A.14 The model with the Taylor rule

This appendix presents the model with the Taylor rule. In the baseline model, we assumed that households face the CIA constraint and that central banks supply money according to constant money growth rules (4) and (5). These assumptions ensure that the NER is given by $S_t = M_t/M_t^*$, and that $\ln S_t$ follows a random walk. Here, we drop these assumptions and instead assume the uncovered interest parity (UIP) condition and the Taylor rule.

We first describe the model incorporating the Taylor rule. We then derive the structural equations used to evaluate its implications. Finally, we assess the first-order autocorrelations and half-lives implied by the model under various parameterizations of the Taylor rule.

A.14.1 Households and firms

The US households maximize $\mathbb{E}_0 \sum_{t=0}^{\infty} \delta^t (\ln c_t - \chi n_t)$ subject to an intertemporal budget constraint, $P_t c_t + \mathbb{E}_t(\Delta_{t,t+1} B_{t+1}) = W_t n_t + B_t + \Pi_t$, which slightly differs from (1). Under this maximization problem, the consumption Euler equation and the labor supply condition continue to hold. Therefore, we have $\hat{w}_t = \hat{c}_t$, $\hat{w}_t^* = \hat{c}_t^*$, and $\hat{q}_t = \hat{c}_t - \hat{c}_t^*$. The Canadian households face a similar intertemporal budget constraint: $P_t^* c_t^* + \mathbb{E}_t(\Delta_{t,t+1} B_{t+1}^*)/S_t = W_t^* n_t^* + B_t^*/S_t + \Pi_t^*$.

Inattentive firms solve the same optimization problem, and the optimal reset prices satisfy the same first-order condition. Therefore, (155) continues to hold for $\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H)$:

$$\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) = (1 - \lambda\delta)(m_{1H}\hat{w}_t - m_{2H}\hat{a}_{it}) + \lambda\delta\mathbb{E}_t[\hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit+1}, \boldsymbol{m}_H) + m_{1H}\pi_{t+1}]. \quad (184)$$

We can similarly derive the optimal reset prices set by Canadian firms for selling their brands in US markets (i.e., $\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F)$):

$$\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F) = (1-\lambda\delta)\left[m_{1F}(\hat{w}_t^* + \hat{q}_t) - m_{2F}\hat{a}_{it}^*\right] + \lambda\delta\mathbb{E}_t\left[\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit+1}, \boldsymbol{m}_F) + m_{1F}\pi_{t+1}\right], \quad (185)$$

as well as the corresponding foreign analogues: $\hat{p}_{Fi}^*(\hat{\boldsymbol{x}}_{Fit}^*, \boldsymbol{m}_F^*)$ and $\hat{p}_{Hi}^*(\hat{\boldsymbol{x}}_{Hit}^*, \boldsymbol{m}_H^*)$.

A.14.2 Monetary policy and the NER

The central bank in the US sets the nominal interest rate R_t according to the Taylor rule:

$$\hat{R}_t = \rho_R \hat{R}_{t-1} + (1-\rho_R)\alpha_\pi \pi_t + \varepsilon_t^R, \quad (186)$$

where $\alpha_\pi > 1$, and the degree of policy inertia ρ_R satisfies $\rho_R \in [0, 1)$. Here, ε_t^R denotes a monetary policy shock. The Canadian central bank sets R_t^* according to the same policy rule. Let $\hat{R}_t^n = \hat{R}_t - \hat{R}_t^*$. By symmetry, the relative policy interest rate $\hat{R}_t^n$ evolves as:

$$\hat{R}_t^n = \rho_R \hat{R}_{t-1}^n + (1-\rho_R)\alpha_\pi \pi_t^n + \varepsilon_t^n, \quad (187)$$

where $\pi_t^n = \pi_t - \pi_t^*$ is relative inflation. With some abuse of notation, we redefine the nominal shock ε_t^n as the difference in monetary policy shocks (i.e., $\varepsilon_t^n = \varepsilon_t^R - \varepsilon_t^{R^*}$).

The UIP condition links monetary policies in the two countries to the NER:

$$\ln S_t = \mathbb{E}_t \ln S_{t+1} - \hat{R}_t^n. \quad (188)$$

Note that a negative nominal shock (a decrease in ε_t^n) lowers the US policy interest rate relative to that in Canada, leading to a depreciation of the NER.

A.14.3 The good-level and aggregate RERs

Even in the model with the Taylor rule, (23)–(25) continue to hold. Using (24) and its foreign analogue, (23) implies

$$\begin{aligned} \hat{q}_{it} &= \hat{q}_t + \lambda(\hat{p}_{it-1}^* - \pi_t^*) + (1-\lambda)\hat{p}_{it}^{opt*} - \lambda(\hat{p}_{it-1} - \pi_t) - (1-\lambda)\hat{p}_{it}^{opt} \\ &= \lambda(\hat{q}_t - \hat{q}_{t-1}) + \lambda(\hat{q}_{t-1} + \hat{p}_{it-1}^* - \hat{p}_{it-1} + \pi_t^n) + (1-\lambda)(\hat{q}_t + \hat{p}_{it}^{opt*} - \hat{p}_{it}^{opt}) \\ &= \lambda\hat{q}_{it-1} + \lambda(\hat{q}_t - \hat{q}_{t-1} + \pi_t^n) + (1-\lambda)\hat{q}_{it}^{opt}. \end{aligned} \quad (189)$$

The last expression can be compared with (85). The second term of the right-hand side includes $\hat{q}_t - \hat{q}_{t-1} + \pi_t^n = \Delta \ln S_t$. However, $\Delta \ln S_t = \varepsilon_t^n$ does not hold in the model with the Taylor rule.

Next, combining (25) with (184) and (185) yields

$$\begin{aligned}
\hat{p}_{it}^{opt} &= \omega \hat{p}_{Hi}(\hat{\boldsymbol{x}}_{Hit}, \boldsymbol{m}_H) + (1-\omega)\hat{p}_{Fi}(\hat{\boldsymbol{x}}_{Fit}, \boldsymbol{m}_F) \\
&= (1-\lambda\delta)\left[\omega m_{1H}\hat{w}_t + (1-\omega)m_{1F}(\hat{w}_t^* + \hat{q}_t) - \omega m_{2H}\hat{a}_{it} - (1-\omega)m_{2F}\hat{a}_{it}^*\right] \\
&+ \lambda\delta \mathbb{E}_t\left[\hat{p}_{it+1}^{opt} + \omega m_{1H}\pi_{t+1} + (1-\omega)m_{1F}\pi_{t+1}\right] \\
&= (1-\lambda\delta)\left[m\hat{c}_t - \omega m_{2H}\hat{a}_{it} - (1-\omega)m_{2F}\hat{a}_{it}^*\right] + \lambda\delta\mathbb{E}_t\left(\hat{p}_{it+1}^{opt} + m\pi_{t+1}\right), \qquad (190)
\end{aligned}$$

where we used (64) and (66). Its foreign analogue is

$$\hat{p}_{it}^{opt*} = (1-\lambda\delta)\left[m\hat{c}_t^* - \omega m_{2H}\hat{a}_{it}^* - (1-\omega)m_{2F}\hat{a}_{it}\right] + \lambda\delta\mathbb{E}_t\left(\hat{p}_{it+1}^{opt*} + m\pi_{t+1}^*\right). \qquad (191)$$

Using (87), we can rewrite the good-level real reset exchange rate as

$$\begin{aligned}
\hat{q}_{it}^{opt} &= \hat{q}_t + \hat{p}_{it}^{opt*} - \hat{p}_{it}^{opt} \\
&= (1-\lambda\delta)\left[(1-m)\hat{q}_t + \psi\varepsilon_{it}^r\right] + \lambda\delta\mathbb{E}_t\left[\hat{q}_{it+1}^{opt} - m\pi_{t+1}^n - (\hat{q}_{t+1} - \hat{q}_t)\right]. \qquad (192)
\end{aligned}$$

We derive the equations for $\hat{q}_t$ and $\hat{q}_t^{opt}$ from (189) and (192). Noting that $\hat{q}_t = \int_{i=0}^1 \hat{q}_{it} di$, $\hat{q}_t^{opt} = \int_{i=0}^1 \hat{q}_{it}^{opt} di$, and $\int_{i=0}^1 \varepsilon_{it}^r di = 0$, we have the system of equations given by

$$\begin{aligned}
\hat{q}_{it} &= \lambda\hat{q}_{it-1} + \lambda(\hat{q}_t - \hat{q}_{t-1} + \pi_t^n) + (1-\lambda)\hat{q}_{it}^{opt}. && (193) \\
\hat{q}_{it}^{opt} &= (1-\lambda\delta)\left[(1-m)\hat{q}_t + \psi\varepsilon_{it}^r\right] + \lambda\delta\mathbb{E}_t\left[\hat{q}_{it+1}^{opt} - m\pi_{t+1}^n - (\hat{q}_{t+1} - \hat{q}_t)\right], && (194) \\
\hat{q}_t &= \frac{\lambda}{1-\lambda}\pi_t^n + \hat{q}_t^{opt}, && (195) \\
\hat{q}_t^{opt} &= (1-\lambda\delta)(1-m)\hat{q}_t + \lambda\delta\mathbb{E}_t\left[\hat{q}_{t+1}^{opt} - m\pi_{t+1}^n - (\hat{q}_{t+1} - \hat{q}_t)\right], && (196) \\
\hat{q}_t &= \mathbb{E}_t\hat{q}_{t+1} - (\hat{R}_t^n - \mathbb{E}_t\pi_{t+1}^n), && (197) \\
\hat{R}_t^n &= \rho_R\hat{R}_{t-1}^n + (1-\rho_R)\alpha_\pi\pi_t^n + \varepsilon_t^n, && (198)
\end{aligned}$$

where (193), (194), and (198) repeat (189), (192), and (187), respectively. Equation (197) expresses the UIP condition in real terms. The system of the equations is used to compute the first-order autocorrelation and the half-lives of the aggregate and the good-level RERs.

A.14.4 Simulation results

We evaluate the effects of m on ρ_q and ρ_{qi} in the model with the Taylor rule. To conduct this evaluation, we need to calibrate the newly introduced parameters: ρ_R, α_π, and σ_r/σ_n. Here, we assume $\rho_R = 0.80$ and $\alpha_\pi = 1.5$. Since the monetary policy specification differs from that in the main analysis, we assign a different value to the standard deviation ratio: $\sigma_r/\sigma_n = 25$. With this ratio, nominal shocks amplified by monetary policy inertia do not dominate real shocks in the volatility of the good-level RER. Here monetary policy shocks in both countries are serially uncorrelated.

Figure A.2 shows the first-order autocorrelations of the aggregate and good-level RERs as functions of m. The solid and dashed lines in the left panel represent ρ_q and ρ_{qi}, respectively. The right panel plots the ratio of ρ_q to ρ_{qi} against m.

Propositions 2 and 3 no longer hold in the model with the Taylor rule. Engel (2019) shows that, in a two-country sticky-price model with the Taylor rule, the first-order autocorrelation of the aggregate RER under full attention is bounded above by ρ_R and λ, that is, $\rho_q \leq \min[\rho_R, \lambda]$. This condition implies that $\rho_q \leq \lambda$ under full attention. The left panel of Figure A.2 confirms that the first-order autocorrelation of the aggregate RER is consistent with Engel's (2019) finding. At $m = 1$, we have $\rho_q = 0.27$ and $\lambda = 0.34$, implying that $\rho_q < \lambda$. Moreover, our simulation results suggest that $\rho_q = 0.27$ and $\rho_{qi} = 0.34$, meaning that $\rho_q < \rho_{qi}$ at $m = 1$. Therefore, the model with the Taylor rule makes predictions that are difficult to reconcile with the data on aggregate and good-level RERs. In the data, we observe that $\rho_q > \rho_{qi} > \lambda$. However, as shown in Figure A.2, the model with the Taylor rule predicts $\rho_q < \rho_{qi} \simeq \lambda$ at $m = 1$.

Nevertheless, as shown in the left panel of Figure A.2, the model of behavioral inattention continue to have a powerful mechanism for improving the model's fit with the data. The downward-sloping curves for ρ_q and ρ_{qi} suggest that behavioral inattention can generate more persistent RERs. The first-order autocorrelations are substantially improved compared to the case of full attention: $\rho_q = 0.65$ and $\rho_{qi} = 0.41$ when we adopt the estimate of $m = 0.11$ from Table 2. Consequently, the ordering of first-order autocorrelations is consistent with the data: $\rho_q > \rho_{qi} > \lambda$. As shown in the right panel of Figure A.2, the ρ_q to ρ_{qi} ratio is hump-shaped and exceeds unity when m is low.

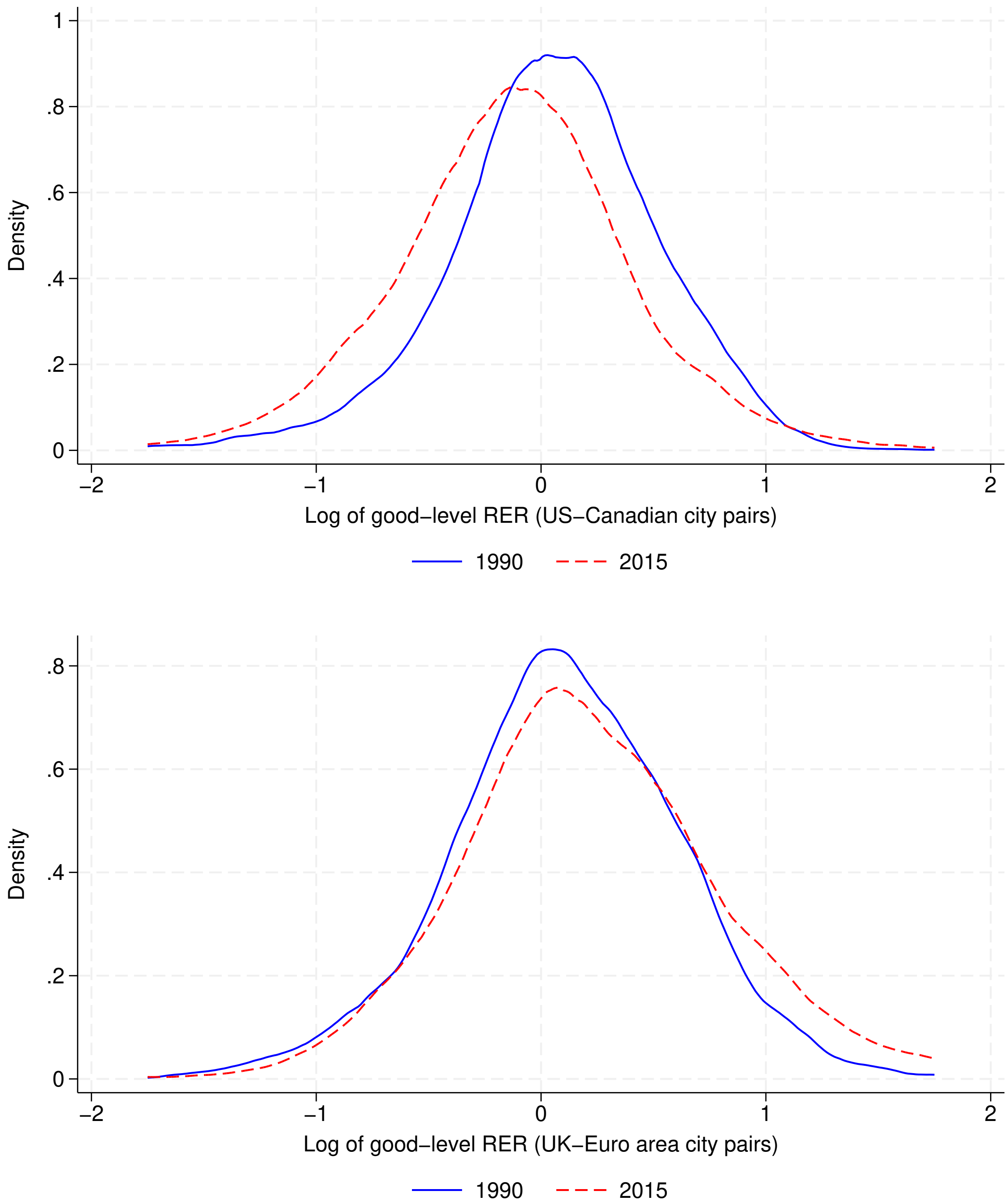
If we introduce persistent monetary policy shocks, the model can reproduce the observed ordering of first-order autocorrelations even under full attention ($m = 1$). However, it still fails to generate sufficiently persistent aggregate and good-level RERs. Assume that ε_t^n follows an AR(1) process: $\varepsilon_t^n = \rho_\varepsilon \varepsilon_{t-1}^n + e_t^n$. Figure A.3 plots the persistence of the aggregate

RER (solid line) and the good-level RER (dashed line) under $\rho_\varepsilon > 0$. In the left panel, we set $\rho_R = \rho_\varepsilon = 0.70$. Even under $m = 1$, the model with $\rho_R = \rho_\varepsilon = 0.70$ predicts that $\rho_q > \rho_{qi} > \lambda$. However, both ρ_q and ρ_{qi} fall short of 0.40 and remain close to $\lambda = 0.34$. In the right panel, we reduce ρ_R to 0.10 and slightly increase ρ_ε to 0.80. This parameterization suggests that ρ_q can be large even under $m = 1$, but ρ_{qi} remains low and close to λ.

Figure A.3 shows that the model of behavioral inattention ($m < 1$) continues to have a powerful mechanism for generating sufficiently persistent RERs at both the aggregate and goods levels. In both panels, the curves for ρ_q and ρ_{qi} are downward sloping with respect to m. Again, if we borrow the estimate of $m = 0.11$ from Table 2, the values of ρ_q and ρ_{qi} are much closer to the data than those under full attention. In the left panel, $\rho_q = 0.77$ and $\rho_{qi} = 0.67$. In the right panel, $\rho_q = 0.80$ and $\rho_{qi} = 0.60$.

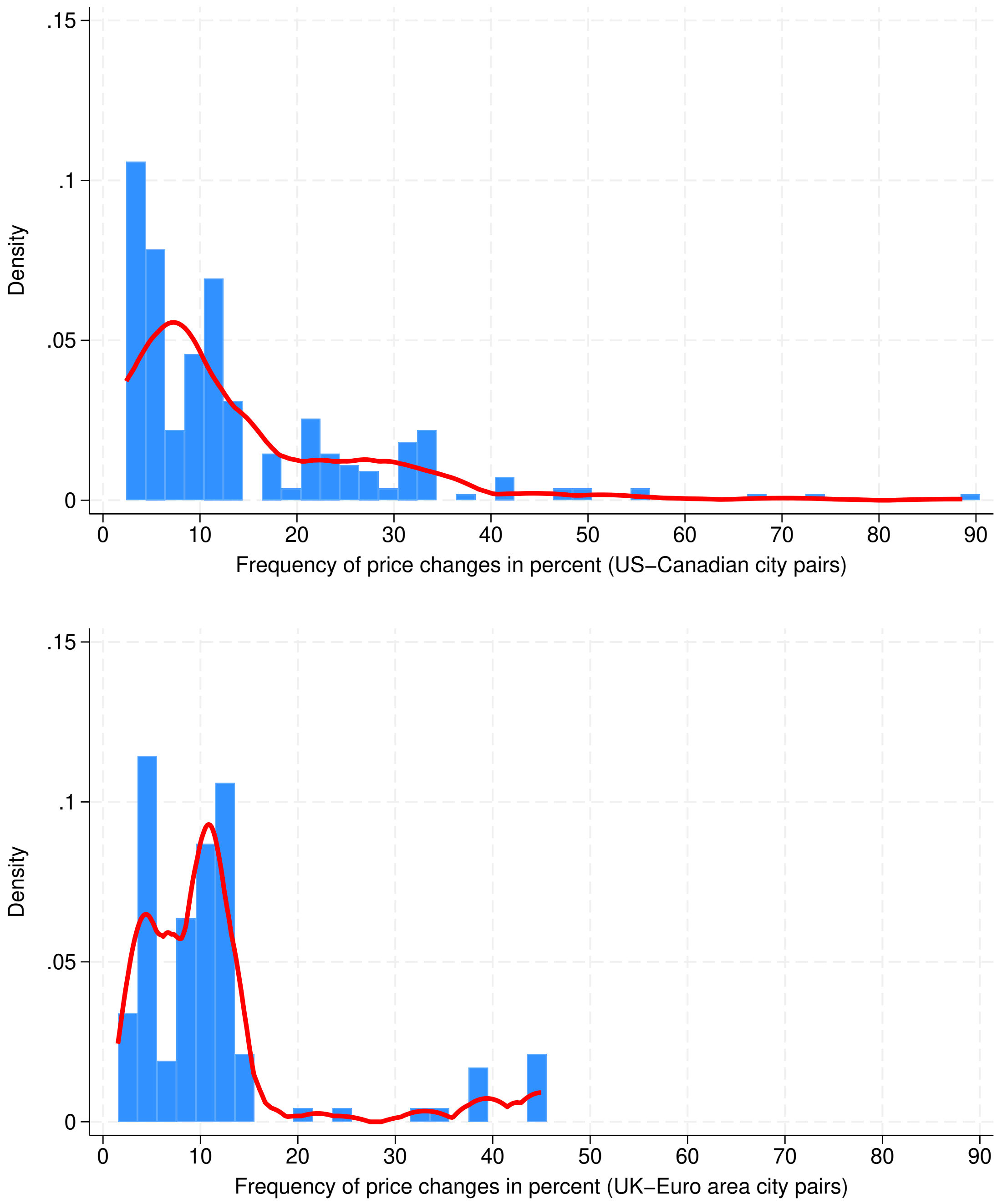
Table A.6 reports half-lives under various specifications of ρ_R, ρ_ε, and m. The upper panel presents the half-lives of the aggregate RER across three parameterizations of $(\rho_R, \rho_\varepsilon)$. For instance, when $(\rho_R, \rho_\varepsilon) = (0.70, 0.70)$, the half-life of the aggregate RER is only 0.76 years under $m = 1$, but rises to between 2.3 and 2.6 years under the estimated degree of attention. The lower panel shows the half-lives of the good-level RER. Again, when $(\rho_R, \rho_\varepsilon) = (0.70, 0.70)$, the half-life of the good-level RER is only 0.65 years under $m = 1$, but increases to 1.5-1.8 years. When $(\rho_R, \rho_\varepsilon) = (0.10, 0.80)$, the half-lives of $\ln q_t$ and $\ln q_{it}$ are roughly consistent with the data. If $(\rho_R, \rho_\varepsilon) = (0.80, 0.10)$, the improvement in predicted half-lives is modest, but the behavioral inattention model performs better than the full attention model.

Figure 1: Empirical distributions of the good-level RERs



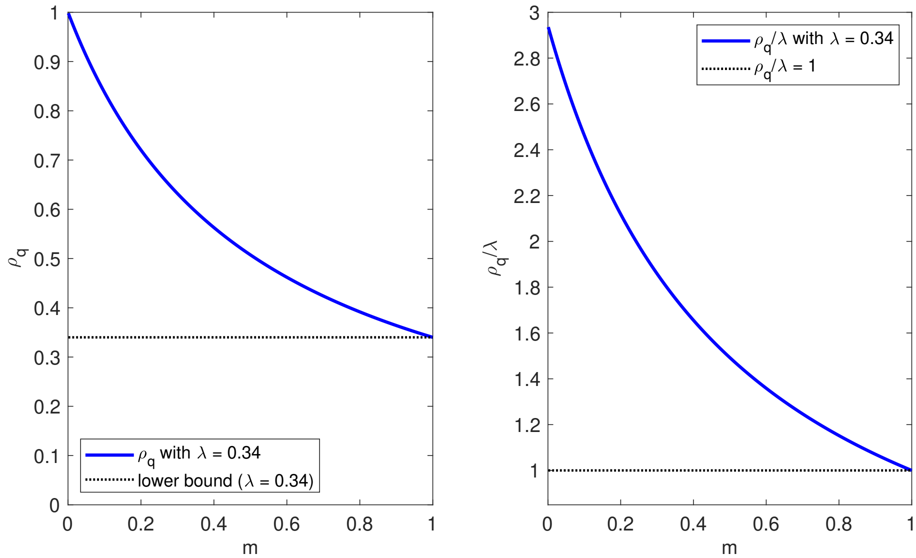
NOTES: The kernel density estimates of the good-level RERs in 1990 and 2015. The upper panel presents the results for the US–Canadian city pairs, while the lower panel displays those for the UK–Euro area city pairs.

Figure 2: Empirical distributions of the monthly frequencies of price changes



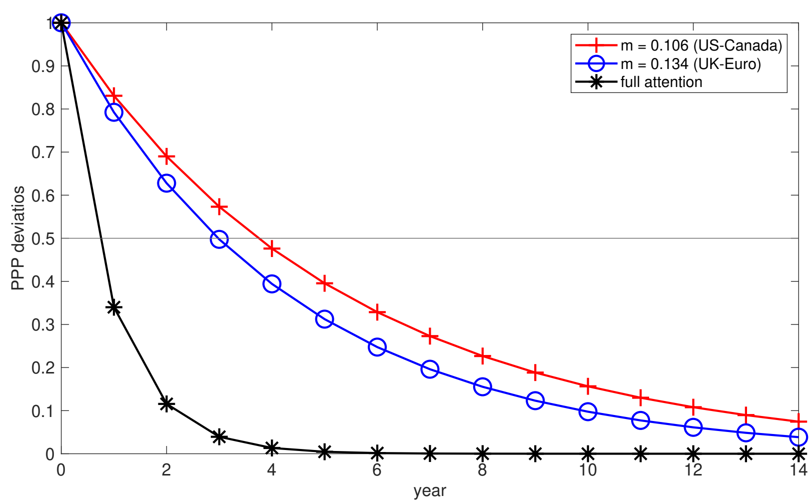
NOTES: The histograms and kernel density estimates of the monthly frequency of price changes used in the analysis. The upper panel presents the distribution for the US–Canadian city pairs, while the lower panel displays that for the UK–Euro area city pairs.

Figure 3: Persistence of the aggregate RER and the ρ_q to λ ratio



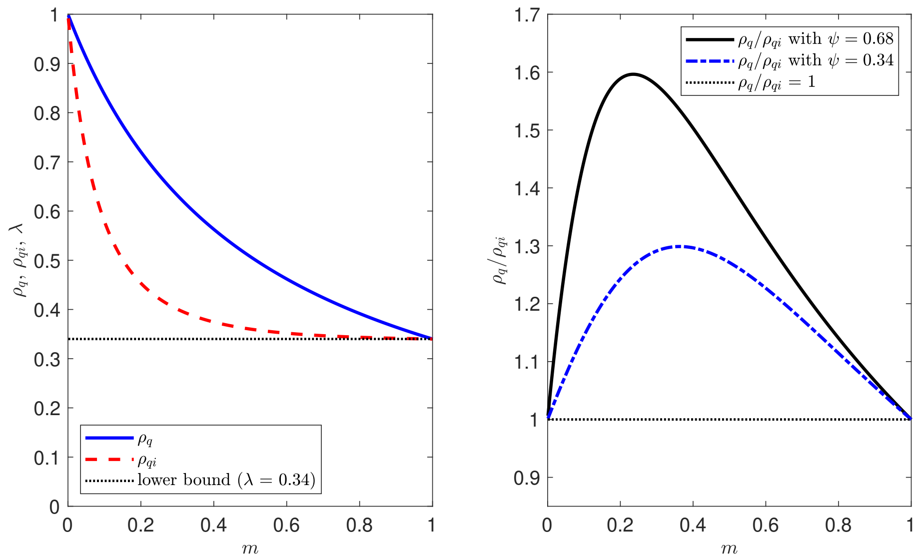
NOTES: The left panel shows the first-order autocorrelations of $\ln q_t$ against m. The dotted line is drawn to compare the persistence under behavioral inattention with that under full attention. The right panel shows the ρ_q to λ ratio against m. The dotted line is included to compare the ρ_q to λ ratio under behavioral inattention with its lower bound of one.

Figure 4: Impulse responses of the aggregate RERs to ε_t^n



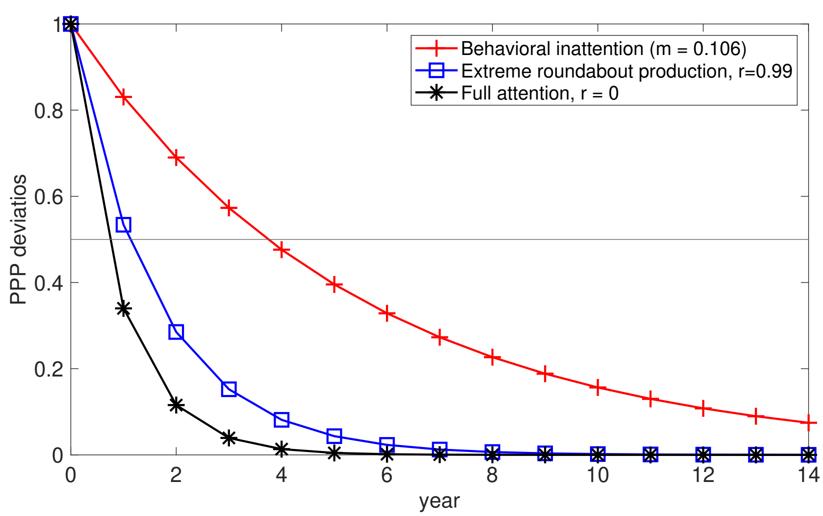
NOTES: The initial responses are normalized to unity. In evaluating the model of behavioral inattention, we use $m = 0.11$ (the estimate from the US–Canadian city pairs) and $m = 0.13$ (the estimate from the UK–Euro area city pairs).

Figure 5: Persistence of the aggregate and the good-level RERs and the ρ_q to ρ_{qi} ratio



NOTES: The left panel compares the first-order autocorrelations of $\ln q_t$ and $\ln q_{it}$. The dotted line is drawn to compare the persistence under behavioral inattention with that under full attention. The right panel shows the ρ_q to ρ_{qi} ratio against m. The solid line represents the ρ_q to ρ_{qi} ratio under the baseline value of $\psi = 0.68$, whereas the dashed-dotted line corresponds to the case where $\psi = 0.34$, half of the baseline value. The dotted line is included to compare the ρ_q to ρ_{qi} ratio under behavioral inattention with its lower bound of one.

Figure 6: Impulse response of the aggregate RERs: Roundabout production



NOTES: The initial responses are normalized to unity. In evaluating the model of roundabout production, we use an extreme value of $r = 0.99$ to allow for the maximum impact of strategic complementarity in the form of roundabout production. In drawing the impulse response function for the model of behavioral inattention, we use the estimate of $m = 0.11$ from the US–Canadian city pairs.

Table 1: Estimation results of (36) under a common λ

	Dependent variable: $\ln \tilde{q}_{ijt}$ with a common λ							
	US–Canadian city pairs				UK–Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\ln \tilde{q}_t$	0.844	0.802	0.812	0.806	0.856	0.851	0.853	0.868
	(0.030)	(0.028)	(0.029)	(0.029)	(0.042)	(0.043)	(0.042)	(0.042)
Observations	389,500	389,500	389,500	389,500	214,115	214,115	213,064	213,064
Adj. R^2	0.225	0.262	0.225	0.262	0.265	0.323	0.265	0.324
city-pair FE	N	Y	N	Y	N	Y	N	Y
Control for η_t^r	N	N	Y	Y	N	N	Y	Y
	The implied degree of attention from the regression							
$\hat{m}$	0.156	0.198	0.188	0.194	0.144	0.149	0.147	0.132

NOTES: The left panel reports estimation results based on 274 items and 64 US–Canadian city pairs, while the right panel presents results based on 301 items and 36 UK–Euro area city pairs. The regressions use data from 1990 to 2015. The dependent variable is $\ln \tilde{q}_{ijt}$ constructed using a common λ where $\lambda = 0.34$. The table reports the regression coefficients on $\ln \tilde{q}_t = (1-\lambda)\ln q_t$. Standard errors, clustered by goods, are shown in parentheses below the coefficients. All specifications include good-specific fixed effects. Specifications (2) and (4) further include city-pair fixed effects, while specifications (3) and (4) control for the log difference in real GDP per hour worked between two currency areas. For US–Canadian city pairs (left panel), the regressions also include dummy variables to control for timing differences in price surveys for Calgary between 2003 and 2014. Each column reports the estimated values of m in the bottom row. "Adj. R^2" denotes the adjusted R-squared.

Table 2: Estimation results of (36) under heterogeneity in price stickiness

	Dependent variable: $\ln \tilde{q}_{ijt}$ with λ_i							
	US–Canadian city pairs				UK–Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\ln \tilde{q}_t^i$	0.894	0.862	0.883	0.880	0.866	0.834	0.864	0.840
	(0.029)	(0.028)	(0.032)	(0.033)	(0.047)	(0.049)	(0.048)	(0.049)
Observations	389,500	389,500	389,500	389,500	171,606	171,606	170,750	170,750
Adj. R^2	0.256	0.294	0.256	0.294	0.246	0.295	0.247	0.295
city-pair FE	N	Y	N	Y	N	Y	N	Y
Control for η_t^r	N	N	Y	Y	N	N	Y	Y
	The implied degree of attention from the regression							
$\hat{m}$	0.106	0.138	0.117	0.120	0.134	0.166	0.136	0.160

NOTES: The left panel shows estimation results based on 274 items and 64 US–Canadian city pairs, while the right panel is based on 236 items and 36 UK–Euro area city pairs. The dependent variable is $\ln \tilde{q}_{ijt}$ constructed using λ_i (i.e., $\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda_i[\ln q_{ijt-1} - \ln(S_t/S_{t-1})]$). The calibrated λ_i follows previous studies: Nakamura and Steinsson (2008) for the US–Canadian city pairs, and Gautier et al. (2024) for the UK–Euro area city pairs. The table reports the coefficients on $\ln \tilde{q}_t^i = (1 - \lambda_i) \ln q_t$. See the notes to Table 1 for further details.

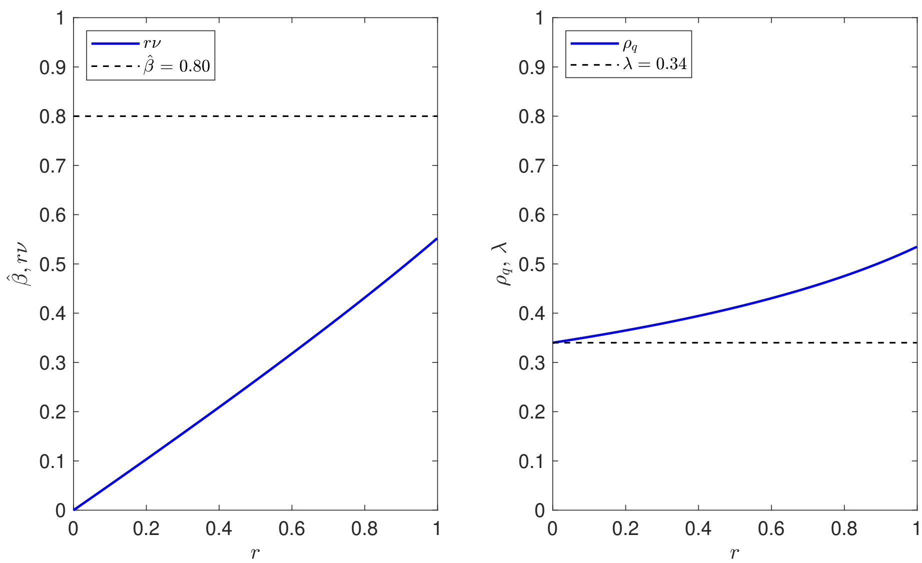
Table 3: Half-lives implied by the estimated degree of attention

	Half-lives of the aggregate RER		
	Predicted half-life	95% CI	Data
US–Canadian city pairs			
$\hat{m} = 0.156$	2.620	[1.989, 4.010]	4.922
$\hat{m} = 0.106$	3.704	[2.524, 7.605]	
UK–Euro area city pairs			
$\hat{m} = 0.144$	2.812	[1.903, 6.129]	2.398
$\hat{m} = 0.134$	2.998	[1.905, 8.868]	
	Half-lives of the good-level RER		
	Predicted half-life	95% CI	Data
US–Canadian city pairs			
$\hat{m} = 0.156$	0.984	[0.851, 1.292]	1.606
$\hat{m} = 0.106$	1.223	[0.963, 2.110]	
UK–Euro area city pairs			
$\hat{m} = 0.144$	1.026	[0.834, 1.773]	1.182
$\hat{m} = 0.134$	1.066	[0.834, 2.399]	

NOTES: The table reports the half-lives predicted by the model of behavioral inattention. The unit of the half-lives is years, and the half-life under full attention is 0.64 years. The upper panel presents the half-lives of the aggregate RER, while the lower panel shows those of the good-level RER. To calculate the predicted half-lives in the table, we use the calibrated values of $\tau = 0.74$, $\varepsilon = 4$, $\sigma_r/\sigma_n = 5$, and $\delta = 0.98$. In all calculations, λ is held constant at $\lambda = 0.34$.

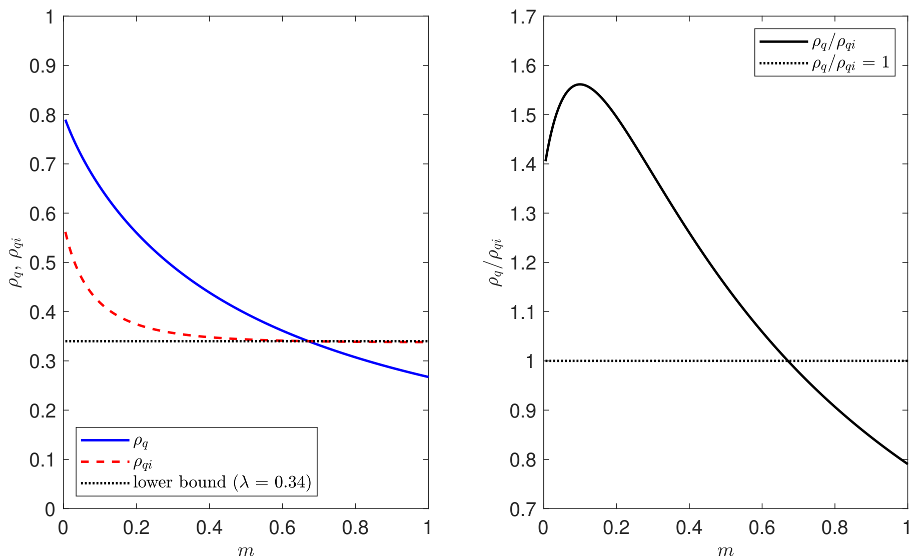
In each panel, we report the half-lives for the US–Canadian city pairs and the UK–Euro area city pairs. The first column of the table reports the half-lives predicted by the model under behavioral inattention, and the second column provides their 95 percent confidence intervals, denoted as "95% CI." We compute the half-lives from $\hat{m}$ and the 95 percent confidence intervals of $\hat{m}$, based on specification (1) of Tables 1 and 2. For comparison, the rightmost column presents the half-lives estimated from the EIU data. See the main text for details on the estimation of the half-lives from the EIU data.

Figure A.1: Regression coefficients and the first-order autocorrelations predicted by the model of roundabout production



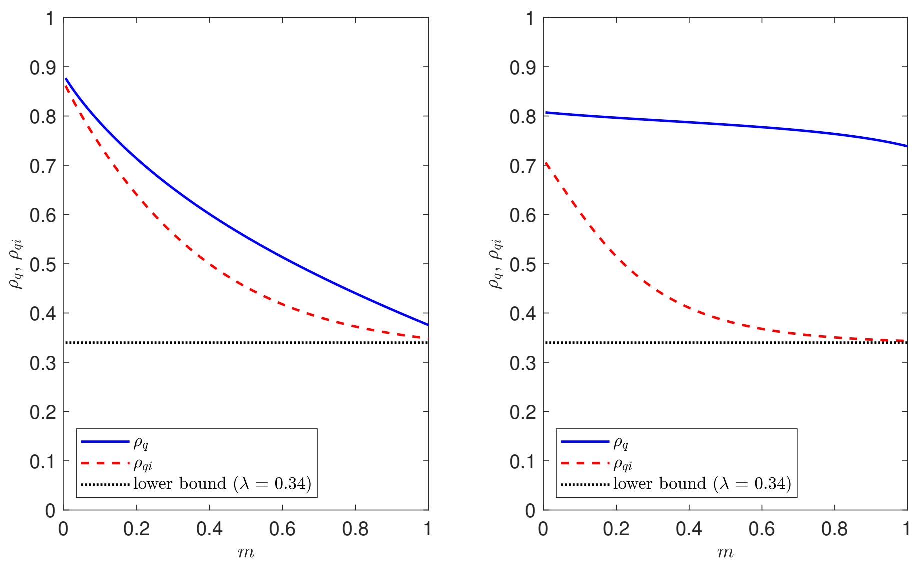
NOTES: The left panel presents the regression coefficient predicted by the model of roundabout production (solid line), which varies with the degree of roundabout production r. The dashed line represents the estimated coefficient from the data, included for comparison. To be conservative, we use the lowest estimate of β from Tables 1 and 2. The right panel reports the first-order autocorrelation implied by the model of roundabout production (solid line), plotted against r. The dashed line represents the persistence under the baseline model without roundabout production ($r = 0$), included for comparison with the case $0 < r < 1$.

Figure A.2: Persistence of the aggregate and the good-level RERs and the ρ_q to ρ_{qi} ratio in the model with the Taylor rule



NOTES: The left panel compares the first-order autocorrelations of $\ln q_t$ and $\ln q_{it}$ in the model with the Taylor rule. In the figure, we set $\rho_R = 0.80$ and $\rho_\varepsilon = 0$. Under full attention, ρ_q and ρ_{qi} decline to 0.27 and 0.34, respectively. The dotted line represents the degree of price stickiness and is included for comparison with the persistence of $\ln q_t$ and $\ln q_{it}$. The right panel plots the ratio of ρ_q to ρ_{qi} against m (solid line). The dotted line is included to compare the ρ_q to ρ_{qi} ratio under behavioral inattention with the case of $\rho_q/\rho_{qi} = 1$.

Figure A.3: Persistence of the aggregate and the good-level RERs in the model with the Taylor rule



NOTES: Both panels compare the first-order autocorrelations of $\ln q_t$ and $\ln q_{it}$ in the model with the Taylor rule. The dotted lines in each panel represent the degree of price stickiness and is included for comparison with the persistence of $\ln q_t$ and $\ln q_{it}$. We set $\rho_R = \rho_\varepsilon = 0.70$ in the left panel and $\rho_R = 0.10$ and $\rho_\varepsilon = 0.80$ in the right panel. In the left panel, full attention under $\rho_R = \rho_\varepsilon = 0.70$ yields $\rho_q = 0.40$ and $\rho_{qi} = 0.35$, leading to $\rho_q > \rho_{qi} > \lambda$ $(= 0.34)$. In the right panel, full attention under $\rho_R = 0.10$ and $\rho_\varepsilon = 0.80$ generates $\rho_q = 0.74$ and $\rho_{qi} = 0.34$, such that $\rho_q \gg \rho_{qi} \simeq \lambda$. Under behavioral inattention ($m = 0.11$), both panels indicate that values of ρ_q and ρ_{qi} are much more consistent with the empirical evidence.

Table A.1: Descriptive statistics for $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t$ (or $\ln \tilde{q}_t^i$)

	Homogeneity in price stickiness			
	US–Canadian city pairs		UK–Euro area city pairs	
	Mean	Standard deviation	Mean	Standard deviation
$\ln \tilde{q}_{ijt}$	-0.028	0.350	0.049	0.388
$\ln \tilde{q}_t$	-0.135	0.098	0.227	0.073
Observations	389,500		214,115	
	Heterogeneity in price stickiness			
	US–Canadian city pairs		UK–Euro area city pairs	
	Mean	Standard deviation	Mean	Standard deviation
$\ln \tilde{q}_{ijt}$	-0.033	0.355	0.026	0.365
$\ln \tilde{q}_t^i$	-0.138	0.110	0.230	0.104
Observations	389,500		171,606	

NOTES: The table reports descriptive statistics for the variables used in the regressions. The upper panel shows statistics for $\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda[\ln q_{ijt-1} - \ln(S_t/S_{t-1})]$ and $\ln \tilde{q}_t = (1-\lambda)\ln q_t$, which are used in the regressions under the assumption of a common λ. The lower panel presents statistics for $\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda_i[\ln q_{ijt-1} - \ln(S_t/S_{t-1})]$ and $\ln \tilde{q}_t^i = (1-\lambda_i)\ln q_t$, used in the regressions allowing for heterogeneity in price stickiness. The left panel provides statistics for the US–Canadian city pairs, and the right panel provides statistics for the UK–Euro area city pairs. The calibrated value of λ is 0.34 for both the US–Canadian and the UK–Euro area city pairs. The calibrated values of λ_i are from Nakamura and Steinsson (2008) for the US–Canadian city pairs and from Gautier et al. (2024) for the UK–Euro area city city pairs.

Table A.2: Estimation results under general CRRA preferences

	Dependent variable: $\ln \tilde{\tilde{q}}_{ijt}$							
	US–Canadian city pairs				UK–Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
λ	common	common	good-specific	good-specific	common	common	good-specific	good-specific
$\ln \tilde{\tilde{q}}_t$	0.782	0.804	0.853	0.870	0.784	0.768	0.791	0.735
	(0.034)	(0.034)	(0.041)	(0.041)	(0.050)	(0.052)	(0.063)	(0.065)
Observations	371,347	371,347	371,347	371,347	201,578	201,578	161,931	161,931
Adj. R^2	0.179	0.206	0.242	0.269	0.217	0.261	0.201	0.237
city-pair FE	N	Y	N	Y	N	Y	N	Y
	The implied degree of attention from the regression							
m	0.218	0.200	0.147	0.130	0.216	0.232	0.209	0.265

NOTES: The estimation equation is (115). The left panel reports the estimation results based on 274 items and 64 US–Canadian city pairs. The right panel shows the estimation results based on 301 items and 36 UK–Euro area city pairs for specifications (1) and (2), and 236 items and 36 UK–Euro area city pairs for specifications (3) and (4). The dependent variable is $\ln \tilde{\tilde{q}}_{ijt} = \ln \tilde{q}_{ijt} - \lambda\delta \ln \tilde{q}_{ijt+1}$ (or $\ln \tilde{\tilde{q}}_{ijt} = \ln \tilde{q}_{ijt} - \lambda_i\delta \ln \tilde{q}_{ijt+1}$), where the degree of price stickiness is calibrated either as a common value (λ) or as a good-specific value (λ_i). In each panel, specifications (1) and (2) report results using the common λ, while specifications (3) and (4) present results allowing for heterogeneity in price stickiness. The table reports the regression coefficients on $\ln \tilde{\tilde{q}}_t = \ln \tilde{q}_t - \lambda\delta \ln \tilde{q}_{t+1}$ (or $\ln \tilde{\tilde{q}}^i_t = \ln \tilde{q}_t - \lambda_i\delta \ln \tilde{q}^i_{t+1}$). We use $\ln \tilde{q}_{t-1}$ (or $\ln \tilde{q}^i_{t-1}$) as an instrument to estimate the parameters. See the notes to Tables 1 and 2 for further details.

Table A.3: Estimation results of (38)

	Dependent variable: $\ln q_{ijt}$							
	US–Canadian city pairs				UK–Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\ln q_t$	0.302	0.300	0.264	0.261	0.077	0.107	0.074	0.103
	(0.021)	(0.020)	(0.025)	(0.022)	(0.029)	(0.026)	(0.029)	(0.027)
Observations	371,347	371,347	371,347	371,347	202,621	202,621	201,578	201,578
P-value for Hansen's J-test	0.932	0.925	0.922	0.926	0.489	0.485	0.488	0.496
Control for η_t^r	N	N	Y	Y	N	N	Y	Y

NOTES: The estimation equation is (38). Following Arellano and Bond (1991), we regress $\Delta \ln q_{ijt}$ on $\Delta \ln q_t$, along with $\Delta \ln q_{ijt-1}$, $\Delta^2 \ln S_t$, and other control variables. In specifications (2) and (4), we impose parameter restrictions such that the coefficients on $\ln q_{ijt-1}$ and $\Delta \ln S_t$ are equal. In specifications (3) and (4), we control for differences in the log of real GDP per hour worked. As suggested by Arellano and Bond (1991), the levels of the lagged dependent variables are used as instruments, depending on the period of observation. "P-value for Hansen's J-test" denotes the p-value for the test of overidentifying restrictions. The degrees of freedom for the chi-squared distribution under the null are 299 in specifications (1) and (3), and 300 in specifications (2) and (4). See the notes to Table 1 for further details.

Table A.4: Estimation results of (36) under a common λ: Splitting samples

	Dependent variable: $\ln \tilde{q}_{ijt}$ constructed with λ							
	US–Canadian city pairs				UK–Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\ln \tilde{q}_t$	0.816	0.877	0.841	0.843	0.854	0.856	0.861	0.838
	(0.031)	(0.033)	(0.030)	(0.030)	(0.042)	(0.046)	(0.042)	(0.043)
Observations	192,253	197,247	370,285	365,603	102,238	111,877	209,396	202,305
Adj. R^2	0.228	0.24	0.225	0.226	0.265	0.297	0.27	0.269
city-pair FE	N	N	N	N	N	N	N	N
Control for η_t^r	N	N	N	N	N	N	N	N
	The implied degree of attention from the regression							
$\hat{m}$	0.184	0.123	0.159	0.157	0.146	0.144	0.139	0.162

NOTES: Each panel presents estimation results based on subsamples split by the distance between two cities. The left panel reports the results for US–Canadian city pairs, while the right panel reports the results for UK–Euro area city pairs. All specifications correspond to specification (1) in Table 1, but the samples vary across specifications (1)-(4).

In specifications (1) and (2), the sample is divided into city pairs below and above the median distance, respectively. Specification (1) includes city pairs with distances below the median, while specification (2) includes those above. Specifications (3) and (4) exclude potential outliers: specification (3) omits city pairs below the 5th percentile of distance, and specification (4) omits those above the 95th percentile. See the notes to Table 1 for further details.

Table A.5: Estimation results of (36) under heterogeneity in price stickiness: Splitting samples

	Dependent variable: $\ln \tilde{q}_{ijt}$ constructed with λ_i							
	US–Canadian city pairs				UK–Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\ln \tilde{q}_t^i$	0.866	0.928	0.898	0.890	0.846	0.876	0.872	0.852
	(0.032)	(0.031)	(0.029)	(0.029)	(0.050)	(0.052)	(0.048)	(0.047)
Observations	192,253	197,247	370,285	365,603	81,967	89,639	167,805	162,135
Adj. R^2	0.257	0.271	0.257	0.257	0.240	0.284	0.252	0.255
city-pair FE	N	N	N	N	N	N	N	N
Control for η_t^r	N	N	N	N	N	N	N	N
	The implied degree of attention from the regression							
$\hat{m}$	0.134	0.072	0.102	0.110	0.154	0.124	0.128	0.148

NOTES: The dependent variable $\ln \tilde{q}_{ijt}$ and the explanatory variable $\ln \tilde{q}_t^i$ are constructed using λ_i. See the notes to Tables 1, 2, and A.4 for further details.

Table A.6: Half-lives predicted by the model with the Taylor rule

	Half-lives of the aggregate RERs			
$(\rho_R, \rho_\varepsilon)$	(0.70, 0.70)	(0.10, 0.80)	(0.80, 0.10)	Data
$m = 1$	0.761	2.288	0.561	
US–Canadian city pairs				
$m = 0.156$	2.258	3.082	1.416	4.922
$m = 0.106$	2.604	3.127	1.671	
UK–Euro area city pairs				
$m = 0.144$	2.330	3.093	1.467	2.398
$m = 0.134$	2.395	3.102	1.514	
	Half-lives of the good-level RERs			
$(\rho_R, \rho_\varepsilon)$	(0.70, 0.70)	(0.10, 0.80)	(0.80, 0.10)	Data
$m = 1$	0.654	0.648	0.640	
US–Canadian city pairs				
$m = 0.156$	1.451	1.163	0.761	1.606
$m = 0.106$	1.753	1.350	0.827	
UK–Euro area city pairs				
$m = 0.144$	1.513	1.202	0.773	1.182
$m = 0.134$	1.568	1.237	0.785	

NOTES: The table reports the half-lives predicted by the model with a Taylor rule. The unit of half-lives is years. The upper panel presents the half-lives of the aggregate RER, while the lower panel shows those of the good-level RER. To calculate the predicted half-lives in the table, we use the calibrated values of $\tau = 0.74$, $\varepsilon = 4$, $\sigma_r/\sigma_n = 25$, and $\delta = 0.98$. In all calculations, λ is held constant at $\lambda = 0.34$.

In each panel, we report the half-lives based on parameters for monetary policy (ρ_R and ρ_ε). The half-lives are computed using the estimated values of m based on specification (1) in Tables 1 and 2, as well as those under full attention ($m = 1$).