

# A Behavioral Explanation for the Puzzling Persistence of the Aggregate Real Exchange Rate\*

Mario J. Crucini<sup>†</sup>, Mototsugu Shintani<sup>‡</sup> and Takayuki Tsuruga<sup>§</sup>

This version: August 2022

## Abstract

At the aggregate level, the evidence that deviations from purchasing power parity (PPP) are too persistent to be explained solely by nominal rigidities has long been a puzzle (Rogoff, 1996). Another puzzle from the micro price evidence of the law of one price (LOP), which is the basic building block of PPP, is that LOP deviations are less persistent than PPP deviations. To reconcile the empirical evidence, we adapt the model of behavioral inattention in Gabaix (2014, 2020) to a simple two-country sticky-price model. We propose a simple test of behavioral inattention and find strong evidence in its favor using micro price data. Calibrating behavioral inattention with our estimates, we show that our model reconciles the two puzzles relating to the PPP and LOP. First, the PPP deviations are more than twice as persistent as PPP deviations implied solely by nominal rigidities. Second, the LOP deviations decrease to less than two-thirds of the PPP deviations in the degree of persistence.

JEL Classification: E31; F31; D40

*Keywords:* Real exchange rates, Law of one price, Rational inattention, Trade cost

---

\*We thank David Cook, Mick Devereux, Ipeei Fujiwara, Yasuo Hirose, Jinill Kim, Akihisa Shibata, Kwansho Shin and conference and seminar participants at the 5th International Macroeconomics and Finance Conference, Winter Meeting of the Econometric Society, Keio University, Otaru University of Commerce, Renmin University of China, Tokyo Institute of Technology, The University of Tokyo, Vanderbilt University, and Yokohama City University for their helpful comments and suggestions. Mototsugu Shintani and Takayuki Tsuruga gratefully acknowledge the financial support of Grant-in-aid for Scientific Research (JP17H02510, JP18K01684, JP20H05631, JP20H05633, and JP21H04397), the Joint Usage/Research Center at the Institute of Social and Economic Research (Osaka University), the Kyoto Institute of Economic Research (Kyoto University), and the International Joint Research Promotion Program in Osaka University.

<sup>†</sup>Krannert School of Management, Purdue University and NBER; e-mail: [mcrucini@purdue.edu](mailto:mcrucini@purdue.edu).

<sup>‡</sup>Faculty of Economics, The University of Tokyo; e-mail: [shintani@e.u-tokyo.ac.jp](mailto:shintani@e.u-tokyo.ac.jp).

<sup>§</sup>Institute of Social and Economic Research, Osaka University and CAMA; e-mail: [tsuruga@iser.osaka-u.ac.jp](mailto:tsuruga@iser.osaka-u.ac.jp).

# 1 Introduction

It is well known that the aggregate real exchange rate, namely the deviation from purchasing power parity (PPP), is highly persistent. In a description of this empirical anomaly known as the PPP puzzle, Rogoff (1996) states, “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). A closely related empirical fact of real exchange rates is the gap in persistence between the PPP deviations and the deviations from the law of one price (LOP) as the basic building block for PPP. For example, Imbs et al. (2005) and Carvalho and Nechio (2011) argue that the good-level real exchange rate (the LOP deviations) is likely to be much less persistent than the aggregate real exchange rate (the PPP deviations).<sup>1</sup> These previous studies emphasize the role of heterogeneity in the speed of price adjustment. As Imbs et al. (2005) argued, “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 aim to explain two empirical anomalies simultaneously: (1) the gap between the observed persistence of the PPP deviations and the persistence predicted from the sticky-price model (e.g., Rogoff, 1996); and (2) the gap between the observed persistence of the PPP deviations and the LOP deviations (e.g., Imbs et al., 2005). To this end, we incorporate behavioral inattention along the lines of Gabaix (2014, 2020) into a simple two-country sticky-price model. In this framework, firm managers bear the cost of paying attention to the aggregate component of the marginal cost of their products. As a result, the full attention to the state of the economy is no longer optimal when firms choose the prices of goods.

The key to solving the PPP puzzle is the complementarity between the PPP and LOP deviations. After deriving a reduced-form solution for the LOP deviations, we show that the LOP deviations are affected by the PPP deviations when firms pay only partial attention to marginal cost. Thus, an increase in the persistence of the PPP deviations makes the LOP deviations more persistent. At the same time, through the aggregation, the more persistent LOP deviations lead to more persistent PPP deviations, further strengthening the link between the PPP and LOP deviations.

The reduced-form solution leads to a direct testable implication. Using micro price data in

---

<sup>1</sup>See Crucini and Shintani (2008) for a comprehensive empirical analysis of the persistence in the LOP deviations.

the US, Canada, and European countries, we implement a simple test for the null hypothesis of full attention against an alternative hypothesis of partial attention. This test is equivalent to asking whether the good-level real exchange rate is uncorrelated with the aggregate real exchange rate after controlling for the common driving forces, such as the nominal exchange rate and country-specific productivity. Using various specifications, we strongly reject the null in favor of our proposed model of behavioral inattention. We also find the estimated degree of attention to be around 0.2, much less than the value of 1.0 under full attention.

In the first theoretical result, our model of behavioral inattention ensures that the persistence in the aggregate real exchange rate exceeds the persistence implied solely by nominal rigidities. While the setting differs greatly, this mechanism is consistent with Ball and Romer (1990) and Woodford (2003). They show that even small frictions in nominal price adjustment lead to a persistent output gap when real rigidities or strategic complementarities are present. In our model of behavioral inattention, only small nominal frictions are needed to generate a highly persistent aggregate real exchange rate. Based on our estimates of the degree of attention, the aggregate real exchange rate is more than twice as persistent as that implied solely by nominal rigidities. In terms of half-lives of the aggregate real exchange rate, the estimated degrees of attention suggest that the half-life becomes longer by 1.98–3.06 (1.40–1.90) years for the real exchange rate between the US and Canada (the UK and Euro area).

In the second theoretical result, our model explains the gap between the highly persistent PPP deviations and the less persistent LOP deviations. This gap arises from the combination of complementarity and the presence of idiosyncratic real shocks to the individual price of goods. We show that both the PPP and the LOP deviations are more persistent when complementarities are present. In contrast, real shocks at the good level (but not country-specific real shocks) reduce persistence only for the LOP deviations and not for the PPP deviations. This is because the aggregation across goods eliminates the effect of real shocks at the good level. As a result, our estimates of the degree of attention imply a substantial gap in persistence between the PPP and LOP deviations. Our model predicts that the persistence of the LOP deviations decreases to less than two-thirds of the persistence of the PPP deviations. In terms of half-lives, the estimated degrees of attention imply that the half-life of the LOP deviations is shorter by 1.64–2.48 (1.18–1.57) years than that of the PPP deviations between the US and Canada (the UK and Euro area).

The first result relates to an extensive literature that has contributed to a better understanding of persistent aggregate real exchange rates. For instance, Chari et al. (2002) argue

that while the sticky-price model can explain the volatility of the aggregate real exchange rate, the persistence is less than that suggested by the data. Benigno (2004) emphasizes the role of monetary policy rules rather than the degree of price stickiness in accounting for the persistent aggregate real exchange rate. Later, Engel (2019) revisits Benigno (2004) and argues for the importance of both monetary policy rules and price stickiness. Blanco and Cravino (2020) introduce a concept of the real exchange rate using only newly reset prices and find that this “reset” exchange rate explains almost all fluctuations in the aggregate real exchange rate. Our solution also closely relates to Bergin and Feenstra (2001) and Kehoe and Midrigan (2007), who introduce strategic complementarity in pricing to two-country sticky-price models in explaining the high persistence of real exchange rates. Our explanation of the PPP puzzle does not contradict these explanations. Instead, we focus on the importance of behavioral inattention in firms’ pricing.

A standard explanation for the second result is the heterogeneity in price adjustment at a good level, which generates a positive bias when aggregating prices. Imbs et al. (2005) point out a positive aggregation bias in dynamic heterogeneous panels, and Carvalho and Nechio (2011) consider the theoretical implication of aggregation based on the sticky-price model in which the degree of price stickiness differs across sectors. Indeed, both statistical aggregation bias and multisector sticky-price models would help toward increasing the persistence of real exchange rates. In contrast, our solution can explain the gap even if the persistence of the LOP deviations is restricted to be common across goods. In this sense, the mechanism in our paper further enhances the ability of existing workhorse models in the macroeconomics literature. Furthermore, our model with behavioral inattention can also explain the following two related findings: the LOP deviations is more persistent than the degrees of price stickiness imply (Kehoe and Mirdigan 2007); and the LOP deviations is as persistent as the PPP deviations when we focus only on macroeconomic shocks (Bergin et al. 2013).

The remainder of the paper is structured as follows. In Section 2, we present a simple two-country open economy model with Calvo pricing and introduce behavioral inattention. Section 3 introduces the reduced-form solution for the LOP deviations and discusses the implications of behavioral inattention. In Section 4, we implement a test of behavioral inattention and quantify its importance. In Section 5, we assess how much the estimated degree of behavioral inattention can improve model predictions. Section 6 concludes.

## 2 The model

The economy consists of two countries. To be specific, we express the US and Canada as the home and foreign countries, respectively. 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  $i$  and brands are indexed by  $z$ . For each good, US brands are indexed by  $z \in [0, 1/2]$  and Canadian brands are indexed by  $z \in (1/2, 1]$ .

We assume that US and Canadian consumers have identical preferences over brands of a particular good and also across goods in the aggregate consumption basket. US preferences over the brands of good  $i$  are given by the constant elasticity-of-substitution (CES) index for good  $i \in [0, 1]$ . The US consumption of good  $i$  is  $c_{it} = \left[ \int_{z=0}^1 c_{it}(z)^{\frac{\varepsilon-1}{\varepsilon}} dz \right]^{\frac{\varepsilon}{\varepsilon-1}}$  and the aggregation across goods gives 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 equations  $c_{it}^* = \left[ \int_{z=0}^1 c_{it}^*(z)^{\frac{\varepsilon-1}{\varepsilon}} dz \right]^{\frac{\varepsilon}{\varepsilon-1}}$  and  $c_t^* = \left[ \int_{i=0}^1 c_{it}^*{}^{\frac{\varepsilon-1}{\varepsilon}} di \right]^{\frac{\varepsilon}{\varepsilon-1}}$ .

### 2.1 Households

The objective of the US agent 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 the budget constraint given by:

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

and a cash-in-advance (CIA) constraint,  $M_t \geq P_t c_t$ . Here,  $\mathbb{E}_0(\cdot)$  denotes the expectation operator conditional on the information available in period 0,  $\delta \in (0, 1)$ , and  $\chi > 0$ . In addition, we suppress the state contingencies for notational convenience. The left-hand side of (1) represents the nominal value of the total wealth of the household. The household allocates the wealth into money balances  $M_t$  for the purchase of consumption goods and 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 supplies hours worked  $n_t$ , receives nominal wages  $W_t$  for hours worked, 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,  $\Pi_t$ . 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 is a CES aggregate over 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 the period  $t$ . The CIA always binds with equality in equilibrium.

Canadian households solve the analogous maximization problem. We assume complete markets for state-contingent financial claims across the US and Canada and the financial claims are denominated in US dollars. Thus, we convert their US dollar bond holdings into Canadian dollars at the spot nominal exchange rate,  $S_t$ . The Canadian households are subject to budget constraint,

$$M_t^* + \frac{\mathbb{E}_t(\Delta_{t,t+1}B_{t+1}^*)}{S_t} = W_t^*n_t^* + \frac{B_t^*}{S_t} + M_{t-1}^* - P_{t-1}^*c_{t-1}^* + T_t^* + \Pi_t^*. \quad (2)$$

and the CIA constraint,  $M_t^* \geq P_t^*c_t^*$ .

The first-order conditions are standard. For the US households, we have  $W_t/P_t = \chi c_t$  and  $\Delta_{t,t+1} = \delta[(c_{t+1}/c_t)^{-1}(P_t/P_{t+1})]$ . For Canadian households, we have  $W_t^*/P_t^* = \chi c_t^*$  and  $\Delta_{t,t+1} = \delta[(c_{t+1}^*/c_t^*)^{-1}S_tP_t^*/(S_{t+1}P_{t+1}^*)]$ . The consumption Euler equations differ because Canadians buy state-contingent bonds denominated in US dollars.

Define the aggregate real exchange rate as  $q_t = S_tP_t^*/P_t$ . The Euler equations imply  $q_{t+1}(c_{t+1}^*/c_{t+1}) = q_t(c_t^*/c_t) = \dots = q_0(c_0^*/c_0)$ . Normalizing  $q_0(c_0^*/c_0)$  to unity yields<sup>2</sup>

$$q_t = \left( \frac{c_t}{c_t^*} \right). \quad (3)$$

## 2.2 Firms

The US firm specializes in the production of brand  $z \in [0, 1/2]$  of good  $i$  and employs  $n_{it}(z)$  hours of labor. They will produce output according to the production function  $y_{it}(z) = a_{it}n_{it}(z)$ , where  $a_{it}$  is labor productivity specific to good  $i$  and to all US firms that produce that good. US firms produce brands on the first-half of the unit continuum,  $z \in [0, 1/2]$ . Thus, for example, all firms that produce beer in the US share the same labor productivity. The production function of Canadian firms is  $y_{it}^*(z) = a_{it}^*n_{it}^*(z)$ , for  $z \in (1/2, 1]$ .

Goods that are shipped between the US and Canada are subject to iceberg trade costs,  $\tau$ . In addition, all goods are perishable. Thus, production of good  $i$  undertaken in the US is exhausted between US and Canadian consumption, with Canadian consumption bearing the iceberg trade cost:

$$c_{it}(z) + (1 + \tau)c_{it}^*(z) = y_{it}(z), \text{ for } z \in [0, 1/2]. \quad (4)$$

Similarly, production of good  $i$  undertaken in Canada is exhausted between Canadian and

---

<sup>2</sup>The condition relies on our preference assumptions. Later, we relax these assumptions in Section 4.

US consumption, with US consumption bearing the iceberg trade cost:

$$(1 + \tau)c_{it}(z) + c_{it}^*(z) = y_{it}^*(z), \quad \text{for } z \in (1/2, 1]. \quad (5)$$

## 2.3 Price setting

We introduce the behavioral inattention of firms into an otherwise standard two-country model with Calvo pricing. Firms have the opportunity to change their prices with a constant probability, as in Calvo (1983) and Yun (1996). Firms set prices in the buyers' currency, referred to in the literature as local currency pricing. We first present the pricing decision of fully attentive firms and then relax this assumption following the approach in Gabaix (2014, 2020). Because the pricing problem of Canadian firms is analogous, we limit our exposition to pricing decisions made by US firms.

### 2.3.1 Fully attentive firms

We first specify the fully attentive firm's pricing decision. Let  $x_t$  be a generic variable. We define the log deviation of  $x_t$  from the steady-state level as  $\hat{x}_t = \ln x_t - \ln \bar{x}$ , where  $\bar{x}$  is the steady-state level of  $x_t$ , so that we express  $x_t = \bar{x} \exp(\hat{x}_t)$ . Using this expression, we write the US firm's real profits of selling goods in the US market as  $[p_{it}(z) - w_t/a_{it}]c_{it}(z) = \{\bar{p}_i(z) \exp[\hat{p}_{it}(z)] - \bar{w} \exp(\hat{w}_t - \hat{a}_{it})\}c_{it}(z)$ , where  $p_{it}(z) = P_{it}(z)/P_t$  is the relative price of brand  $z$  of good  $i$  and  $w_t = W_t/P_t$  is the real wage. The demand by US consumers for a particular brand of good  $i$ ,  $c_{it}(z) = (P_{it}(z)/P_{it})^{-\varepsilon}c_{it}$ . In terms of the log deviation, this equation is written as  $c_{it}(z) = (\bar{p}_i(z)/\bar{p}_i)^{-\varepsilon} \{-\varepsilon[\exp(\hat{p}_{it}(z)) - \exp(\hat{p}_{it})]\}c_{it}$ , where  $p_{it} = P_{it}/P_t$ .

We assume that the firms cannot change their price with a probability  $\lambda$ . This parameter captures the degree of price stickiness. Along with the assumption that steady-state inflation is zero, a fully attentive US firm chooses  $\hat{p}_{it}(z)$  to maximize the objective function:

$$v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} \times \frac{P_t}{P_{t+k}} \left\{ \bar{p}_i(z) \exp[\hat{p}_{it}(z)] - \bar{w} \exp \left( \hat{w}_{t+k} + \sum_{l=1}^k \pi_{t+l} - \hat{a}_{it+k} \right) \right\} c_{it,t+k}(z), \quad (6)$$

where

$$c_{it,t+k}(z) = \left( \frac{\bar{p}_i(z)}{\bar{p}_i} \right)^{-\varepsilon} \exp \left\{ -\varepsilon \left[ \hat{p}_{it}(z) - \sum_{l=1}^k \pi_{t+l} - \hat{p}_{it+k} \right] \right\} c_{it+k} \quad (7)$$

is the demand for brand  $z$  of good  $i$  in period  $t + k$ , conditional on the firm having last reset the price in period  $t$ .<sup>3</sup> Here,  $v_{it}(z)$  is the present discount 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 (6), the second line represents the real profits in each period. The marginal cost is the real wage divided by the labor productivity specific to that good. However, because of sticky prices, real wages are adjusted with  $\sum_{l=1}^k \pi_{t+l}$  accumulated from periods  $t$  to  $t + k$ , where  $\pi_t = \ln(P_t/P_{t-1})$  denotes inflation. Real profits in each period are discounted by the stochastic discount factor  $\delta_{t,t+k} = \delta^k(c_{t+k}/c_t)^{-1}$  satisfying  $\delta_{t,t+k}P_t/P_{t+k} = \Delta_{t,t+k}$ . In (7), relative prices are also adjusted by inflation accumulated from period  $t$  to  $t + k$ . Note that this objective function is for the US firms indexed by  $z \in [0, 1/2]$ .

The US firm's real profits of selling goods in the Canadian market are analogously defined. Let  $p_{it}^*(z)$  be the relative price in Canadian markets given by  $p_{it}^*(z) = P_{it}^*(z)/P_t^*$ . A fully attentive US firm chooses  $\hat{p}_{it}^*(z)$  to maximize<sup>4</sup>

$$v_{it}^*(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} q_{t+k} \quad (8)$$

$$\times \frac{P_t^*}{P_{t+k}^*} \left\{ \bar{p}_i^*(z) \exp[\hat{p}_{it}^*(z)] - (1 + \tau) \frac{\bar{w}}{\bar{q}} \exp\left(\hat{w}_{t+k} - \hat{q}_{t+k} + \sum_{l=1}^k \pi_{t+l}^* - \hat{a}_{it+k}\right) \right\} c_{it,t+k}^*(z),$$

where

$$c_{it,t+k}^*(z) = \left( \frac{\bar{p}_i^*(z)}{\bar{p}_i^*} \right)^{-\varepsilon} \exp\left\{ -\varepsilon \left[ \hat{p}_{it}^*(z) - \sum_{l=1}^k \pi_{t+l}^* - \hat{p}_{it+k}^* \right] \right\} c_{it+k}^*. \quad (9)$$

Here,  $\pi_t^* = \ln(P_t^*/P_{t-1}^*)$  and  $p_{it}^* = P_{it}^*/P_t^*$ . In (8), the second line represents real profits in each period. The cost of providing a unit of the good to a Canadian consumer is higher by the amount of the iceberg trade cost  $\tau$ . The real exchange rate in the second line of the equation converts the cost in terms of Canadian goods to compare it to the relative price  $p_{it}^*(z)$ . When discounting the US firm's real profits in each period,  $q_{t+k}$  in the first line of (8) converts these profits in terms of the US goods.

### 2.3.2 Inattentive firms

We now consider the firm's maximization problem when a firm is less than fully attentive to the state variables that enter into its objective function. This problem is called the "sparse max" because Gabaix (2014) originally developed the model in which the economic agents

<sup>3</sup>The derivation is provided in Appendix A.1.

<sup>4</sup>The derivation is again in Appendix A.1.



respond to only a limited number of variables out of numerous variables.

In our model, a firm’s marginal cost is a function of aggregate variables including the real wage and the real exchange rate, as well as microeconomic variables, such as good-specific productivity shocks. For simplicity, we assume that the firm is fully attentive to its own productivity but possibly less attentive to the aggregate variables.

Let us introduce the “attention-augmented” objective function. Define  $m_H \in [0, 1]$  as the degree of attention to set prices of home-produced goods in the US market, where the subscript  $H$  represents the place of production.<sup>5</sup> The attention-augmented objective function is given by

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

where  $\hat{\boldsymbol{\mu}}_{Ht} = (\hat{\mu}_{Ht}, \hat{\mu}_{Ht+1}, \dots)'$  and  $\hat{\mu}_{Ht+k} = \hat{w}_{t+k} + \sum_{l=1}^k \pi_{t+l}$ .<sup>6</sup> In the limit case of  $m_H = 0$ , firms fully ignore changes in the aggregate components of the firm’s cost function,  $\hat{\mu}_{Ht+k}$ . In the opposite limit case of  $m_H = 1$ , the attention-augmented objective function reduces to (6), namely, the full attention case. Because the firm is fully attentive to its own productivity, there is a unit coefficient on  $\hat{a}_{it+k}$ .<sup>7</sup>

In the sparse max, the inattentive firm sets its optimal price to maximize (10):

$$\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) = \arg \max_{\hat{p}_{it}(z)} v_{Hi}(\hat{p}_{it}(z), \hat{\boldsymbol{\mu}}_{Ht}, m_H), \quad (11)$$

given  $m_H$ .

In Gabaix (2014), agents choose the degree of attention endogenously. More attentiveness increases expected profits, a benefit, but being more attentive is costly. We employ the

---

<sup>5</sup>Likewise, we define  $m_H^*$  as the degree of attention to set prices of home-produced goods in the Canadian market. We represent the degree of attention to set prices of foreign-produced goods by  $m_F^*$  when selling the goods in the Canadian market and by  $m_F$  when selling the goods in the US market.

<sup>6</sup>For the aggregate component of the marginal costs of selling goods in the other markets, the definitions are  $\hat{\mu}_{Ht+k}^* = \hat{w}_{t+k} - \hat{q}_{t+k} + \sum_{l=1}^k \pi_{t+l}^*$ ,  $\hat{\mu}_{Ft+k}^* = \hat{w}_{t+k}^* + \sum_{l=1}^k \pi_{t+l}^*$ , and  $\hat{\mu}_{Ft+k} = \hat{w}_{t+k} + \hat{q}_{t+k} + \sum_{l=1}^k \pi_{t+l}$ , respectively.

<sup>7</sup>In the attention-augmented objective function, we do not explicitly introduce  $m_H$  as a coefficient on  $\sum_{l=1}^k \pi_{t+l}$  in (7). This is because we examine the log-linearized first-order condition for the optimal prices. When we take the log-linearization, the presence of  $m_H$  in (7) does not matter for the first-order terms. Further, nor do we explicitly introduce “cognitive discounting” as in Gabaix (2020). Gabaix (2020) assumes that the effect of  $k$  period-ahead economic variables on the agent’s expectations is weakened relative to the rational agent’s expectations, in addition to the degree of attention. In the present model setup, however, we can show that the presence of cognitive discounting does not matter for our results.

quadratic cost function,

$$\mathcal{C}(m_H) = \frac{\kappa}{2} m_H^2,$$

where  $\kappa \geq 0$ . Given the cost function, the firm chooses the optimal allocation of attention by solving

$$\max_{m_H \in [0,1]} \mathbb{E} \{ v_{Hi} [\hat{p}_{Hi}(\hat{\mu}_{Ht}, m_H), \hat{\mu}_{Ht}, 1] - \mathcal{C}(m_H) \}, \quad (12)$$

where  $\mathbb{E}(\cdot)$  represents the unconditional expectations. In (12), we evaluate  $v_{Hi}(\cdot)$  at  $\hat{p}_{it}(z) = \hat{p}_{Hi}(\hat{\mu}_{Ht}, m_H)$  in the first argument and at  $m_H = 1$  in the third argument. That is, the profit function is the true function since  $m_H = 1$  in the third argument but it is evaluated at the inattentive firm's action because  $m_H$  in  $\hat{p}_{Hi}(\hat{\mu}_{Ht}, m_H)$  is not equal to one in general.

Following Gabaix (2014), we define the sparse max for  $v_{it}(z)$  as follows. The firm's choices divide into two steps. In the first step, the firm chooses the degree of attention  $m_H$  based on the linear-quadratic approximation of (12):

$$m_H = \arg \min_{m_H \in [0,1]} \frac{1}{2} (1 - m_H)^2 \Lambda_H + \frac{\kappa}{2} m_H^2, \quad (13)$$

where

$$\Lambda_H = - \left\{ \frac{\partial^2 v_{Hi} [\hat{p}_{Hi}(\mathbf{0}, 1), \mathbf{0}, 1]}{\partial \hat{p}_{Hit}^2} \right\} Var(\hat{\mu}_{Ht}). \quad (14)$$

The solution of the first step is given by  $m_H = \Lambda_H / (\Lambda_H + \kappa)$ . In the second step, the firm chooses the optimal price (11), given the solution of the first step.<sup>8</sup>

In this sparse max, the choice of  $m_H = \Lambda_H / (\Lambda_H + \kappa) = 0$  is excluded as long as  $Var(\hat{\mu}_{Ht}) > 0$ , which implies  $\Lambda_H > 0$ . Gabaix (2014) showed that in the case of a quadratic cost function, the selected degree of attention is zero if and only if there is no uncertainty in the variables to which the economic agents pay only partial attention.<sup>9</sup> In addition, in the special case of  $\kappa = 0$ ,  $m_H = \Lambda_H / (\Lambda_H + \kappa) = 1$  is selected because  $\kappa = 0$  means that there is no cost of paying attention. For these reasons, in the following analysis, we focus on the case of  $m_H \in (0, 1]$ . As we discuss later, these assumptions are convenient for our objective of accounting for the PPP puzzle because they ensure the stationarity of the PPP and LOP deviations.

Note that we can also introduce inattention to the idiosyncratic productivity and derive the expressions similar to (13) and (14). Nevertheless, we focus on the case that firms are fully

---

<sup>8</sup>In Appendix A.2, we derive (13) and (14) that are relevant to US firms selling in US markets. The appendix also describes the remaining sparse max for US firms selling abroad and Canadian firms selling in the Canadian and US markets.

<sup>9</sup>Gabaix (2014) discusses the properties of the selected degree of attention using not only the quadratic cost function but also the other functional forms of the cost function.

attentive to their productivity but are inattentive to the aggregate component for three reasons. First, in general, the level of uncertainty matters for the size of the degree of attention. Naturally, the volatility of the idiosyncratic shock can be much higher than the aggregate shocks. In this case,  $\Lambda_H$  for the idiosyncratic productivity would be much larger than  $\Lambda_H$  for the aggregate component of the marginal cost, meaning that the degree of attention to the idiosyncratic variable is closer to unity than that to the aggregate variable. Second, firms may have easier access to information on their variables rather than the macroeconomic variables. In this case,  $\kappa$  for their productivity may be much lower than  $\kappa$  for the aggregate shock, such as monetary shocks. Thus, the degree of attention to the idiosyncratic variable is again closer to unity. Finally, our test of behavioral inattention in Section 4 can still detect the inattention to the aggregate variable even if we allow inattention to their idiosyncratic productivity. In other words, our test is fully robust to the presence of inattention to their idiosyncratic productivity. In this sense, the full attention to idiosyncratic productivity is a convenient assumption for focusing on the presence of inattention to the aggregate variable.

## 2.4 Equilibrium

The monetary authority in each country determines the national stock of money. Following Kehoe and Midrigan (2007), we assume that the log of the money supply follows a random walk:

$$\ln M_t = \ln M_{t-1} + \varepsilon_t^M, \quad (15)$$

$$\ln M_t^* = \ln M_{t-1}^* + \varepsilon_t^{M^*}, \quad (16)$$

where  $\varepsilon_t^M$  and  $\varepsilon_t^{M^*}$  are zero-mean i.i.d. shocks. Importantly, the stochastic processes, combined with (3) and the CIA constraints, imply a nominal exchange rate that follows a random walk, which is empirically plausible. In particular, we have  $S_t = M_t/M_t^*$  from (3) and the CIA constraints. This equation leads to  $\ln S_t = \ln S_{t-1} + \varepsilon_t^n$ , where  $\varepsilon_t^n = \varepsilon_t^M - \varepsilon_t^{M^*}$  is the shock to the nominal exchange rate. We simply call  $\varepsilon_t^n$  the nominal shock.

For simplicity, we assume that the log labor productivity also follows a zero-mean i.i.d.

process:<sup>10</sup>

$$\ln a_{it} = \varepsilon_{it}, \quad (17)$$

$$\ln a_{it}^* = \varepsilon_{it}^*. \quad (18)$$

The difference in labor productivity is  $\ln(a_{it}/a_{it}^*) = \varepsilon_{it}^r$ , where  $\varepsilon_{it}^r = \varepsilon_{it} - \varepsilon_{it}^*$ . We refer to the shock to the difference in productivity as the real shock.

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  $n_t = \int_i \int_{z=0}^{1/2} n_{it}(z) dz di$  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 Theoretical implications for LOP deviations

In this section, we derive the reduced-form solution to the good-level real exchange rate. Taking the first-order condition with respect to  $\hat{p}_{it}(z)$  from (10) and log-linearizing the condition around the steady state yield the optimal price:

$$\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) = (1 - \lambda\delta) \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k (m_H \hat{\mu}_{Ht+k} - \hat{a}_{it+k}) \quad (19)$$

$$= m_H \hat{w}_t - (1 - \lambda\delta) \hat{a}_{it}. \quad (20)$$

Here, the expression reflects the forward-looking properties in the Calvo pricing and  $\lambda$  affects the extent to which firms place weights on the expected marginal cost. Derivation of the

---

<sup>10</sup>Later we consider an alternative stochastic process for productivity, but the empirical results from the test of behavioral inattention are unaffected.

second equality is provided in Appendix A.3. The optimal price set by US firms for the Canadian market is given by:

$$\hat{p}_{Hi}^*(\hat{\boldsymbol{\mu}}_{Ht}^*, m_H^*) = m_H^*(\hat{w}_t - \hat{q}_t) - (1 - \lambda\delta)\hat{a}_{it}. \quad (21)$$

Similarly, the prices set by Canadian firms for the Canadian market and the US market are respectively given by:

$$\hat{p}_{Fi}^*(\hat{\boldsymbol{\mu}}_{Ft}^*, m_F^*) = m_F^*\hat{w}_t^* - (1 - \lambda\delta)\hat{a}_{it}^*, \quad (22)$$

and

$$\hat{p}_{Fi}(\hat{\boldsymbol{\mu}}_{Ft}, m_F) = m_F(\hat{w}_t^* + \hat{q}_t) - (1 - \lambda\delta)\hat{a}_{it}^*. \quad (23)$$

Turning to the price index for good  $i$ , we log-linearize the CES index for good  $i$  sold in the US market:

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

Here,  $\hat{p}_{it}^{opt}$  denotes the weighted average of the optimal reset prices:

$$\hat{p}_{it}^{opt} = \omega\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) + (1 - \omega)\hat{p}_{Fi}(\hat{\boldsymbol{\mu}}_{Ft}, m_F), \quad (25)$$

where  $\omega = (1 + (1 + \tau)^{1-\varepsilon})^{-1} \in [1/2, 1]$  is the degree of home bias. The home bias is strictly larger than 1/2 in the presence of the iceberg trade costs ( $\tau > 0$ ). The log-linearized price index for good  $i$  sold in the Canadian markets is

$$\hat{p}_{it}^* = \lambda(\hat{p}_{it-1}^* - \pi_t^*) + (1 - \lambda)\hat{p}_{it}^{opt*}, \quad (26)$$

where

$$\hat{p}_{it}^{opt*} = \omega\hat{p}_{Fi}^*(\hat{\boldsymbol{\mu}}_{Ft}^*, m_H) + (1 - \omega)\hat{p}_{Hi}^*(\hat{\boldsymbol{\mu}}_{Ht}^*, m_F). \quad (27)$$

In (27), we employ the assumption of a symmetry between the US and Canada. That is, the degrees of attention in the production of domestically consumed goods are identical in the US and Canada, such that  $m_F^* = m_H$ . Likewise, the degrees of attention in the production of exported goods are also identical, such that  $m_H^* = m_F$ .

Recall that the PPP deviation, or the aggregate real exchange rate, is defined by  $q_t = S_t P_t^*/P_t$ . Similarly, the LOP deviation, or the good-level real exchange rate, is defined by  $\hat{q}_{it} = S_t P_{it}^*/P_{it}$ . Using  $p_{it}$  and  $p_{it}^*$ ,  $\hat{q}_{it}$  is expressed as

$$\hat{q}_{it} = \hat{q}_t + \hat{p}_{it}^* - \hat{p}_{it}. \quad (28)$$

We combine (20) - (28), and the CIA constraints to obtain the expression for the good-level real exchange rate. The following proposition summarizes the dynamics of the good-level real exchange rate.

**Proposition 1** *Under the preferences given by  $U(c, n) = \ln c - \chi n$ , the CIA constraints, the stochastic processes of money supply (15) and (16), the stochastic processes of the labor productivity (17) and (18), and the Calvo pricing with the degree of price stickiness  $\lambda \in (0, 1)$ , the stochastic process of the good-level real exchange rate is given by:*

$$\ln q_{it} = \lambda \ln q_{it-1} + (1 - m)(1 - \lambda) \ln q_t + \lambda \varepsilon_t^n + (1 - \lambda)(1 - \lambda\delta)\psi \varepsilon_{it}^r, \quad (29)$$

where  $m \in (0, 1]$  represents the degree of attention:

$$m = \omega m_H + (1 - \omega) m_F. \quad (30)$$

and  $\psi = 2\omega - 1$ . The two random shocks  $\varepsilon_{it}^r$  and  $\varepsilon_t^n$  are given by  $\varepsilon_{it}^r = \varepsilon_{it} - \varepsilon_{it}^* \sim i.i.d.(0, \sigma_r^2)$  and  $\varepsilon_t^n = \varepsilon_t^M - \varepsilon_t^{M*} \sim i.i.d.(0, \sigma_n^2)$ , respectively.

**Proof.** See Appendix A.4. ■

This stochastic process for the good-level real exchange rate generalizes the simple stochastic process considered by Kehoe and Midrigan (2007) who emphasized the importance of nominal shocks. They showed that under the fully attentive rational expectations model, the good-level real exchange rate follows an autoregressive process of order one (AR(1)) driven by the nominal shock  $\varepsilon_t^n$ :

$$\ln q_{it} = \lambda \ln q_{it-1} + \lambda \varepsilon_t^n. \quad (31)$$

This equation is a special case of (29) with  $m = 1$  and  $\psi = 0$ .<sup>11</sup> To gain some intuition behind (31), recall that  $\ln q_{it} = \ln S_t + \ln P_{it}^* - \ln P_{it}$ . Suppose that money supply increases unexpectedly in the US. While the unexpected increase in the domestic money supply keeps  $P_{it}^*$  constant, it increases  $S_t$  and  $P_{it}$ . Notice that the nominal exchange rate is free to adjust, while the adjustment of  $P_{it}$  is slow because of sticky prices. As a result, the increase in  $P_{it}$  only partially offsets the increase in  $S_t$ . The extent of the offsetting effect depends on  $\lambda$ . If  $\lambda \rightarrow 0$ , a change in  $P_{it}$  perfectly offsets the increase in  $S_t$ , meaning that the nominal shock is irrelevant for the real exchange rate. If  $\lambda \rightarrow 1$ ,  $P_{it}$  never moves, meaning that the good-level real exchange rate keeps track of the nominal exchange rate, which follows a random walk.

<sup>11</sup>Note that  $\psi = 0$  if and only if  $\tau = 0$  (i.e., no trade cost) from the definition of  $\omega = 1/(1 + (1 + \tau)^{1-\varepsilon})$ .

Let us compare the stochastic processes for the good-level real exchange rates between the cases  $m = 1$  and  $0 < m < 1$ .<sup>12</sup> For comparison purposes, we maintain the assumption of  $\psi = 0$ . If firms are only partially attentive to the aggregate component of the marginal cost (i.e.,  $0 < m < 1$ ), the good-level real exchange rate has the aggregate real exchange rate on the right-hand side:

$$\ln q_{it} = \lambda \ln q_{it-1} + (1 - m)(1 - \lambda) \ln q_t + \lambda \varepsilon_t^n. \quad (32)$$

The intuition behind the appearance of the aggregate real exchange rate in (32) lies in the responses of relative prices  $\hat{p}_{it}$  and  $\hat{p}_{it}^*$  to aggregate shocks. If firms become less attentive to the aggregate components of the marginal cost, relative prices are more invariant to aggregate shocks. The more invariant a relative price, the more the firm anchors its nominal prices to the aggregate price level. The link between the good-level and the aggregate price indices leads to a link between the good-level and the aggregate real exchange rates.

It should be noted that there is a single common driving force in both (31) and (32) because the aggregate real exchange rate that additionally enters in (32) is also driven by the nominal shock. Indeed, aggregating  $\ln q_{it}$  over  $i$  yields<sup>13</sup>

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

Using (33), we can see that the impact multiplier of nominal shocks on the good-level real exchange rate increases from  $\lambda$  in (31) to  $\lambda \times (1 + \frac{(1-m)(1-\lambda)}{1-(1-m)(1-\lambda)})$  in (32). In other words, the behavioral inattention changes the stochastic process of the good-level real exchange rate but not the source of its variations.

When  $\psi > 0$ , a real shock represented by  $\varepsilon_{it}^r$  appears in the stochastic process as an additional driving force. Stressing the importance of real shocks, Crucini et al. (2010b, 2013) extended the Kehoe and Midrigan (2007) model to incorporate idiosyncratic productivity

---

<sup>12</sup>Note that  $m$  is the mean of the degrees of attention  $m_H$  and  $m_F$ . Because  $0 < \omega < 1$  holds for  $\tau \in [0, \infty)$ ,  $m = 1$  holds only if all US and Canadian firms are completely attentive to the aggregate component of their marginal costs.

<sup>13</sup>To derive the stochastic process, we integrate (29) across good  $i$ . In aggregation,  $\int_{i=0}^1 \ln q_{it} di = \ln q_t$  holds from the definition of the good-level real exchange rate. From the definition of  $q_{it}$ ,  $\ln q_{it} = \ln q_t + \ln p_{it}^* - \ln p_{it}$ . For the US relative prices, the integral of the relative price 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 relative 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).

shocks. Under the fully attentive rational expectations, their model implies:

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

This equation is a special case of (29) under  $m = 1$ . To understand the intuition behind the role of real shocks, again recall that  $\ln q_{it} = \ln S_t + \ln P_{it}^* - \ln P_{it}$ . Positive productivity shocks in the US firms producing good  $i$  reduce both  $P_{it}^*$  and  $P_{it}$  because the US firms sell their goods in both countries. However, the home bias generated by trade cost will decrease  $P_{it}$  more than  $P_{it}^*$ . This effect results in an appreciation of  $q_{it}$ . For the case of  $0 < m < 1$ , (33) continues to hold unless aggregate real shocks are introduced. In the process of aggregating the good-level real exchange rates, all idiosyncratic real shocks are washed out in the integral over  $i$  because  $\int_{i=0}^1 \varepsilon_{it}^r di = 0$ .

To summarize, behavioral inattention generates a new term that affects the good-level real exchange rate, namely, the aggregate real exchange rate. Time-dependent pricing models of the good-level and aggregate real exchange rates without the behavioral inattention have been theoretically developed and empirically assessed by Kehoe and Midrigan (2007) and Crucini et al. (2010b, 2013), among many others. However, the importance of the behavioral inattention has not been tested in the context of LOP deviations. In the next section, we use a rich international dataset on good-level real exchange rates to test the model of behavioral inattention.

## 4 A test of behavioral inattention

### 4.1 Methodology

In this section, we consider testing the null hypothesis of  $m = 1$  (full attention), against the alternative hypothesis of  $m < 1$  (partial attention). Define  $\ln \tilde{q}_{it} = \ln \left[ q_{it} / (q_{it-1} S_t / S_{t-1})^\lambda \right] = \ln q_{it} - \lambda \ln q_{it-1} - \lambda \Delta \ln S_t$  and  $\ln \tilde{q}_t = \ln(q_t / q_t^\lambda) = (1 - \lambda) \ln q_t$ . We estimate the following panel regression model:

$$\ln \tilde{q}_{it} = \alpha + \beta \ln \tilde{q}_t + \gamma' X_{it} + u_{it}, \quad (35)$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are regression coefficients,  $X_{it}$  is a vector of the control variables, and  $u_{it}$  is the error term. This regression is motivated by (29) because it can be written as

$$\ln \tilde{q}_{it} = (1 - m) \ln \tilde{q}_t + (1 - \lambda)(1 - \lambda\delta)\psi \varepsilon_{it}^r, \quad (36)$$



where the nominal shock  $\varepsilon_t^n$  in (29) is replaced by  $\Delta \ln S_t$  because the (log) nominal exchange rate follows a random walk with an increment  $\varepsilon_t^n$ . To implement the regression, we construct  $\ln \tilde{q}_{it}$  and  $\ln \tilde{q}_t$  using the micro evidence of  $\lambda$ . The error term  $u_{it}$  represents an i.i.d. real shock  $\varepsilon_{it}^r$  and is uncorrelated with the regressor  $\ln \tilde{q}_t = (1 - \lambda) \ln q_t$  because  $\varepsilon_{it}^r$  does not show up in (33). Therefore, we estimate (35) using ordinary least squares (OLS). The control variables  $X_{it}$  here may include time-invariant fixed effects or other time-varying components, such as common productivity differentials across countries, which we discuss later.

The key idea is the equivalence of testing the full attention hypothesis and checking the statistical significance of the coefficient on  $\ln \tilde{q}_t$  in (35) because (36) suggests that  $\beta = 1 - m = 0$  if firms are fully attentive. When the null hypothesis of  $\beta = 0$  is rejected in favor of the alternative hypothesis of  $\beta > 0$ , the data are consistent with the presence of inattentive firms. Note that because the nominal exchange rate is the common driving force of the good-level and aggregate real exchange rates ( $\ln q_{it}$  and  $\ln q_t$ ), the two variables are expected to be highly correlated to each other. In our regression, however, both good-level and real exchange rates are modified so that two variables ( $\ln \tilde{q}_{it}$  and  $\ln \tilde{q}_t$ ) are correlated only when the degree of attention is less than unity. As an important by-product of the regression (35), the degree of attention  $m$  can be obtained as  $\hat{m} = 1 - \hat{\beta}$ , where  $\hat{\beta}$  is an OLS estimator of  $\beta$ .

Some remarks are in order. First, we can generalize the stochastic process of labor productivity from a simple i.i.d. process to a more realistic process that allows for a nonstationary stochastic trend and a stationary but serially correlated component. Let us assume that labor productivity is given by:

$$\ln a_{it} = \xi_t + \eta_t + \varepsilon_{it}, \quad (37)$$

$$\ln a_{it}^* = \xi_t + \eta_t^* + \varepsilon_{it}^*. \quad (38)$$

Here, the labor productivity consists of three components: a global component  $\xi_t$ , a country-specific component  $\eta_t$  (or  $\eta_t^*$ ), and a good-specific component  $\varepsilon_{it}$  (or  $\varepsilon_{it}^*$ ). In this generalized setting, global and country-specific components follow  $\xi_t - \xi_{t-1} = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}^{\xi}$ ,  $\eta_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}^{\eta}$ , and  $\eta_t^* = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}^{\eta^*}$ , respectively, where  $\varepsilon_t^{\xi}$ ,  $\varepsilon_t^{\eta}$ , and  $\varepsilon_t^{\eta^*}$  are i.i.d. shocks. This error structure implies that the productivities in both countries are nonstationary, but share a common stochastic trend (or the two variables are cointegrated). Because only the relative labor productivity  $\ln a_{it} - \ln a_{it}^*$  matters in the dynamics of LOP deviations, the global component becomes irrelevant in our analysis. However, this is not the case for the country-specific component, in which case, regression (35) requires modification. For example, if  $\eta_t$

and  $\eta_t^*$  each follow an AR(1) process with AR coefficient  $\rho_\eta$  and firms are fully attentive to  $\eta_t$  and  $\eta_t^*$ , (36) is modified to

$$\ln \tilde{q}_{it} = (1 - m) \ln \tilde{q}_t + \frac{(1 - \lambda)(1 - \lambda\delta)}{1 - \lambda\delta\rho_\eta} \psi \eta_t^r + (1 - \lambda)(1 - \lambda\delta) \psi \varepsilon_{it}^r, \quad (39)$$

where  $\eta_t^r = \eta_t - \eta_t^*$ . The only difference between (36) and (39) is that the latter includes the new control variable  $\eta_t^r$ . We can obtain a similar equation even if we include additional lags in the process of the country-specific component.

Second, even if we additionally introduce the inattention to the idiosyncratic productivity, our regression framework of testing behavioral inattention (to the aggregate variable) remains valid. The stochastic process for the good-level real exchange rate (29) needs to be slightly modified because  $\ln q_{it}$  becomes less sensitive to  $\varepsilon_{it}^r$ . A smaller coefficient on  $\varepsilon_{it}^r$  in (29), however, does not change our regression equation (35).

Third, our model assumes quasi-linear preferences  $U(c, n) = \ln c - \chi n$ . In the case of the more general constant-relative-risk-aversion (CRRA) form, firms expect a dynamic path for the labor supply from the time of price setting to the infinite future. As a result, the good-level real exchange rate does not have a simple reduced-form solution. However, we can still conduct a test of behavioral inattention and obtain the degree of attention by employing the instrumental variables estimator under CRRA preferences. Appendix A.5 derives the estimation equation together with the empirical results.

Finally, an alternative way of testing  $m = 1$  is to regress  $\ln q_{it}$  directly on  $\ln q_t$  with additional regressors  $\ln q_{it-1}$  and  $\Delta \ln S_t$ . The estimation equation is given by:

$$\ln q_{it} = \alpha + \beta \ln q_t + \gamma' X_{it} + u_{it}, \quad (40)$$

where  $\beta$  in (40) corresponds to  $(1 - m)(1 - \lambda)$  in (29). The control variables  $X_{it}$  in (40) now include  $\ln q_{it-1}$  and  $\Delta \ln S_t$ . Note that  $\beta = (1 - m)(1 - \lambda) = 0$  corresponds to  $m = 1$  provided  $\lambda < 1$ . Therefore, a test of  $\beta = 0$  against  $\beta > 0$  in (40) is equivalent to the test of the fully attentive hypothesis. In Appendix A.6, we report the empirical results based on this alternative strategy.

## 4.2 Data

We use the retail price data from the *Worldwide Cost of Living Survey* compiled by the Economist Intelligence Unit (EIU), which entails an extensive annual survey of international

retail prices in a variety of cities. The survey reports all the prices of individual goods in local currency terms, conducted by a single agency in a consistent manner over time. The coverage of goods and services is substantial in breadth and thus overlaps with the typical urban consumption basket tabulated by national statistical agencies.<sup>14</sup> Recent studies using 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).

Our analysis focuses on the US–Canadian city pairs and the UK and Euro area city pairs. For the US–Canadian city pairs, the data contain prices of 274 goods and services in multiple cities from 1990 to 2015. There are 16 US and four Canadian cities.<sup>15</sup> This results in 64 unique cross-country city pairs. However, because some US cities have many missing values in the early 90s, our data are an unbalanced panel.<sup>16</sup> Nevertheless, the total number of observations available for our regressions exceeds 350,000.

For the UK–Euro area city pairs, there are two UK cities and 19 Euro area cities.<sup>17</sup> The data include 301 goods and services from 1990 to 2015. As in the US–Canadian city pairs, the panel is unbalanced. The number of observations exceeds 200,000 from 38 UK–Euro area city pairs.

We compute the log of  $q_{ijt}$  for each year ( $t = 1990, \dots, 2015$ ), each good ( $i = 1, 2, \dots$ ), and each international city pair ( $j = 1, 2, \dots$ ). The prices used to construct the good-level real exchange rates are the prices in a city expressed in the local currency unit. We use the spot nominal exchange rates from the EIU data to convert prices to common currency units. The EIU records the nominal exchange rate vis-à-vis the US dollar at the end of the week of the price survey. Thus, the nominal exchange rate 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 were different from those in the remaining Canadian cities from 2003 to 2014.<sup>18</sup> The nominal exchange rates in cities of other countries are common in the EIU data.

---

<sup>14</sup>See Rogers (2007) for details on the comparison between EIU data and the consumer price index data from national statistical agencies.

<sup>15</sup>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.

<sup>16</sup>In particular, the survey data in 1990 and 1991 do not include the price data of goods and services in Honolulu, and Lexington and Minneapolis have only been included in the list of cities since 1998.

<sup>17</sup>The UK cities are London and Manchester. The Euro area cities are Amsterdam, Athens, Barcelona, Berlin, Brussels, Dublin, Dusseldorf, Frankfurt, Hamburg, Helsinki, Lisbon, Luxembourg, Lyon, Madrid, Milan, Munich, Paris, Rome, and Vienna.

<sup>18</sup>As we discuss later, we adjust our regressions to account for this difference in timing.

Figure 1 plots two kernel density estimates of the bilateral good-level real exchange rates pooling all goods and services, one for the first year of the sample (1990) and the other for the last year of the sample (2015). The upper panel of the figure shows the distribution of the good-level real exchange rates for the US–Canadian city pairs, and the lower panel shows those for the UK–Euro area city pairs. For our regressions and empirical tests that follow, we augment the micro price data with the aggregate bilateral real exchange rate computed from the official consumer price indices, which the EIU also reports.

When we allow for the general stochastic process of labor productivity (37) and (38), we need to control for the difference in the country-specific components in the labor productivity  $\eta_t^r (= \eta_t - \eta_t^*)$  in (39). As a proxy for  $\eta_t^r$ , we utilize the difference in real GDP per hour worked between two countries taken from *OECD.Stat*.

### 4.3 Estimation results

Table 1 provides the estimation results of (35) for the test of behavioral inattention. The left panel shows the results for the US–Canadian city pairs, whereas the right panel presents those for the UK–Euro area city pairs. To construct  $\ln \tilde{q}_{ijt}$  and  $\ln \tilde{q}_t$ , we calibrate  $\lambda$  using a value suggested by previous studies. Nakamura and Steinsson (2008) report that the median frequency of price changes in the US is 8.7 percent. Gautier et al. (2022) find that the average frequency of price changes is 8.5 percent for consumer prices in the 11 Euro area countries.<sup>19</sup> We transform monthly frequencies of price changes into the *inf*frequencies of price changes at an annual rate to compute the value of  $\lambda$ . The transformation leads to  $\lambda \simeq 0.34$  in both cases.<sup>20,21</sup> Because the degree of price stickiness at the macroeconomic level is similar between the US and the Euro area countries, our regression (35) assumes that  $\lambda = 0.34$ . The table reports the estimated coefficient on  $\ln \tilde{q}_t$  with the standard errors. We include the good-specific fixed effects in the regressions by default. This is because variations in the good-specific fixed effect are substantially larger in the LOP deviations than in the city-pair-

---

<sup>19</sup>Both Nakamura and Steinsson (2008) and Gautier et al. (2022) remove the impact of sales on the frequencies of price changes. In addition, Nakamura and Steinsson (2008) remove the impact of product substitutions on the frequency of price changes.

<sup>20</sup>We transform the monthly frequency of price changes into the annual infrequency of price changes as follows. Let  $f$  be the monthly frequency of price changes. If the price of a good is kept unchanged for 12 months under our assumption of sticky prices, the probability of not being able to change prices within a year is  $(1 - f)^{12}$ . We substitute  $f = 0.087$  in Nakamura and Steinsson (2008) and  $f = 0.085$  in Gautier et al. (2022) into the above formula.

<sup>21</sup>Alternatively, we can use  $\lambda$  implied from Bils and Klenow (2004) and Klenow and Kryvtsov (2008) for the US–Canadian city pairs and Alvarez et al. (2006) for the UK–Euro area city pairs. Our main findings are robust, even if we replace  $\lambda$  with the values implied from their estimates.

specific fixed effect.<sup>22</sup> For robustness, we allow for adding the city-pair-specific fixed effects to the regression and/or controlling for the country-specific component of labor productivity  $\eta_t^r$ , motivated by (39). In regressions for the US–Canadian city pairs, we also control for the difference in timing of the price survey in Calgary by adding dummy variables that take a value of one if a city pair includes Calgary in 2003, 2004, ..., or 2014.<sup>23</sup>

Overall,  $\hat{\beta}$  is around 0.8, or equivalently, the estimated degree of attention  $\hat{m}(= 1 - \hat{\beta})$  is around 0.2. The standard error of the coefficient indicates that the test strongly rejects the null hypothesis of full attention ( $\beta = 0$ ) against the alternative hypothesis of partial attention ( $\beta > 0$ ).<sup>24</sup> Comparisons between the left and right panels reveal that the estimated coefficients on  $\ln \tilde{q}_t$  for the US–Canadian city pairs are larger than those for the UK–Euro area city pairs. Taking specification (1) as an example, the first row of Table 1 indicates that the estimated  $\beta$  is 0.84 for the US–Canadian city pairs while the estimated  $\beta$  is 0.79 for the UK–Euro area city pairs. Our results for the test of behavioral inattention are robust to the presence of city-pair-specific fixed effects (see specifications (2) and (4)) and to the inclusion of the log-difference in labor productivity as a control variable (see specifications (3) and (4)).<sup>25</sup> Interpreted through the lens of our theoretical model, these results suggest that firms are not fully attentive to the aggregate components of marginal costs in making their pricing decisions. The bottom of the table also reports that the estimated degrees of attention,  $\hat{m}_t$ , are lower in the US–Canadian city pairs than in the UK–Euro area city pairs. For example, in specification (1),  $\hat{m} = 0.16$  in the US–Canadian city pairs and  $\hat{m} = 0.21$  in the UK–Euro area city pairs.

Table 2 points to the estimation results when the calibrated  $\lambda$  differs across goods. While our model assumes the degree of price stickiness common across goods, the previous empirical studies on sticky prices find heterogeneity in the frequency of price changes. Thus, we transform each good-specific monthly frequency of price changes reported by Nakamura and Steinsson (2008) into the good-specific annual infrequency of price changes and construct

---

<sup>22</sup>Crucini and Telmer (2020) emphasize the importance of good-specific fixed effects using the analysis of variance of the EIU data.

<sup>23</sup>The difference in timing of the price survey causes the aggregate real exchange rate to be city-pair- and year-specific. More specifically, let  $q_t^k$  and  $S_t^k$  be the aggregate real exchange rate and the nominal exchange rate for a city pair  $k$  that involves Calgary in a year from 2003 to 2014. Here,  $\ln q_t^k$  is given by  $\ln q_t^k = \ln S_t^k + \ln P_t^* - \ln P_t$ . We can express  $\ln q_t^k$  as  $\ln q_t^k = (\ln S_t^k - \ln S_t) + \ln q_t$  and  $\ln q_{ijt}^k$  as  $\ln q_{ijt}^k = (\ln S_t^k - \ln S_t) + \ln q_{ijt}$  where the variables without the superscript  $k$  are variables in the other city pairs. Therefore, this dummy variable can control for the presence of  $\ln S_t^k - \ln S_t$  arising from the difference in timing of the price survey in Calgary.

<sup>24</sup>We report the standard errors clustered by goods, but the null hypothesis is also rejected even if the standard errors are clustered by city pairs or years.

<sup>25</sup>While we do not report the result, we also allow for a fixed effect specific to both good  $i$  and city pair  $j$ . We find that the estimated  $\beta$  is not substantially different.

$\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda_i \ln q_{ijt-1} - \lambda_i \Delta \ln S_t$  and  $\ln \tilde{q}_t^i = (1 - \lambda_i) \ln q_t$  for good  $i$ .<sup>26</sup> Unfortunately, good-specific frequencies of price changes are not available in Gautier et al. (2022). Thus, we apply the good-specific monthly frequencies of price changes provided by Nakamura and Steinsson (2008) to the UK–Euro city pairs.

Even when we allow for the good-specific degree of price stickiness, the null hypothesis of full attention is again significantly rejected. Regarding the estimated degrees of attention,  $\hat{m}$  tends to be lower when we allow for the good-specific  $\lambda$ . In particular, if we take specification (1) for comparison,  $\hat{m}$  reduces from 0.16 to 0.11 for the US–Canadian city pairs and from 0.21 to 0.16 for the UK–Euro area city pairs. We observe similar reductions in  $\hat{m}$  under the remaining specifications.

## 5 Explaining the PPP puzzle

In the previous section, we provided strong evidence for behavioral inattention using micro price data. We now turn to the implications of this finding for the PPP puzzle.

### 5.1 Persistence of the aggregate real exchange rate

Let  $\rho_q$  be the first-order autocorrelation of aggregate real exchange rates. Because the AR coefficient in (33) corresponds to the first-order autocorrelation, let us rewrite (33) as:

$$\ln q_t = \rho_q \ln q_{t-1} + \rho_q \varepsilon_t^n, \quad (41)$$

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 in Proposition 1,*

$$\rho_q \geq \lambda, \quad (42)$$

*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 (42) holds with the equality if and only if  $m = 1$ . ■

---

<sup>26</sup>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.

Proposition 2 implies that the presence of behavioral inattention helps with the resolution of Rogoff’s (1996) PPP puzzle. That is, the aggregate real exchange rate is more persistent than the degree of price stickiness implies. Without behavioral inattention (i.e.,  $m = 1$ ),  $\rho_q$  is equal to  $\lambda$ . However, if firms are inattentive (i.e.,  $m < 1$ ),  $\rho_q$  becomes strictly greater than  $\lambda$ . In the extreme case of  $m \rightarrow 0$ , the aggregate real exchange rate can even follow a random walk, as  $\rho_q \rightarrow 1$ . Therefore, even when the nominal frictions are small, the model with a small  $m$  can explain a highly persistent aggregate real exchange rate.

We rule out the case of flexible price ( $\lambda = 0$ ) in Propositions 1 and 2 because (41) suggests that  $\lambda = 0$  leads to no PPP deviations, even in the short run ( $\ln q_t = 0$  for all  $t$ ). Our model thus requires nominal rigidities as the external source of the persistence of the aggregate real exchange rate. We can understand this feature of our model in the context of real rigidities in Ball and Romer (1990) or as a form of strategic complementarity as in Woodford (2003). Using a closed-economy model, Ball and Romer (1990) show that real rigidities are insufficient to create real effects of nominal shocks. They argue that a combination of real rigidities and a small friction in the nominal price adjustment matters for the real effect of a nominal shock. In our model, a combination of behavioral inattention and a small friction in the nominal price adjustment could generate a substantially persistent aggregate real exchange rate.

Figure 2 shows how the persistence of aggregate real exchange rate changes as  $m$  changes. The left panel plots  $\rho_q$  against  $m \in (0, 1]$  when  $\lambda$  is calibrated at 0.34. For reference, the figure also plots the line of the lower bound of  $\rho_q$ :  $\lambda = 0.34$ . Starting from  $\rho_q = \lambda$  when  $m = 1$ ,  $\rho_q$  increases monotonically as  $m$  decreases. The persistence becomes closer to unity as  $m$  approaches zero. The right panel illustrates the  $\rho_q$  to  $\lambda$  ratio, which is defined as:

$$\frac{\rho_q}{\lambda} = \frac{1}{1 - (1 - m)(1 - \lambda)}. \quad (43)$$

This ratio measures the extent to which inattention amplifies the persistence of the aggregate real exchange rate explained solely by nominal rigidities under full attention. The figure suggests that the  $\rho_q$  to  $\lambda$  ratio can be quite large depending on  $m$ .

The estimated degrees of attention suggest that behavioral inattention makes the PPP deviations more than twice as persistent as what is predicted only by the degree of price stickiness. In the left panel of Figure 2,  $\rho_q = 0.34$  if  $m = 1$ . However, if we take  $m = 0.16$  as a calibrated value for the US–Canadian city pairs (i.e., specification (1) of Table 1 for the US–Canadian city pairs), the left panel of Figure 2 indicates that  $\rho_q = 0.76$  when  $m = 0.16$ . We also see from the right panel of the same figure that these calibrated values generate the  $\rho_q$  to  $\lambda$  ratio that exceeds two. In particular, the  $\rho_q$  to  $\lambda$  ratio is 2.24 when  $m = 0.16$ .

When we use  $m = 0.21$  (i.e., specification (1) of Table 1 for the UK–Euro area city pairs),  $\rho_q$  becomes 0.71 and the  $\rho_q$  to  $\lambda$  ratio is 2.09.

Let us evaluate how much the estimated degree of inattention can explain the half-life of the aggregate real exchange rate. The upper panel of Table 3 compares the predicted half-lives of the aggregate real exchange rate with the half-lives observed in our data. Taking the aggregate real exchange rate used for the regression equation (35), we estimate half-lives from the AR(1) model for  $\ln q_t$ .<sup>27</sup> As shown in the third column of the upper panel (the column headed “Data”), the observed half-life for the US–Canadian city pairs is 4.92 years, consistent with three to five years in Rogoff (1996). The aggregate real exchange rate for the UK–Euro area city pairs exhibits a slightly low half-life of 2.43 years.<sup>28</sup>

The model with behavioral inattention can explain the observed persistence of the aggregate real exchange rate quite well. When we use  $\hat{m} = 0.16$  for the US–Canadian city pairs and  $\hat{m} = 0.21$  for the UK–Euro area city pairs (i.e., specification (1) in Table 1), the predicted half-lives are 2.62 years and 2.04 years, respectively. While the predicted half-lives are slightly shorter than the half-lives observed in the data, the ranges implied from the 95 percent confidence intervals of  $\hat{m}$  overlap the observed half-life in both the US–Canadian city pairs and the UK–Euro area city pairs. If we take  $\hat{m} = 0.11$  for the US–Canadian city pairs and  $\hat{m} = 0.16$  for the UK–Euro area city pairs (i.e., specification (1) in Table 2), the predictions of the model with behavioral inattention further improves. In particular, the predicted half-lives are 3.70 years for the US–Canadian city pairs and 2.54 years for the UK–Euro area city pairs. Again, the ranges implied from the 95 percent confidence intervals overlap the observed half-life in the corresponding city pairs.

We emphasize that the predictions greatly improve compared to the model with full attention. When  $m = 1$ , the first-order autocorrelation of the aggregate real exchange rate is only 0.34 because  $\rho_q = \lambda = 0.34$ . This low persistence of the aggregate real exchange rate translates into a very short half-life of 0.64 years. Thus, the behavioral inattention lengthens the half-life of the aggregate real exchange rate in the US–Canadian city pairs by 1.98 – 3.06 years (i.e.,  $2.62 - 0.64 = 1.98$  years and  $3.70 - 0.64 = 3.06$  years). Likewise, in the UK–Euro area city pairs, the half-life becomes longer by 1.40–1.90 years (i.e.,  $2.04 - 0.64 = 1.40$  years and  $2.54 - 0.64 = 1.9$  years).

---

<sup>27</sup>We calculate the half-lives for the AR(1) process from the standard formula given by  $-\ln(2)/\ln \rho$ , where  $\rho$  is the AR(1) coefficient.

<sup>28</sup>Note that we have multiple aggregate real exchange rates for the UK–Euro area city pairs because the consumer price indices differ across Euro-area countries. The half-life of 2.43 years reported in Table 3 for the UK–Euro area city pairs results from the mean of estimated half-lives in each country pair to which the UK–Euro area city pairs belong.



## 5.2 Persistence of the good-level real exchange rate

We next turn to the good-level real exchange rate. We let  $\rho_{qi}$  be the first-order autocorrelation of the good-level real exchange rate implied by (29). The following proposition describes the relationship between the persistence of the good-level real exchange rates and that of the aggregate real exchange rate, as predicted by the model.

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

$$\rho_q \geq \rho_{qi}, \quad (44)$$

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

**Proof.** See Appendix A.7. ■

Proposition 3 explains the stylized fact that good-level real exchange rates are much less persistent than the aggregate real exchange rate. Importantly, we obtain this aggregation result without relying on the “aggregation bias” pointed out by Imbs et al. (2005). They emphasized that the heterogeneity in the persistence of the good-level real exchange rates induces a positive bias in the persistence of the aggregate real exchange rate. Using multi-sector sticky-price models with heterogeneity in the degree of price stickiness, Carvalho and Nechio (2011) successfully explain the positive bias. By contrast, our model intentionally assumes homogeneity in the persistence across goods. Nevertheless, our model can qualitatively explain the gap in persistence between the aggregate and the good-level real exchange rates.

Once again, the value of  $m$  plays a crucial role in generating the gap between  $\rho_q$  and  $\rho_{qi}$ . This point can be further investigated from the  $\rho_q$  to  $\rho_{qi}$  ratio defined by:

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

where

$$A = (1 - \lambda)^2 (1 - \lambda \delta)^2 \psi^2 \frac{1 - \rho_q^2}{\rho_q^2 (1 - \lambda^2)} \left( \frac{\sigma_r}{\sigma_n} \right)^2. \quad (46)$$

The derivation is in Appendix A.7. Similar to the  $\rho_q$  to  $\lambda$  ratio in (43), the  $\rho_q$  to  $\rho_{qi}$  ratio indicates that  $\rho_q = \rho_{qi}$  if  $m = 1$ . Therefore, combined with the result from (43), full attention leads to the complete failure to explain the PPP puzzle:  $\rho_q = \rho_{qi} = \lambda$ .

While the behavioral inattention ( $m < 1$ ) is necessary for  $\rho_{qi} < \rho_q$ , it is not sufficient for explaining the gap between  $\rho_q$  and  $\rho_{qi}$ . What is additionally needed is real friction. More specifically, trade cost ( $\tau$ ) needs to be strictly positive and the elasticity of substitution across brands ( $\varepsilon$ ) needs to be larger than one for the  $\rho_q$  to  $\rho_{qi}$  ratio (45) to be strictly greater than one. If  $\tau = 0$  or  $\varepsilon \rightarrow 1$ , there is no home bias ( $\omega = 1/(1 + (1 + \tau)^{1-\varepsilon}) = 1/2$ ) so that  $\psi = 2\omega - 1 = 0$ . According to (46), either  $\tau = 0$  or  $\varepsilon \rightarrow 1$  makes  $A$  zero and thus (45) becomes one. Likewise,  $\sigma_r/\sigma_n$ , namely the standard deviation ratio of real shocks ( $\varepsilon_{it}^r$ ) to nominal shocks ( $\varepsilon_t^n$ ), in (46) needs to be strictly positive. If the nominal shock fully dominates the real shock such that  $\sigma_r/\sigma_n \rightarrow 0$ ,  $A$  is again zero, such that the model fails to generate the gap between  $\rho_q$  and  $\rho_{qi}$ .

To assess the effect of  $m$  on the gap between  $\rho_q$  and  $\rho_{qi}$ , we calibrate the parameters in (45) and (46). For the parameters of real frictions, we set  $\tau$  to 74 percent from Anderson and van Wincoop (2004) and  $\varepsilon$  to 4 from Broda and Weinstein (2006).<sup>29,30</sup> Using these values, we obtain the degree of home bias  $\omega$  of 0.84, which is roughly consistent with the parameter for home bias used in the literature.<sup>31</sup> The resulting calibrated value of  $\psi$  becomes 0.68. Crucini et al. (2013) found that  $\sigma_r/\sigma_n = 5$  is a sensible estimate of the standard deviation ratio, based on the sectoral real exchange rate data in Europe. The households' discount factor  $\delta$  is set to 0.98, and the degree of price stickiness  $\lambda$  is again set to 0.34.

Figure 3 shows the extent to which the good-level real exchange rate becomes less persistent than the aggregate real exchange rate against the degree of attention. The left panel plots  $\rho_{qi}$  against  $m \in (0, 1]$  in the dashed line. It also includes the curve for  $\rho_q$  taken from the solid line in Figure 2. As suggested by Proposition 3, the curve for  $\rho_{qi}$  is always located below the curve for  $\rho_q$ . Recall that the lower bound of  $\rho_q$  is  $\lambda (= 0.34)$  at  $m = 1$ . This property is preserved for  $\rho_{qi}$  because  $\rho_q = \rho_{qi} = \lambda$  hold at  $m = 1$ . The right panel represents the  $\rho_q$  to  $\rho_{qi}$  ratio along with the lower bound of unity. The panel indicates that the  $\rho_q$  to  $\rho_{qi}$  ratio is hump shaped against  $m \in (0, 1]$ . The  $\rho_q$  to  $\rho_{qi}$  ratio is one, when  $m \rightarrow 0$  or  $m = 1$ . We reconfirm this from the left panel of the same figure. When  $m$  is either zero or one, we have  $\rho_q = \rho_{qi}$  so that  $\rho_q/\rho_{qi} = 1$  holds.<sup>32</sup> However, when  $0 < m < 1$ , the  $\rho_q$  to  $\rho_{qi}$  ratio exceeds

<sup>29</sup>Using US data, Anderson and van Wincoop (2004) argue that the transportation costs are 21 percent and that the border-related trade barriers are 44 percent. Using these values, they calculate total international trade costs as  $0.74 (= 1.21 \times 1.44 - 1)$ .

<sup>30</sup>Broda and Weinstein (2006) report that the medians of the 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 of the SITC.

<sup>31</sup>For example, Chari et al. (2002) calibrate the degree of home bias as 0.76, while Steinsson (2008) uses 0.94.

<sup>32</sup>Analytically, the results can be understood as follows. When  $m = 1$ ,  $\rho_q/\rho_{qi} = 1$  immediately follows from

unity.

The estimated degrees of attention suggest that inattention reduces the persistence of the good-level real exchange rate to less than two-thirds of that of the aggregate real exchange rate. Suppose that  $m = 0.16$ , which is the estimated degree of attention in the US–Canadian city pairs (See specification (1) of Table 1). The left panel of Figure 3 shows that  $\rho_{qi}$  is around 0.49 whereas  $\rho_q$  is 0.76. We can see from the right panel that the  $\rho_q$  to  $\rho_{qi}$  ratio is 1.55. Equivalently, we see that  $\rho_{qi}$  is less than two-thirds of  $\rho_q$  (i.e.,  $0.49/0.76 < 2/3$ ). We can also confirm that the estimated degree of attention in the UK–Euro area city pairs generates similar results. When we take  $m = 0.21$  from specification (1) of Table 1 for the UK–Euro area city pairs,  $\rho_{qi} = 0.45$ ,  $\rho_q = 0.71$ , and thus  $\rho_{qi}/\rho_q = 1.57$ . In other words,  $\rho_{qi}$  in the UK–Euro area city pairs is also less than two-thirds of  $\rho_q$  ( $0.45/0.71 < 2/3$ ).

The lower panel of Table 3 presents the predicted half-lives of the good-level real exchange rate, together with the half-lives observed in the data. The rightmost column reports the median half-lives of the good-level real exchange rates estimated from our dataset.<sup>33</sup> In the data over 1990–2015, we find that the half-life of the median goods is 1.61 years for the US–Canadian city pairs and 1.25 years for the UK–Euro area city pairs, both of which are much shorter than the half-life of the aggregate real exchange rate shown in the same column of the upper panel. The estimated half-lives are also consistent with previous studies who use the EIU data. For example, Crucini and Shintani (2008) find the half-life of the median goods to range from 1.03 to 1.61 years based on the EIU data from 1990–2005. Bergin et al. (2013) also use the EIU data and construct the good-level real exchange rates of twenty cities in industrial countries (including 16 European cities ) relative to the New York city during 1990 and 2007. When they estimate the AR(1) model for the good-level real exchange rates, the average half-life is 1.15 years.

The model with behavioral inattention explains the half-lives of the good-level real exchange rate quite well. When we calibrate  $m$  at  $m = 0.16$  for the US–Canadian city pairs and  $m = 0.21$  for the UK–Euro area city pairs (i.e.,  $\hat{m}$  in specification (1) of Table 1), the half-life is 0.98 years and 0.86 years, respectively (See the first column of the lower panel of Table

---

(45). When  $m \rightarrow 0$ ,  $\rho_q \rightarrow 1$  holds from the definition of  $\rho_q = \lambda/[1 - (1 - m)(1 - \lambda)]$ . As  $\rho_q \rightarrow 1$ ,  $A \rightarrow 0$ , which leads to  $\rho_q/\rho_{qi} \rightarrow 1$ .

<sup>33</sup>We estimate the panel AR(1) model of  $\ln q_{ijt}$  for each good  $i$ , using the generalized method of moments estimator of Arellano and Bond (1991). The first-order autocorrelation estimated from the panel AR(1) model is transformed into the half-life. Typically, the good-by-good panel consists of more than 1,450 observations in the US–Canadian city pairs and more than 900 observations for the UK–Euro area city pairs. Our median half-lives reported in Table 3 are chosen from half-lives in which the number of observations exceeds 500 in the estimation of the panel AR(1) model.

3).<sup>34</sup> When we take calibrate  $m$  at  $m = 0.11$  for the US–Canadian city pairs and  $m = 0.16$  for the UK–Euro area city pairs (i.e.,  $\hat{m}$  in specification (1) of Table 2), the half-life becomes closer to the data: 1.22 years for the US–Canadian city pairs and 0.97 for the UK–Euro area city pairs. Note that if we allow for a range of half-life implied from 95 percent confidence intervals of  $\hat{m}$ , the range of the predicted half-life overlaps with the observed half-life in both the US–Canadian city pairs and the UK–Euro area city pairs.

The success of our model with behavioral inattention in predicting half-lives of both the aggregate and the good-level real exchange rates means that our model can explain the gap in persistence between the aggregate and the good-level real exchange rates. In the data, the difference in half-lives between the aggregate and the good-level real exchange rates is 3.32 years for the US–Canadian city pairs and 1.18 years for the UK–Euro area city pairs. The corresponding differences predicted by the model is 1.64–2.48 years in the US–Canadian city pairs (i.e.,  $2.62 - 0.98 = 1.64$  years when using  $\hat{m}$  in Table 1 and  $3.70 - 1.22 = 2.48$  years when using  $\hat{m}$  in Table 2). We can also confirm similar results for the UK–Euro area city pairs: The difference in half-lives are 1.18–1.57 years ( $2.04 - 0.86 = 1.18$  years when using  $\hat{m}$  in Table 1 and  $2.54 - 0.97 = 1.57$  years when using  $\hat{m}$  in Table 2).

Before closing this section, two remarks are in order. First, it is straightforward to combine Propositions 2 and 3 to obtain the  $\rho_{qi}$  to  $\lambda$  ratio that measures the amplification from  $\lambda$  to  $\rho_{qi}$ . In particular, using (43) and (45), we have

$$\frac{\rho_{qi}}{\lambda} = \frac{1 - (1 - m)(1 - \lambda) \left(\frac{A}{1+A}\right)}{1 - (1 - m)(1 - \lambda)} \geq 1. \quad (47)$$

Given that the equality is excluded as long as  $A > 0$  and  $m < 1$ , the persistence of the good-level real exchange rate exceeds  $\lambda$ . The result is also consistent with Kehoe and Midrigan’s (2007) finding that even the persistence of the good-level real exchange rate is more persistent than what is predicted only by the degree of price stickiness. In equation, the relationship can be summarized as  $\rho_q > \rho_{qi} > \lambda$ . Translating this relationship into our theoretical prediction of the half-lives in Table (3), our model of behavioral inattention produces 2.62–3.70 years for the aggregate real exchange rate, 0.98–1.22 years for the good-level real exchange rate, and 0.64 years for the half-life predicted only by the degree of price stickiness.

Second, our model of behavioral inattention can reproduce the findings by Bergin et al. (2013), who analyze the persistence of good-level real exchange rate *conditional on shocks*. Using a vector error correction model for each good, they find that the good-level real ex-

---

<sup>34</sup>We calculate the half-life of the good-level real exchange rate by  $-\ln(2)/\ln(\rho_{qi})$  because  $\rho_{qi}$  asymptotically equivalent to the AR(1) coefficient.

change rate is as persistent as the aggregate real exchange rate, conditional on macroeconomic shock. We can analyze  $\hat{q}_{it}$  conditional on macroeconomic shock by setting  $\sigma_r = 0$ . As we discuss earlier,  $\sigma_r = 0$  implies  $A = 0$ . Therefore, (45) and (47) implies that  $\rho_q = \rho_{qi} > \lambda$ , which is consistent with the empirical finding by Bergin et al. (2013).

## 6 Conclusion

In this paper, we offer a possible explanation for two empirical anomalies. First, the observed PPP deviations are much more persistent than the theoretical predictions given by the standard model of nominal rigidities in prices. Second, the micro price evidence suggests that the deviations from the LOP are often less persistent than the PPP deviations. To reconcile the PPP and LOP evidence, we adapted the model of behavioral inattention in Gabaix (2014) to a simple two-country, sticky-price model. We showed that pricing by inattentive firms generates the complementarity between the LOP and PPP deviations, which is the key to accounting for the puzzling behavior of real exchange rates.

Using international price data, we implemented a test of behavioral inattention and quantified its importance. We found strong evidence consistent with behavioral inattention. The complementarity in our model with behavioral inattention produces an aggregate real exchange rate that is more than twice as persistent as the real exchange rate explained only by sticky prices. Our model also predicts that the persistence of the LOP deviations is less than two-thirds of the persistence of the PPP deviations. We showed that the model quantitatively replicates the observed half-lives of both the aggregate and the good-level real exchange rates.

Based upon our examination of the behavioral inattention hypothesis, it seems plausible that it plays a comparable role to other real rigidities in the existing real exchange rate literature while also amplifying some prominent existing mechanisms such as sticky prices. The avenues for further exploration appear to be quite promising.

## References

- [1] Álvarez, L. J., E. Dhyne, M. Hoeberichts, C. Kwapil, H. Le Bihan, P. Lünnemann, 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] Ball, L. and D. Romer (1990). "Real rigidities and the non-neutrality of money," *Review of Economic Studies*, vol. 57(2), pp. 183-203.
- [6] Benigno, G. (2004). "Real exchange rate persistence and monetary policy rules," *Journal of Monetary Economics*, vol. 51(3), pp. 473-502.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Blanco, A. and J. Cravino (2020). "Price rigidities and the relative PPP," *Journal of Monetary Economics*, vol. 116, pp. 104-116.
- [11] Broda, C. and D. E. Weinstein (2006). "Globalization and the gains from variety," *Quarterly Journal of Economics*, vol. 121(2), pp. 541-585.
- [12] Calvo, G. A. (1983). "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, vol. 12(3), pp. 383-398.
- [13] Carvalho, C. and F. Nechio (2011). "Aggregation and the PPP puzzle in a sticky-price model," *American Economic Review*, vol. 101(6), pp. 2391-424.
- [14] 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.

- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] Crucini, M. J., C. I. Telmer (2020). “Microeconomic sources of real exchange rate variation,” *Review of Economic Dynamics*, vol. 38, pp. 22-40.
- [21] Crucini, M. J. and H. Yilmazkuday (2014). “Understanding long-run price dispersion,” *Journal of Monetary Economics*, vol. 66, pp. 226-240.
- [22] 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.
- [23] Engel, C. (2019). “Real exchange rate convergence: The roles of price stickiness and monetary policy,” *Journal of Monetary Economics*, vol. 103, pp. 21-32.
- [24] Engel, C. and J. H. Rogers (2004). “European product market integration after the Euro,” *Economic Policy*, vol. 19(39), pp. 348-384.
- [25] Gabaix, X. (2014). “A sparsity-based model of bounded rationality,” *Quarterly Journal of Economics*, vol. 129(4), pp. 1661-1710.
- [26] Gabaix, X. (2020). “A behavioral New Keynesian model,” *American Economic Review*, vol. 110(8), pp. 2271-2327.
- [27] Galí, J. (2015). *Monetary policy, Inflation, and the Business Cycle: an Introduction to the New Keynesian Framework and Its Applications*, Princeton University Press.

- [28] 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 (2022). “New facts on consumer price rigidity in the Euro area,” Unpublished manuscript.
- [29] 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.
- [30] 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.
- [31] Kehoe, P. J. and V. Midrigan (2007). “Sticky prices and sectoral real exchange rates,” Federal Reserve Bank of Minneapolis, Research Department.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] Rogoff, K. (1996). “The purchasing power parity puzzle,” *Journal of Economic Literature*, vol. 34(2), pp. 647-668.
- [36] Steinsson, J. (2008). “The dynamic behavior of the real exchange rate in sticky price models,” *American Economic Review*, vol. 98(1), pp. 519-533.
- [37] Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*: Princeton University Press.
- [38] Yun, T. (1996). “Nominal price rigidity, money supply endogeneity, and business cycles,” *Journal of Monetary Economics*, vol. 37(2), pp. 345-370.



# A Appendix

## A.1 The derivation of the objective function for the pricing decision

To derive (6) and (7), we begin with the standard expression.<sup>35</sup> The objective function of US firms that sell their brand in US markets is given by:

$$v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} (1/P_{t+k}) \left[ P_{it}(z) - \frac{W_{t+k}}{a_{it+k}} \right] c_{it,t+k}(z), \quad (48)$$

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

$$c_{it,t+k}(z) = \left[ \frac{P_{it}(z)}{P_{it+k}} \right]^{-\varepsilon} c_{it+k}, \quad (49)$$

where  $z \in [0, 1/2]$ . Using the definitions of  $p_{it}(z)$ ,  $w_t$ , and  $p_{it}$ , we rewrite (48) as:

$$v_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} \frac{P_t}{P_{t+k}} \left[ p_{it}(z) - \frac{w_{t+k} P_{t+k}}{a_{it+k} P_t} \right] c_{it,t+k}(z). \quad (50)$$

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, by assumption,  $P_{t+k}/P_t$  and  $a_{it}$  are both unity in the steady state. Rewriting (50) yields (6):

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

where  $P_{t+k}/P_t = \prod_{l=1}^k P_{t+l}/P_{t+l-1} = \exp \left[ \sum_{l=1}^k \ln(P_{t+l}/P_{t+l-1}) \right] = \exp \left[ \sum_{l=1}^k \pi_{t+l} \right]$ . For the demand function, we can rewrite (49) as  $c_{it,t+k}(z) = [P_{it}(z)/P_{it+k}]^{-\varepsilon} c_{it+k} = [(P_{it}(z)/P_t)/(P_{it+k}/P_{t+k}) \times (P_t/P_{t+k})]^{-\varepsilon} c_{it+k} = [(p_{it}(z)/p_{it+k})(P_t/P_{t+k})]^{-\varepsilon} c_{it+k}$ . Using the log deviation, we can derive (7):

$$c_{it,t+k}(z) = \left( \frac{\bar{p}_i(z)}{\bar{p}_i} \right)^{-\varepsilon} \exp \left\{ -\varepsilon \left[ \hat{p}_{it}(z) - \sum_{l=1}^k \pi_{t+l} - \hat{p}_{it+k} \right] \right\} c_{it+k}. \quad (51)$$

We next work on the derivation of (8) and (9). When US firms sell their brands in Canadian markets, they set the price in the local currency. Under this assumption, the

<sup>35</sup>For example, see Galí (2015).

objective function of these firms is

$$v_{it}^*(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} (1/P_{t+k}) \left[ S_{t+k} P_{it}^*(z) - (1 + \tau) \frac{W_{t+k}}{a_{it+k}} \right] c_{it+k}^*(z), \quad (52)$$

subject to the demand function by Canadian consumers:

$$c_{it+k}^*(z) = \left[ \frac{P_{it}^*(z)}{P_{it+k}^*} \right]^{-\varepsilon} c_{it+k}^*, \quad (53)$$

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 (52) 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  $x_t = \bar{x} \exp(\hat{x}_t)$  and assuming the zero-inflation steady state, we obtain (8):

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

Equation (9) can be derived from (53) in the same way as the derivation of (7) from (49).

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, their objective function is

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

for  $z \in (1/2, 1]$ . Similarly, when Canadian firms sell their brands in US markets, the objective function is

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

for  $z \in (1/2, 1]$ .

## A.2 The sparse max

Following Gabaix (2014), we assume that firms choose the degree of attention. Equations (13) and (14) correspond to the case of US firms that sell their goods in the US market. The US firms' objective function for choosing  $m_H$  is based on the second-order Taylor expansion of  $\mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1] - \mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, 1), \hat{\boldsymbol{\mu}}_{Ht}, 1]$  around  $\hat{\boldsymbol{\mu}}_{Ht} = \mathbf{0}$ , which is the loss of profits of choosing the price distorted by partial attention. In this appendix, we derive (13) and (14).

To obtain (13) and (14), we first take the approximation of  $\mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1]$  around  $\hat{\boldsymbol{\mu}}_{Ht} = \mathbf{0}$ . Here, the profit of the firm is evaluated at  $m_H = 1$  (which appears in the last argument in  $v_{Hi}(\cdot)$ ), but the price is distorted by  $m_H \neq 1$ . We next evaluate the price in the approximated equation at  $m_H = 1$ . The second-order approximation of

$\mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1]$  around  $\hat{\boldsymbol{\mu}}_{Ht} = \mathbf{0}$  is

$$\begin{aligned} & \mathbb{E}v_{Hi}[\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1] \\ \simeq & v_{Hi}^0 + \frac{1}{2} \left\{ \sum_{k=0}^{\infty} \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2} \left[ \frac{\partial \hat{p}_{Hit}(\mathbf{0}, m_H)}{\partial \hat{\boldsymbol{\mu}}_{Ht+k}} \right]^2 + \sum_{k=0}^{\infty} \frac{\partial^2 v_{Hi}^0}{\partial \hat{\boldsymbol{\mu}}_{Ht+k}^2} \right\} \mathbb{E}\hat{\boldsymbol{\mu}}_{Ht+k}^2 \\ & + \sum_{k=0}^{\infty} \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z) \partial \hat{\boldsymbol{\mu}}_{Ht+k}} \left[ \frac{\partial \hat{p}_{Hit}(\mathbf{0}, m_H)}{\partial \hat{\boldsymbol{\mu}}_{Ht+k}} \right] \mathbb{E}\hat{\boldsymbol{\mu}}_{Ht+k}^2, \end{aligned} \quad (54)$$

where  $v_{Hi}^0 = v_{Hi}[\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1]_{\hat{\boldsymbol{\mu}}_{Ht}=\mathbf{0}} = v_{Hi}(\hat{p}_{Hi}(\mathbf{0}, m_H), \mathbf{0}, 1)$ . For the second derivatives,  $\partial^2 v_{Hi}^0 / \partial \hat{p}_{it}(z)^2 = \partial^2 v_{Hi}(\hat{p}_{Hi}(\mathbf{0}, m_H), \mathbf{0}, 1) / \partial \hat{p}_{it}(z)^2$  and  $\partial^2 v_{Hi}^0 / \partial \hat{\boldsymbol{\mu}}_{Ht+k}^2 = \partial^2 v_{Hi}(\hat{p}_{Hi}(\mathbf{0}, m_H), \mathbf{0}, 1) / \partial \hat{\boldsymbol{\mu}}_{Ht+k}^2$ .

We use the first-order condition for pricing of inattentive firms to simplify (54). The first-order condition is  $\partial v_{Hi}[\hat{p}_{it}(z), \hat{\boldsymbol{\mu}}_{Ht}, m_H] / \partial \hat{p}_{it}(z) = 0$ . Taking the partial derivative of the first-order conditions with respect to  $\hat{\boldsymbol{\mu}}_{Ht+k}$  for  $k = 0, 1, 2, \dots$  and evaluating them at  $\hat{\boldsymbol{\mu}}_{Ht} = \mathbf{0}$ :

$$\frac{\partial^2 v_{Hi}(\hat{p}_{Hit}(\mathbf{0}, m_H), \mathbf{0}, m_H)}{\partial \hat{p}_{it}(z) \partial \hat{\boldsymbol{\mu}}_{Ht+k}} = - \frac{\partial^2 v_{Hi}(\hat{p}_{Hit}(\mathbf{0}, m_H), \mathbf{0}, m_H)}{\partial \hat{p}_{it}(z)^2} \frac{\partial \hat{p}_{Hit}(\mathbf{0}, m_H)}{\partial \hat{\boldsymbol{\mu}}_{Ht+k}}, \quad \text{for } k = 0, 1, 2, \dots \quad (55)$$

Let us focus on  $\partial \hat{p}_{Hit}(\mathbf{0}, m_H) / \partial \hat{\boldsymbol{\mu}}_{Ht+k}$  in the right-hand side of (55). The optimal price is given by  $\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) = m_H \hat{w}_t - (1 - \lambda\delta) \hat{a}_{it} = m_H \hat{\boldsymbol{\mu}}_{Ht} - (1 - \lambda\delta) \hat{a}_{it}$  (see (20)). Thus,

$$\frac{\partial \hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H)}{\partial \hat{\boldsymbol{\mu}}_{Ht+k}} = \begin{cases} m_H & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases} \quad (56)$$

When we evaluate the profits in (55) at  $m_H = 1$  but not the prices, (55) can be substituted into (54). Then, together with (56), we now simplify (54) to:

$$\begin{aligned} & \mathbb{E}v_{Hi}[\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1] \\ \simeq & v_{Hi}^0 + \frac{1}{2} \left[ \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2} m_H^2 - 2 \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2} m_H \right] \mathbb{E}(\hat{\boldsymbol{\mu}}_{Ht}^2) + \frac{1}{2} \sum_{k=0}^{\infty} \frac{\partial^2 v_{Hi}^0}{\partial \hat{\boldsymbol{\mu}}_{Ht+k}^2} \mathbb{E}(\hat{\boldsymbol{\mu}}_{Ht+k}^2). \end{aligned} \quad (57)$$

We further need the second-order approximation of  $\mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1]$  where the price is not distorted by  $m_H$ . Evaluating (57) at  $m_H = 1$  yields

$$\mathbb{E}v_{Hi}[\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, 1), \hat{\boldsymbol{\mu}}_{Ht}, 1] \simeq v_{Hi}^0 - \frac{1}{2} \left[ \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2} \right] \mathbb{E}(\hat{\boldsymbol{\mu}}_{Ht}^2) + \frac{1}{2} \sum_{k=0}^{\infty} \frac{\partial^2 v_{Hi}^0}{\partial \hat{\boldsymbol{\mu}}_{Ht+k}^2} \mathbb{E}(\hat{\boldsymbol{\mu}}_{Ht+k}^2). \quad (58)$$

Combining (57) and (58),  $\mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1] - \mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, 1), \hat{\boldsymbol{\mu}}_{Ht}, 1]$  around

$\hat{\mu}_{Ht} = 0$  is

$$\begin{aligned}
& \mathbb{E}v_{Hi}[\hat{p}_{Hi}(\hat{\mu}_{Ht}, m_H), \hat{\mu}_{Ht}, 1] - \mathbb{E}v_{Hi}[\hat{p}_{Hi}(\hat{\mu}_{Ht}, 1), \hat{\mu}_{Ht}, 1] \\
& \simeq \frac{1}{2} (m_H^2 - 2m_H + 1) \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2} \mathbb{E}(\hat{\mu}_{Ht}^2) \\
& = \frac{1}{2} (1 - m_H)^2 \frac{\partial^2 v_{Hi}^0}{\partial \hat{p}_{it}(z)^2} \mathbb{E}(\hat{\mu}_{Ht}^2) \\
& = -\frac{1}{2} (1 - m_H)^2 \Lambda_H,
\end{aligned} \tag{59}$$

where we used (14) for the last equality.

While firms can reduce the loss of paying partial attention (59) by paying more attention, they also have to pay costs of increasing attention, which we specify as a quadratic cost function  $\mathcal{C}(m_H) = (\kappa/2)m_H^2$ . Formally, the choice of attention for US firms that sell their goods in US markets is characterized by:

$$\min_{m_H \in [0,1]} \frac{1}{2} [(1 - m_H)^2 \Lambda_H] + \frac{\kappa}{2} m_H^2, \text{ where } \Lambda_H = - \left\{ \frac{\partial^2 v_{Hi}}{\partial \hat{p}_{it}^2(z) [0, \mathbf{0}, 1]} \right\} Var(\hat{\mu}_{Ht}).$$

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

$$\min_{m_H^* \in [0,1]} \frac{1}{2} (1 - m_H^*)^2 \Lambda_H^* + \frac{\kappa}{2} (m_H^*)^2, \text{ where } \Lambda_H^* = - \left\{ \frac{\partial^2 v_{Hi}^*[0, \mathbf{0}, 1]}{\partial \hat{p}_{Hit}^{*2}} \right\} Var(\hat{\mu}_{Ht}^*).$$

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

$$\min_{m_F^* \in [0,1]} \frac{1}{2} (1 - m_F^*)^2 \Lambda_F^* + \frac{\kappa}{2} (m_F^*)^2, \text{ where } \Lambda_F^* = - \left\{ \frac{\partial^2 v_{Fi}^*[0, \mathbf{0}, 1]}{\partial \hat{p}_{Fit}^{*2}} \right\} Var(\hat{\mu}_{Ft}^*).$$

By symmetry, we can easily show that  $\Lambda_F^* = \Lambda_H$ , which reconfirms  $m_F^* = m_H$ . the sparse max for Canadian firms selling their goods in US markets is

$$\min_{m_F \in [0,1]} \frac{1}{2} (1 - m_F)^2 \Lambda_F + \frac{\kappa}{2} m_F^2, \text{ where } \Lambda_F = - \left\{ \frac{\partial^2 v_{Fi}[0, \mathbf{0}, 1]}{\partial \hat{p}_{Fit}^2} \right\} Var(\hat{\mu}_{Ft}).$$

Again, by symmetry, we have  $\Lambda_F = \Lambda_H^*$  and  $m_F = m_H^*$ .

### A.3 The optimal prices under behavioral inattention

Using the definition of  $\hat{\mu}_{Ht+k}$ , we rewrite the log-linearized first-order condition (19) as

$$p_{Hi}(\hat{\mu}_{Ht}, m_H) = (1 - \lambda\delta)\mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k (m_H \hat{w}_{t+k} - \hat{a}_{it+k}) + m_H (1 - \lambda\delta) \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k \sum_{l=1}^k \pi_{t+l}, \quad (60)$$

where we used  $\mu_{Ht+k} = \hat{w}_{t+k} + \sum_{l=1}^k \pi_{t+l}$ .

We separately arrange the terms in the right-hand side of (60). First, note that

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

Next, the remaining terms are

$$\begin{aligned} & m_H (1 - \lambda\delta) \mathbb{E}_t \sum_{k=0}^{\infty} (\lambda\delta)^k \sum_{l=1}^k \pi_{t+l} \\ = & m_H (1 - \lambda\delta) \mathbb{E}_t \left[ \begin{array}{l} (\lambda\delta)\pi_{t+1} \\ + (\lambda\delta)^2 \pi_{t+1} + (\lambda\delta)^2 \pi_{t+2} \\ + (\lambda\delta)^3 \pi_{t+1} + (\lambda\delta)^3 \pi_{t+2} + (\lambda\delta)^3 \pi_{t+3} \\ + \dots \end{array} \right] \\ = & m_H (1 - \lambda\delta) \mathbb{E}_t \left\{ (\lambda\delta) \left[ \sum_{k=0}^{\infty} (\lambda\delta)^k \right] \pi_{t+1} + (\lambda\delta)^2 \left[ \sum_{k=0}^{\infty} (\lambda\delta)^k \right] \pi_{t+2} + (\lambda\delta)^3 \left[ \sum_{k=0}^{\infty} (\lambda\delta)^k \right] \pi_{t+3} + \dots \right\} \\ = & m_H \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k \pi_{t+k}, \end{aligned}$$

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

(60) becomes

$$\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) = (m_H \hat{w}_t - \hat{a}_{it}) + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda \delta)^k \{m_H (\Delta \hat{w}_{t+k} + \pi_{t+k}) - \Delta \hat{a}_{it+k}\}. \quad (61)$$

Now, under the assumption of  $U(c, n) = \ln c - \chi n_t$ , the first-order conditions of US households ( $W_t/P_t = \chi c_t$ ) imply  $\hat{w}_t = \hat{c}_t$ . 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 (15), we have  $\Delta \hat{w}_t + \pi_t = \ln M_t/M_{t-1} = \varepsilon_t^M$ . As a result, (61) becomes

$$\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) = (m_H \hat{w}_t - \hat{a}_{it}) - \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda \delta)^k \Delta a_{it+k}.$$

If the stochastic process  $\hat{a}_{it}$  is given by (17),  $\mathbb{E}_t \sum_{k=1}^{\infty} (\lambda \delta)^k \Delta a_{it+k} = -\lambda \delta \hat{a}_{it}$ . Therefore,

$$\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) = m_H \hat{w}_t - (1 - \lambda \delta) \hat{a}_{it},$$

which is (20) in the main text.

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

$$\begin{aligned} \hat{p}_{Hi}^*(\hat{\boldsymbol{\mu}}_{Ht}^*, m_H^*) &= [m_H^* (\hat{w}_t - \hat{q}_t) - \hat{a}_{it}] + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda \delta)^k [m_H (\Delta \hat{w}_{t+k} - \Delta \hat{q}_{t+k} + \pi_{t+k}^*) - \Delta \hat{a}_{it+k}] \\ &= (m_H^* \hat{w}_t^* - \hat{a}_{it}) + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda \delta)^k [m_H (\Delta \hat{w}_{t+k}^* + \pi_{t+k}^*) - \Delta \hat{a}_{it+k}], \end{aligned} \quad (62)$$

where we used the log-linearized equation of (3):  $\hat{q}_t = \hat{c}_t - \hat{c}_t^* = \hat{w}_t - \hat{w}_t^*$ . This equation has the same structure as (61). Using the CIA constraint, (16), and (17), the above equation can be simplified to

$$\hat{p}_{Hi}^*(\hat{\boldsymbol{\mu}}_{Ht}^*, m_H^*) = m_H^* (\hat{w}_t - \hat{q}_t) - (1 - \lambda \delta) \hat{a}_{it},$$

which is equivalent to (21).

The remaining optimal prices, namely  $\hat{p}_{Fi}^*(\hat{\boldsymbol{\mu}}_{Ft}^*, m_F^*)$  and  $\hat{p}_{Fi}(\hat{\boldsymbol{\mu}}_{Ft}, m_F)$  are analogously derived.

## A.4 Proof of Proposition 1

We begin by (28),  $\hat{q}_{it} = \hat{q}_t + \hat{p}_{it}^* - \hat{p}_{it}$ . From (3), the log deviation of the real exchange rate is

$$\hat{q}_t = \hat{c}_t - \hat{c}_t^*. \quad (63)$$

Thus,  $\hat{q}_{it}$  can be rewritten as:

$$\hat{q}_{it} = (\hat{p}_{it}^* - \hat{c}_t^*) - (\hat{p}_{it} - \hat{c}_t). \quad (64)$$

In what follows, we focus on  $\hat{p}_{it} - \hat{c}_t$  and  $\hat{p}_{it}^* - \hat{c}_t^*$  to derive (29). Equation (24) implies

$$\begin{aligned} \hat{p}_{it} - \hat{c}_t &= \lambda(\hat{p}_{it-1} - \pi_t) + (1 - \lambda)\hat{p}_{it}^{opt} - \hat{c}_t \\ &= \lambda(\hat{p}_{it-1} - \hat{c}_{t-1}) - \lambda(\Delta\hat{c}_t + \pi_t) + (1 - \lambda)(\hat{p}_{it}^{opt} - \hat{c}_t). \end{aligned} \quad (65)$$

Note that  $\Delta\hat{c}_t + \pi_t$  in (65) is equal to  $\varepsilon_t^M$  because of the CIA constraint of US households and the money supply process (15). Substituting this result yields

$$\hat{p}_{it} - \hat{c}_t = \lambda(\hat{p}_{it-1} - \hat{c}_{t-1}) - \lambda\varepsilon_t^M + (1 - \lambda)(\hat{p}_{it}^{opt} - \hat{c}_t). \quad (66)$$

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

$$\hat{p}_{it}^* - \hat{c}_t^* = \lambda(\hat{p}_{it-1}^* - \hat{c}_{t-1}^*) - \lambda\varepsilon_t^{M*} + (1 - \lambda)(\hat{p}_{it}^{opt*} - \hat{c}_t^*). \quad (67)$$

Substituting (66) and (67) into (64) yields an expression for  $\hat{q}_{it}$ :

$$\hat{q}_{it} = \lambda\hat{q}_{it-1} + \lambda\varepsilon_t^n + (1 - \lambda) [(\hat{p}_{it}^{opt*} - \hat{c}_t^*) - (\hat{p}_{it}^{opt} - \hat{c}_t)], \quad (68)$$

where  $\varepsilon_t^n = \varepsilon_t^M - \varepsilon_t^{M*}$ .

We next focus on the expression inside the bracket on the right-hand side of (68). Using (20), (23), (25), and (63), we rewrite  $\hat{p}_{it}^{opt}$  as

$$\hat{p}_{it}^{opt} = m\hat{c}_t - (1 - \lambda\delta)[\omega\hat{a}_{it} + (1 - \omega)\hat{a}_{it}^*], \quad (69)$$

where the relative price index is determined by the aggregate demand  $\hat{c}_t$  and the weighted average of labor productivity. In the equation, we also use the degree of attention  $m$  defined



by (30). We then subtract  $\hat{c}_t$  from both sides of (69) to get

$$\hat{p}_{it}^{opt} - \hat{c}_t = -(1-m)\hat{c}_t - (1-\lambda\delta)[\omega\hat{a}_{it} + (1-\omega)\hat{a}_{it}^*]. \quad (70)$$

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

$$\hat{p}_{it}^{opt*} - \hat{c}_t^* = -(1-m)\hat{c}_t^* - (1-\lambda\delta)[\omega\hat{a}_{it}^* + (1-\omega)\hat{a}_{it}], \quad (71)$$

where  $m = \omega m_F^* + (1-\omega)m_H^* = \omega m_H + (1-\omega)m_F$ . Combining (70) and (71), we have

$$\begin{aligned} (\hat{p}_{it}^{opt*} - \hat{c}_t^*) - (\hat{p}_{it}^{opt} - \hat{c}_t) &= (1-m)(\hat{c}_t - \hat{c}_t^*) + (1-\lambda\delta)(2\omega-1)(\hat{a}_{it} - \hat{a}_{it}^*) \\ &= (1-m)\hat{q}_t + (1-\lambda\delta)\psi\varepsilon_{it}^r, \end{aligned}$$

where  $\hat{q}_t = \hat{c}_t - \hat{c}_t^*$  from (63),  $\varepsilon_{it}^r = \varepsilon_{it} - \varepsilon_{it}^* = a_{it} - a_{it}^*$  from (17) and (18), and  $\psi = 2\omega - 1$ .

Substituting the above equation into (68) yields

$$\hat{q}_{it} = \lambda\hat{q}_{it-1} + (1-\lambda)(1-m)\hat{q}_t + \lambda\varepsilon_{it}^n + (1-\lambda)(1-\lambda\delta)\psi\varepsilon_{it}^r. \quad (72)$$

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  $\ln \bar{q}_i = \ln \bar{q} + \bar{p}_i^* - \bar{p}_i = 0$ . Therefore, (72) is equivalent to (29) in Proposition 1.

## A.5 The model with CRRA preferences

Let us assume more general CRRA preferences:  $U(c, n) = c^{1-\sigma}/(1-\sigma) - \chi n^{1+\varphi}/(1+\varphi)$ . We modify the first-order conditions for households to allow for the degree of relative risk aversion. Under  $\sigma \neq 1$ , the first-order conditions imply  $S_t = (M_t/M_t^*)^\sigma (P_t/P_t^*)^{1-\sigma}$ .

If we maintain the assumption that the money supply follows a random walk, the equation for  $S_t$  leads to the nominal exchange rate growth that is predictable through the inflation of the two countries.<sup>36</sup> Because this is inconsistent with the exchange-rate disconnect puzzle, we replace this assumption by the new assumption on the money growth rate:

$$\Delta \ln M_t = \frac{\sigma-1}{\sigma}\pi_t + \frac{1}{\sigma}\varepsilon_t^M, \quad (73)$$

$$\Delta \ln M_t^* = \frac{\sigma-1}{\sigma}\pi_t^* + \frac{1}{\sigma}\varepsilon_t^{M*}. \quad (74)$$

<sup>36</sup>In particular, the nominal exchange rate growth is given by  $\Delta s_t = \sigma\varepsilon_t^n + (1-\sigma)(\pi_t - \pi_t^*)$ , meaning that  $\pi_t - \pi_t^*$  can help forecast  $\Delta s_t$ .

Under (73) and (74), the nominal exchange rate continues to follow a random walk.<sup>37</sup>

Using the CIA constraints, we can rewrite (73) and (74) as:

$$\sigma \Delta \hat{c}_{t+k} + \pi_{t+k} = \varepsilon_{t+k}^M, \quad (75)$$

$$\sigma \Delta \hat{c}_{t+k}^* + \pi_{t+k}^* = \varepsilon_{t+k}^{M^*}, \quad (76)$$

for  $k > 0$ . Later, we utilize (75) and (76) for deriving the estimation equation.

### A.5.1 The derivation of the estimation equation

To derive the estimation equation, we follow the same procedure as the derivation of (29). When  $\sigma \neq 1$ , the international risk-sharing condition (3) is replaced by  $q_t = (c_t/c_t^*)^\sigma$ . Combining its log-linearized expression with (28),  $\hat{q}_{it}$  can be written as

$$\hat{q}_{it} = (\hat{p}_{it}^* - \sigma \hat{c}_t^*) - (\hat{p}_{it} - \sigma \hat{c}_t). \quad (77)$$

We focus on  $\hat{p}_{it} - \sigma \hat{c}_t$  and  $\hat{p}_{it}^* - \sigma \hat{c}_t^*$  and obtain the expression for  $\hat{q}_{it}$  using (77). Note that (24) remains valid even under the CRRA preferences. Therefore, we subtract  $\sigma \hat{c}_t$  from both sides of (24) and arrange terms to get

$$\hat{p}_{it} - \sigma \hat{c}_t = \lambda (\hat{p}_{it-1} - \sigma \hat{c}_{t-1}) - \lambda \varepsilon_t^M + (1 - \lambda) (\hat{p}_{it}^{opt} - \sigma \hat{c}_t), \quad (78)$$

where we replace  $\sigma \Delta \hat{c}_t + \pi_t$  by  $\varepsilon_t^M$  using (75). Analogously, (26) remains valid under the CRRA preferences. Using (26), we have

$$\hat{p}_{it}^* - \sigma \hat{c}_t^* = \lambda (\hat{p}_{it-1}^* - \sigma \hat{c}_{t-1}^*) - \lambda \varepsilon_t^{M^*} + (1 - \lambda) (\hat{p}_{it}^{opt*} - \hat{c}_t^*). \quad (79)$$

Therefore, the good-level real exchange rate is

$$\hat{q}_{it} = \lambda \hat{q}_{it-1} + \lambda \varepsilon_t^n + (1 - \lambda) [(\hat{p}_{it}^{opt*} - \sigma \hat{c}_t^*) - (\hat{p}_{it}^{opt} - \sigma \hat{c}_t)]. \quad (80)$$

Equations (78)–(80) generalize (66)–(68), respectively.

We next focus on the expression inside the brackets on the right-hand side of (80). For the case of  $\sigma \neq 1$ , we recalculate the log optimal prices:  $\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H)$ ,  $\hat{p}_{Hi}^*(\hat{\boldsymbol{\mu}}_{Ht}^*, m_H^*)$ ,

---

<sup>37</sup>To see this, note that the nominal exchange rate growth is given by:  $\Delta s_t = \sigma (\Delta \ln M_t - \Delta \ln M_t^*) + (1 - \sigma) (\pi_t - \pi_t^*)$ . Substituting (73) and (74) into the above equation yields  $\Delta s_t = (\sigma - 1) (\pi_t - \pi_t^*) + (1 - \sigma) (\pi_t - \pi_t^*) + \varepsilon_t^M - \varepsilon_t^{M^*} = \varepsilon_t^n$ .

$\hat{p}_{Fi}^*(\hat{\boldsymbol{\mu}}_{Ft}^*, m_F^*)$ , and  $\hat{p}_{Fi}(\hat{\boldsymbol{\mu}}_{Ft}, m_F)$ . While (61) continues to hold under the CRRA preferences,  $\hat{w}_t$  is no longer equal to  $\hat{c}_t$  but is now given by  $\hat{w}_t = \sigma\hat{c}_t + \varphi\hat{n}_t$ . Accordingly, we rewrite (61) as

$$\begin{aligned}\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) &= m_H(\sigma\hat{c}_t + \varphi\hat{n}_t) - \hat{a}_{it} \\ &\quad + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k [m_H(\sigma\Delta\hat{c}_{t+k} + \pi_{t+k} + \varphi\Delta\hat{n}_{t+k}) - \Delta\hat{a}_{it+k}]. \\ &= m_H\sigma\hat{c}_t - (1 - \lambda\delta)\hat{a}_{it} + m_H\varphi \left( \frac{1 - \lambda\delta}{1 - \lambda\delta L^{-1}} \right) \hat{n}_t.\end{aligned}\quad (81)$$

where  $L$  is the lag operator. In the second equality, we used (75) and replaced  $\sigma\Delta\hat{c}_{t+k} + \pi_{t+k}$  by  $\varepsilon_{t+k}^M$ , which greatly simplifies the equation.

Equation (81) differs from (20) in the presence of the forward-looking terms for the labor supply. Equations for  $\hat{p}_{Hi}^*(\hat{\boldsymbol{\mu}}_{Ht}^*, m_H^*)$ ,  $\hat{p}_{Fi}^*(\hat{\boldsymbol{\mu}}_{Ft}^*, m_F^*)$ , and  $\hat{p}_{Fi}(\hat{\boldsymbol{\mu}}_{Ft}, m_F)$  are

$$\hat{p}_{Hi}^*(\hat{\boldsymbol{\mu}}_{Ht}^*, m_H^*) = m_H^*\sigma\hat{c}_t^* - (1 - \lambda\delta)\hat{a}_{it} + m_H^*\varphi \left( \frac{1 - \lambda\delta}{1 - \lambda\delta L^{-1}} \right) \hat{n}_t, \quad (82)$$

$$\hat{p}_{Fi}^*(\hat{\boldsymbol{\mu}}_{Ft}^*, m_F^*) = m_F^*\sigma\hat{c}_t^* - (1 - \lambda\delta)\hat{a}_{it}^* + m_H^*\varphi \left( \frac{1 - \lambda\delta}{1 - \lambda\delta L^{-1}} \right) \hat{n}_t^*, \quad (83)$$

$$\hat{p}_{Fi}(\hat{\boldsymbol{\mu}}_{Ft}, m_F) = m_F\sigma\hat{c}_t - (1 - \lambda\delta)\hat{a}_{it}^* + m_F\varphi \left( \frac{1 - \lambda\delta}{1 - \lambda\delta L^{-1}} \right) \hat{n}_t^*, \quad (84)$$

respectively.

Using (25) and (27),  $\hat{p}_{it}^{opt} - \sigma\hat{c}_t$ , and  $\hat{p}_{it}^{opt*} - \sigma\hat{c}_t^*$  are given by:

$$\begin{aligned}\hat{p}_{it}^{opt} - \sigma\hat{c}_t &= -(1 - m)\sigma\hat{c}_t - (1 - \lambda\delta)[\omega\hat{a}_{it} + (1 - \omega)\hat{a}_{it}^*] \\ &\quad + \varphi \frac{1 - \lambda\delta}{1 - \lambda\delta L^{-1}} [\omega m_H \hat{n}_t + (1 - \omega)m_F \hat{n}_t^*],\end{aligned}\quad (85)$$

$$\begin{aligned}\hat{p}_{it}^{opt*} - \sigma\hat{c}_t^* &= -(1 - m)\sigma\hat{c}_t^* - (1 - \lambda\delta)[\omega\hat{a}_{it}^* + (1 - \omega)\hat{a}_{it}] \\ &\quad + \varphi \frac{1 - \lambda\delta}{1 - \lambda\delta L^{-1}} [\omega m_H \hat{n}_t^* + (1 - \omega)m_F \hat{n}_t],\end{aligned}\quad (86)$$

respectively. In (86), we assumed that  $m_F^* = m_H$  and  $m_H^* = m_F$ .

Plugging (85) and (86) into (80) yields

$$\begin{aligned}\hat{q}_{it} &= \lambda\hat{q}_{it-1} + (1 - \lambda)(1 - m)\hat{q}_t + \lambda\varepsilon_t^n + (1 - \lambda)(1 - \lambda\delta)\psi\varepsilon_{it}^r \\ &\quad - \varphi\psi_m \frac{(1 - \lambda)(1 - \lambda\delta)}{1 - \lambda\delta L^{-1}} (\hat{n}_t - \hat{n}_t^*),\end{aligned}\quad (87)$$

where  $\psi_m = \omega m_H - (1 - \omega) m_F$ . Equation (87) differs from (72) in that there are the forward-looking terms for labor supply. If  $\varphi = 0$ , these forward-looking terms disappear, and the equation coincides with (72). Under our assumptions,  $\sigma$  does not appear in (87). More importantly,  $\sigma$  does not affect the coefficient on the aggregate real exchange rate.

As in the proof of Proposition 2, the symmetry between the two countries implies that  $\bar{q}_i = \bar{q} = 1$ , and thus  $\hat{q}_{it} = \ln q_{it}$  and  $\hat{q}_t = \ln q_t$ . Likewise, the symmetry implies the same steady-state labor supply between the two countries:  $\bar{n} = \bar{n}^*$ , leading to  $\hat{n}_t - \hat{n}_t^* = \ln n_t - \ln n_t^*$ . Substitution of these equations into the above equation leads to

$$\begin{aligned} \ln q_{it} = \lambda \ln q_{it-1} &+ (1 - \lambda)(1 - m) \ln q_t + \lambda \varepsilon_t^n + (1 - \lambda)(1 - \lambda\delta)\psi\varepsilon_{it}^r \\ &- \varphi\psi_m \frac{(1 - \lambda)(1 - \lambda\delta)}{1 - \lambda\delta L^{-1}} (\ln n_t - \ln n_t^*), \end{aligned} \quad (88)$$

which generalizes (29).

To derive the estimation equation for our empirical analysis, we use the definition of  $\tilde{q}_{it}$  and  $\tilde{q}_t$  and further rewrite (88) as

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

or equivalently,

$$\begin{aligned} \ln \tilde{q}_{it} - \lambda\delta\mathbb{E}_t \ln \tilde{q}_{it+1} &= (1 - m)(\ln \tilde{q}_t - \lambda\delta\mathbb{E}_t \ln \tilde{q}_{t+1}) \\ &- (1 - \lambda)(1 - \lambda\delta)\varphi\psi_m(\ln n_t - \ln n_t^*) + (1 - \lambda)(1 - \lambda\delta)\psi\varepsilon_{it}^r, \end{aligned} \quad (89)$$

where  $\mathbb{E}_t \varepsilon_{it+1}^r = 0$ .

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

$$\ln \tilde{\tilde{q}}_{it} = \alpha + \beta \ln \tilde{\tilde{q}}_t + \gamma' X_{it} + u_{it}, \quad (90)$$

where  $X_{it}$  includes the log-difference in labor supply  $\ln n_t - \ln n_t^*$  and  $\gamma$  includes  $-(1 - \lambda)(1 - \lambda\delta)\varphi\psi_m$  as an element. Note that OLS is no longer a valid estimation because  $u_{it}$  now includes forecast error  $\ln \tilde{q}_{it+1} - \mathbb{E}_t \ln \tilde{q}_{it+1}$  and  $\ln \tilde{q}_{t+1} - \mathbb{E}_t \ln \tilde{q}_{t+1}$ . We thus use the instrument for estimation. For the data source of  $\ln n_t - \ln n_t^*$ , we take the indices of total hours worked from *OECD.Stat* with the base year 2010.

Table A.1 reports the estimation results of (90). The left panel presents the results of the US–Canadian city pairs, while the right panel points to the results of the UK–Euro area

city pairs. In both panels, we assume a common  $\lambda$  in specifications (1) and (2) and the good-specific  $\lambda$  in specifications (3) and (4). Specifications (2) and (4) includes the city-pair-specific fixed effects as additional explanatory variable. We instrument  $\ln \tilde{q}_t$  by  $\ln \tilde{q}_{t-1}$  in all specifications. In all cases, the null hypothesis of full attention, namely  $\beta = 0$ , is significantly rejected. The estimated values of  $m$  are much smaller than one, suggesting robustness to changes in the assumption of preferences.

## A.6 Estimation results for (40)

Table A.2 reports the estimation results for (40). Unlike the case of (35), the presence of a lagged dependent variable on the right-hand side of (40) implies a dynamic panel structure. Therefore, dynamic panel regression estimators, such as the generalized method of moments estimator of Arellano and Bond (1991), need to be employed in place of OLS (see, e.g., Crucini and Shintani, 2008 and Crucini et al., 2010a). The left panel of the table presents the estimation results of the US–Canadian city pairs, while the right panel shows those of the UK–Euro area city pairs. In specifications (2) and (4), we impose the restriction that the coefficients on  $\ln q_{it-1}$  and  $\Delta \ln S_t$  as control variables are the same as each other. This is because (29) indicates that  $\ln q_{it-1}$  and  $\varepsilon_t^n = \Delta \ln S_t$  have the same coefficient. Specifications (3) and (4) differ from specifications (1) and (2) in that the regressions include  $\eta_t^r$  as a control variable.

The table indicates that, in all regressions, the null hypothesis that  $\beta = 0$  in (40) is statistically rejected. In addition, the estimates of  $\beta$  are all positive, consistent with the theory. Therefore, even if we directly regress  $\ln q_{ijt}$  on  $\ln q_t$ , the estimation results are consistent with the behavioral inattention of  $0 < m < 1$ .

## A.7 Persistence of the good-level real exchange rate

### A.7.1 Proof of Proposition 3

As a preparation, we rewrite (41) in terms of the log deviation:

$$\hat{q}_t = \rho_q \hat{q}_{t-1} + \rho_q \varepsilon_t^n. \quad (91)$$

The variance of  $\hat{q}_t$  is given by  $\sigma_q^2 = [\rho_q^2 / (1 - \rho_q^2)] \sigma_n^2$ , so

$$\sigma_n^2 = \frac{1 - \rho_q^2}{\rho_q^2} \sigma_q^2. \quad (92)$$

In Appendix A.4, we have shown that the log deviation of the LOP deviations is

$$\hat{q}_{it} = \lambda \hat{q}_{it-1} + \theta \hat{q}_t + \lambda \varepsilon_t^n + \tilde{\psi} \varepsilon_{it}^r, \quad (93)$$

where  $\theta = (1 - \lambda)(1 - m)$  and  $\tilde{\psi} = (1 - \lambda)(1 - \lambda\delta)\psi$ .

Let the covariances denote  $R_0 = \mathbb{E}\hat{q}_t\hat{q}_{it}$  and  $R_1 = \mathbb{E}\hat{q}_t\hat{q}_{it-1}$ . They are written as

$$R_0 = \lambda R_1 + \theta \sigma_q^2 + \lambda \rho_q \sigma_n^2, \quad (94)$$

$$R_1 = \rho_q R_0, \quad (95)$$

which can be derived from (91) and (93).

We further simplify (94) and (95). Substitute (92) and (95) into (94) to get

$$R_0 = \lambda \rho_q R_0 + \left[ \theta + \frac{\lambda(1 - \rho_q^2)}{\rho_q} \right] \sigma_q^2. \quad (96)$$

Note that, using the definition of  $\rho_q$ , the expression inside the brackets can be simplified as<sup>38</sup>

$$\theta + \frac{\lambda(1 - \rho_q^2)}{\rho_q} = 1 - \lambda \rho_q. \quad (97)$$

Using (97), (96) and (95) become

$$R_0 = \sigma_q^2, \quad (98)$$

$$R_1 = \rho_q \sigma_q^2, \quad (99)$$

respectively.

We next work on the variance and the autocovariance that are denoted as  $\sigma_{qi}^2 = \mathbb{E}\hat{q}_{it}^2$  and  $\gamma_1 = \mathbb{E}\hat{q}_{it}\hat{q}_{it-1}$ , respectively. Using (93), (98), and (99), we have

$$\sigma_{qi}^2 = \lambda \gamma_1 + \theta \sigma_q^2 + \lambda \rho_q \sigma_n^2 + \tilde{\psi}^2 \sigma_r^2, \quad (100)$$

$$\gamma_1 = \lambda \sigma_{qi}^2 + \theta \rho_q \sigma_q^2. \quad (101)$$

---

<sup>38</sup>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  $\rho_q$  to this equation again, we obtain (97).

We obtain (100) and (101) from tedious algebra. Regarding (100), we use (93) and (98)

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

Simplifying the last two terms on the right-hand side of the above equation yields (100):

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

The third equality results from (91), and the fourth equality is from the definition of  $\rho_q$ . Regarding (101), use (93) and (99) to get

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

We further simplify  $\sigma_{qi}^2$  and  $\gamma_1$ . Using (92) and (101), (100) becomes

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

Recall that, from (97), the expression inside the brackets is  $1 - \lambda\rho_q$ . This implies,

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

where we use  $1 - \lambda\rho_q(1 - \theta) = 1 - \lambda^2$  given  $\rho_q = \lambda/(1 - \theta)$ . Therefore,  $\sigma_{qi}^2$  and  $\gamma_1$  are

$$\sigma_{qi}^2 = \sigma_q^2 + \frac{\tilde{\psi}^2}{1 - \lambda^2} \sigma_r^2. \quad (102)$$

$$\begin{aligned} \gamma_1 &= \lambda \left[ \sigma_q^2 + \frac{\tilde{\psi}^2}{1 - \lambda^2} \sigma_r^2 \right] + \theta\rho_q\sigma_q^2 \\ &= (\lambda + \theta\rho_q)\sigma_q^2 + \frac{\lambda\tilde{\psi}^2}{1 - \lambda^2} \sigma_r^2 \\ &= \rho_q\sigma_q^2 + \frac{\lambda\tilde{\psi}^2}{1 - \lambda^2} \sigma_r^2. \end{aligned} \quad (103)$$

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

$$\rho_{qi} = \gamma_1/\sigma_{qi}^2 = \omega_\rho\rho_q + (1 - \omega_\rho)\lambda, \quad (104)$$

where  $\omega_\rho$  is defined as

$$\omega_\rho = \frac{\sigma_q^2}{\sigma_q^2 + \frac{\tilde{\psi}^2}{1 - \lambda^2} \sigma_r^2} = \frac{1}{1 + A} \in [0, 1]$$

because

$$A = \frac{\tilde{\psi}^2}{1 - \lambda^2} \frac{\sigma_r^2}{\sigma_q^2} \geq 0. \quad (105)$$

Equation (104) means that  $\rho_{qi}$  is the weighted average of  $\rho_q$  and  $\lambda$ . When we combine Proposition 2, namely  $\rho_q \geq \lambda$ , with (104), it immediately follows that  $\rho_q \geq \rho_{qi} \geq \lambda$ .

### A.7.2 Derivation of (45) and (46)

Using  $\rho_q = \lambda/(1 - \theta)$ , eliminate  $\lambda$  from (104):

$$\rho_{qi} = \omega_\rho\rho_q + (1 - \omega_\rho)(1 - \theta)\rho_q = \rho_q [1 - \theta(1 - \omega_\rho)]. \quad (106)$$

Recall (92) and the definition of  $\tilde{\psi}$ . Then, (105) becomes (46):

$$A = (1 - \lambda)^2(1 - \lambda\delta)^2\psi^2 \frac{1 - \rho_q^2}{\rho_q^2(1 - \lambda^2)} \left( \frac{\sigma_r}{\sigma_n} \right)^2. \quad (107)$$



From  $\omega_\rho = 1/(1+A)$ ,  $1 - \omega_\rho = A/(1+A)$ . In addition, recall that  $\theta = (1-\lambda)(1-m)$ . Therefore, (106) implies

$$\frac{\rho_q}{\rho_{qi}} = \frac{1}{1 - (1-\lambda)(1-m)\frac{A}{1+A}}. \quad (108)$$

Figure 1: Empirical distributions of the good-level real exchange rates

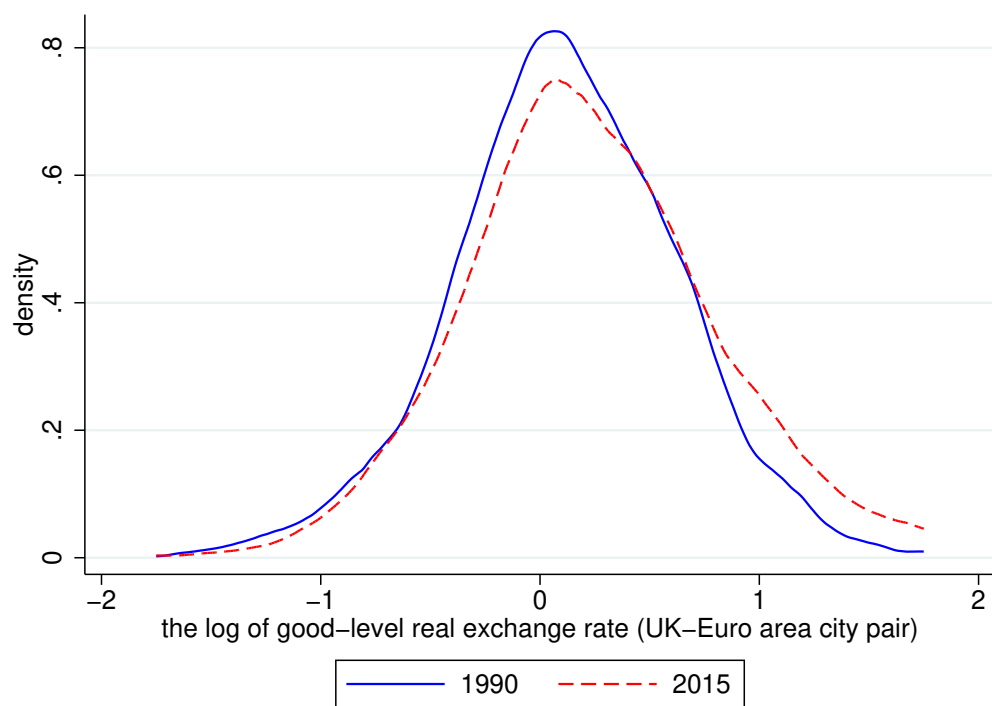
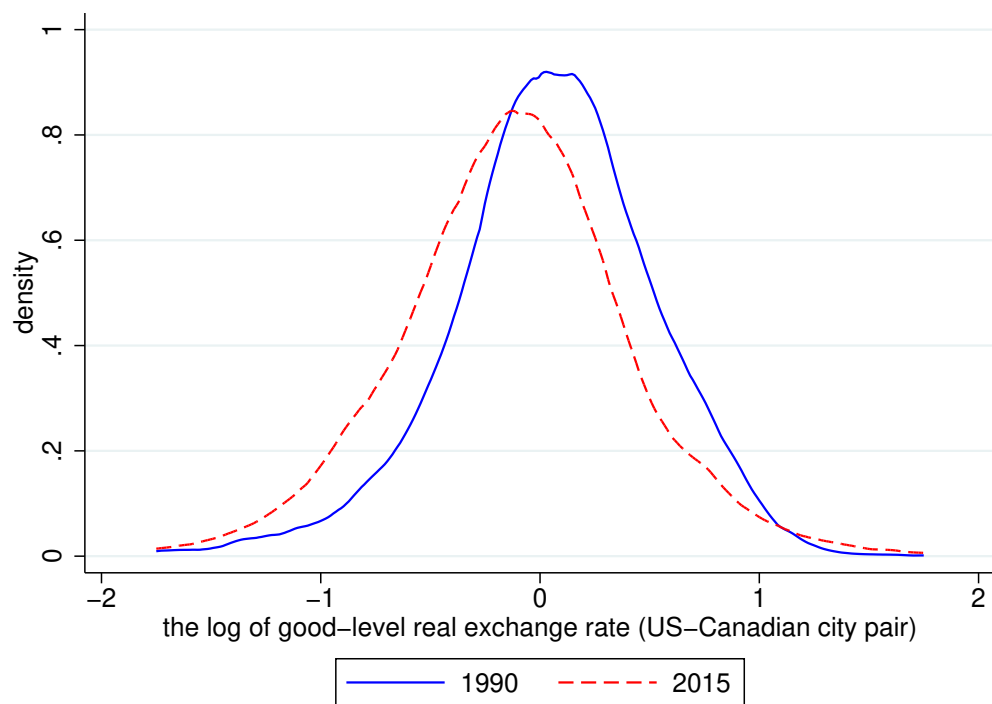


Figure 2: Persistence of the aggregate real exchange rate and the  $\rho_q$  to  $\lambda$  ratio

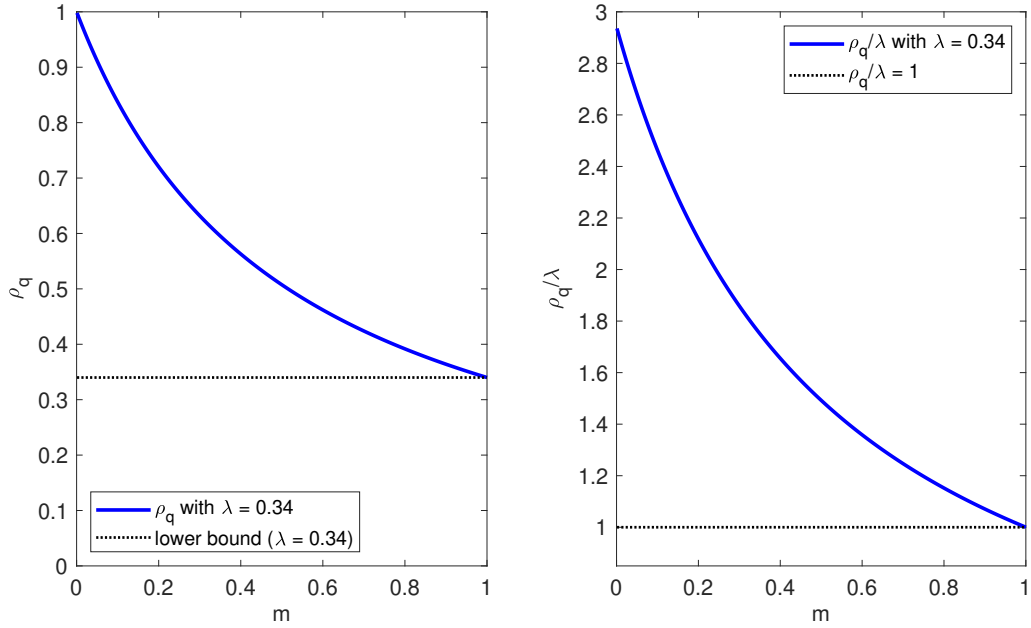


Figure 3: Persistence of the aggregate and the good-level real exchange rates and the  $\rho_q$  to  $\rho_{qi}$  ratio

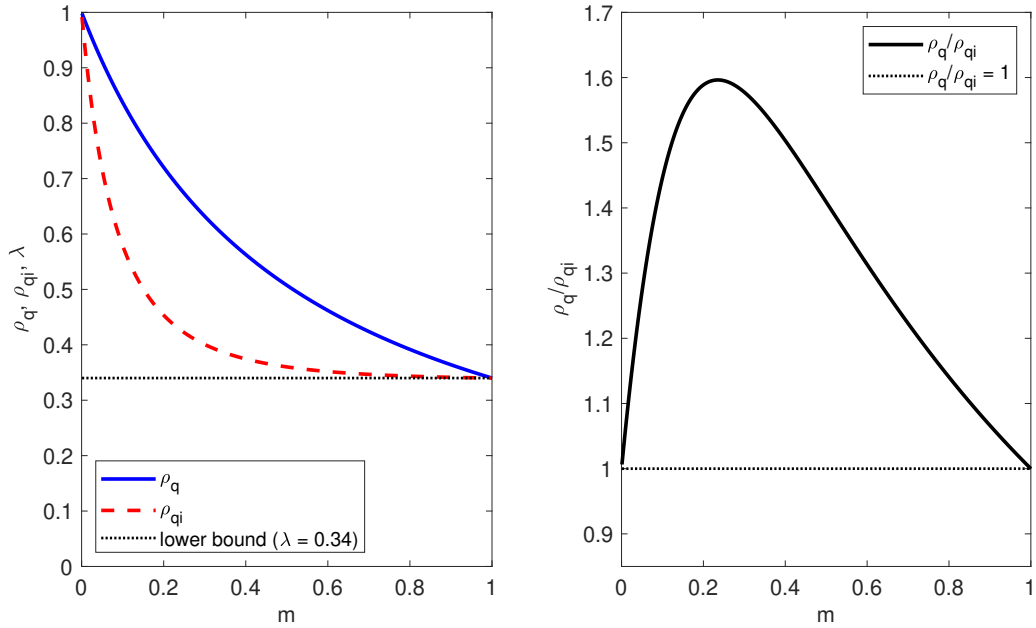


Table 1: Estimation results of (35) under common  $\lambda$

	Dependent variable: $\ln \tilde{q}_{ijt}$ with $\lambda = 0.34$							
	US-Canadian city pairs				UK-Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\ln \tilde{q}_t$	0.844 (0.030)	0.802 (0.028)	0.812 (0.029)	0.806 (0.029)	0.791 (0.0409)	0.749 (0.0420)	0.782 (0.0408)	0.779 (0.0409)
Observations	389,500	389,500	389,500	389,500	225,922	225,922	224,871	224,871
Adj. $R^2$	0.225	0.262	0.225	0.262	0.260	0.319	0.261	0.321
city-pair FE	N	Y	N	Y	N	Y	N	Y
Control for $\eta_t^i$	N	N	Y	Y	N	N	Y	Y
The implied degree of attention from the regression								
$\hat{\eta}$	0.156	0.198	0.188	0.194	0.209	0.251	0.218	0.221

NOTES: 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 38 UK-Euro area city pairs. The data used for regressions covers periods from 1990 to 2015. The table reports the regression coefficients on  $(1 - \lambda) \ln q_t$ . The numbers in parentheses below the coefficients show the standard errors clustered by goods. Each specification includes good-specific fixed effects. Specifications (2) and (4) include city-pair fixed effects. In specifications (3) and (4), we control for the difference in the log of real GDP per hour worked. Regressions for US-Canadian city pairs in the left panel also include dummy variables that control for the difference in timing of the price survey in Calgary from 2003 to 2014. Each column reports the estimate of  $m$  in the bottom row. The “Adj.  $R^2$ ” denotes the adjusted  $R$ -squared.

Table 2: Estimation results of (35) under good-specific  $\lambda$

	Dependent variable: $\ln \tilde{q}_{ijt}$ with good-specific $\lambda$							
	US-Canadian city pairs				UK-Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\ln \tilde{q}_t^i$	0.894 (0.029)	0.862 (0.028)	0.883 (0.032)	0.880 (0.033)	0.838 (0.043)	0.789 (0.044)	0.827 (0.042)	0.803 (0.043)
Observations	389,500	389,500	389,500	389,500	208,073	208,073	207,095	207,095
Adj. $R^2$	0.256	0.294	0.256	0.294	0.274	0.317	0.275	0.317
city-pair FE	N	Y	N	Y	N	Y	N	Y
Control for $\eta_t^i$	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	0.162	0.211	0.173	0.197

NOTES: The dependent variable is  $\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda_i [\ln q_{ijt-1} - \ln(S_t/S_{t-1})]$ , where  $\lambda_i$  is good specific and calibrated based on Nakamura and Steinsson (2008). The table reports the regression coefficients on  $\ln \tilde{q}_t^i = \ln(q_t/q_t^{\lambda_i})$ . See the notes to Table 1 for the remaining details.

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

	Half-lives of the aggregate real exchange rate		
	Predicted half-life	95% CI	Data
Based on $\hat{m}$ from Table 1			
US–Canada ( $\hat{m} = 0.156$ )	2.620	[1.989, 4.010]	4.922
UK–Euro area ( $\hat{m} = 0.209$ )	2.035	[1.555, 3.109]	2.425
Based on $\hat{m}$ from Table 2			
US–Canada ( $\hat{m} = 0.106$ )	3.704	[2.524, 7.605]	4.922
UK–Euro area ( $\hat{m} = 0.162$ )	2.535	[1.774, 4.933]	2.425
	Half-lives of the good-level real exchange rate		
	Predicted half-life	95% CI	Data
Based on $\hat{m}$ from Table 1			
US–Canada ( $\hat{m} = 0.156$ )	0.984	[0.851, 1.292]	1.606
UK–Euro area ( $\hat{m} = 0.209$ )	0.860	[0.766, 1.091]	1.247
Based on $\hat{m}$ from Table 2			
US–Canada ( $\hat{m} = 0.106$ )	1.223	[0.963, 2.110]	1.606
UK–Euro area ( $\hat{m} = 0.162$ )	0.965	[0.808, 1.500]	1.247

NOTES: The table reports the predicted half-lives of the aggregate and good-level real exchange rates. The unit of half-lives is a year and the half-life under full attention is 0.64 years. The upper panel presents the half-lives of the aggregate real exchange rate, and the lower panel shows those of the good-level real exchange rate. To calculate the predicted half-lives in the table, we use the calibrated values of  $\tau = 0.74$ ,  $\varepsilon = 4$ ,  $\sigma_\tau/\sigma_{\Delta s} = 5$ , and  $\delta = 0.98$ . In all calculations,  $\lambda$  is kept constant at  $\lambda = 0.34$ .

In each panel, we report half-lives for US–Canadian city pairs and those for the UK–Euro area city pairs. The second column of the table reports the half-lives predicted by the model with behavioral inattention, and the second column is their 95% confidence intervals denoted by “95% CI.” We compute the half-lives from  $\hat{m}$  and the 95% confidence intervals of  $\hat{m}$  based on specification (1) of Tables 1 and 2. For comparisons, the most-right column presents the half-lives estimated from the data. For the details of the estimation of the half-lives, see the main text.

Table A.1: Estimation results of (90)

	Dependent variable: $\ln \tilde{q}_{ijt}$							
	US-Canadian city pairs				UK-Euro area city pairs			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\lambda$	0.34	0.34	good-specific	good-specific	0.34	0.34	good-specific	good-specific
$\ln \tilde{q}_i$	0.782 (0.034)	0.804 (0.034)	0.853 (0.041)	0.87 (0.041)	0.703 (0.047)	0.642 (0.049)	0.751 (0.053)	0.678 (0.055)
Observations	371,347	371,347	371,347	371,347	212,712	212,712	196,181	196,181
Adj. $R^2$	0.179	0.206	0.242	0.269	0.213	0.257	0.238	0.269
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	0.297	0.358	0.249	0.322

NOTES: In the estimation equation, we assume the CRRRA preferences. The dependent variable is  $\ln \tilde{q}_{ijt} = \ln \tilde{q}_{ijt} - \lambda \delta \ln \tilde{q}_{ijt+1}$ , where  $\lambda$  is calibrated at either common or good-specific  $\lambda$ . In each panel, specifications (1) and (2) report the results using common  $\lambda$  while specifications (3) and (4) report those using good-specific  $\lambda$ . The table reports the regression coefficients on  $\ln \tilde{q}_i = \ln \tilde{q}_i - \lambda \delta \tilde{q}_{i+1}$ . We use the instrument  $\ln \tilde{q}_{t-1}$  to estimate parameters. See the notes to Table 1 for the remaining details.

Table A.2: Estimation results of (40)

	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 (0.021)	0.300 (0.020)	0.264 (0.025)	0.261 (0.022)	0.082 (0.029)	0.112 (0.026)	0.078 (0.029)	0.108 (0.026)
Observations	371,347	371,347	371,347	371,347	213,755	213,755	212,712	212,712
P-value for Hansen's J-test	0.932	0.925	0.922	0.926	0.474	0.498	0.473	0.469
Control for $\eta_t^f$	N	N	Y	Y	N	N	Y	Y

NOTES: The estimation equation is given by (40). Following Arellano and Bond (1991), we regress  $\Delta \ln q_{ijt}$  on the lagged dependent variable  $\Delta \ln q_{ijt-1}$ ,  $\Delta \ln S_t$ ,  $\Delta^2 \ln S_t$  along with the other control variables used in the benchmark regressions. In specifications (2) and (4), we impose the parameter restrictions that the coefficients on  $\ln q_{ijt-1}$  and  $\Delta \ln S_t$  are the same. In specifications (3) and (4), we control for the difference 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 for instruments, depending on the period of observation. "P-value for Hansen's J-test" denotes the p-value for the test of the over-identifying 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 the remaining details.