Discussion Paper No. 1121

ISSN (Print) 0473-453X
ISSN (Online) 2435-0982

A BEHAVIORAL EXPLANATION FOR THE PUZZLING PERSISTENCE OF THE AGGREGATE REAL EXCHANGE RATE

Mario J. Crucini
Mototsugu Shintani
Takayuki Tsuruga

March 2021

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

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

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

This version: March 2021

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 from US and Canadian cities. 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 explained only by sticky prices. 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, Ippei Fujiwara, Yasuo Hirose, Jinill Kim, Akihisa Shibata, Kwanho Shin and conference and seminar participants at the 5th International Macroeconomics and Finance Conference, the North American 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 (15H05728, 15H05729, 18K01684, 20H01482, 20H05631, and 20H05633), 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.

†Department of Economics, Vanderbilt University and NBER; e-mail: `mario.crucini@vanderbilt.edu`.

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

§Institute of Social and Economic Research, Osaka University; and CAMA; e-mail: `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 aggregate real exchange rate (the PPP deviations) is likely to be much more persistent than the good-level real exchange rate (the LOP deviations).[1] 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 LOP and PPP 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.

[1] See Crucini and Shintani (2008) for a comprehensive empirical analysis of the persistence in the LOP deviations.

The reduced-form solution leads to a direct testable implication. Using US and Canadian micro price data, 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 0.17, 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), who 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 with 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 persistence of the aggregate real exchange rate rises to 0.74 from 0.34 in the full attention case.

In the second theoretical result of our model, we explain the gap between the highly persistent PPP deviations and the less persistent LOP deviations. This gap arises from the combination of the complementarity and the presence of idiosyncratic real shocks to the individual price of goods. We show that both the aggregate and the good-level real exchange rates 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 also imply a substantial gap in persistence between the PPP and LOP deviations. In fact, we predict the persistence of the LOP deviations to be 57 percent lower than the persistence of the PPP deviations.

Our results relate 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 rule out these explanations. Instead, we focus on the importance of behavioral inattention in pricing of firms.

A popular explanation for the gap in persistence between aggregate (CPI-based) and good-level real exchange rates 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 multi-sector 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.

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. In what follows, the US and Canada represent 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:

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

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. 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, Π_t. In (1), the aggregate price P_t is given by $P_t = \left[\int P_{it}^{1-\varepsilon} di\right]^{\frac{1}{1-\varepsilon}}$, where P_{it} is the price index for good i. This is a CES aggregate over brands: $P_{it} = \left[\int P_{it}(z)^{1-\varepsilon} dz\right]^{\frac{1}{1-\varepsilon}}$. The left-hand side of (1) represents the nominal value of the total wealth of the household. The household allocates the wealth into state-contingent nominal bond holdings B_{t+1} brought into period $t+1$ and money balances M_t. 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,

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

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 $\Delta_{t,t+1} = \delta[(c_{t+1}/c_t)^{-1}(P_t/P_{t+1})]$, where $\Delta_{t,t+1}$ is the nominal stochastic discount factor. For Canadian households, we have $W_t^*/P_t^* = \chi c_t^*$ and $\Delta_{t,t+1} = \delta[(c_{t+1}^*/c_t^*)^{-1} S_t P_t^*/(S_{t+1}P_{t+1}^*)]$. The consumption Euler equations differ because Canadians buy state-contingent bonds denominated in US dollars.

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

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

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, τ. 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:

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

Similarly, production of good i undertaken in Canada is exhausted between Canadian and US consumption, with US consumption bearing the iceberg trade cost:

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

2.3 Price setting

We introduce the 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,

[2]The condition relies on our preference assumptions. Later, we relax these assumptions in Section 4.

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 λ. 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:

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

where

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

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.[3] 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 $t+k$,

[3] The derivation is provided in Appendix A.1.

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 $p^*_{it}(z) = P^*_{it}/P^*_t$. A fully attentive US firm chooses $\hat{p}^*_{it}(z)$ to maximize[4]

$$v^*_{it}(z) = \mathbb{E}_t \sum_{k=0}^{\infty} \lambda^k \delta_{t,t+k} q_{t+k} \tag{8}$$
$$\times \frac{P^*_t}{P^*_{t+k}} \left\{ \bar{p}^*_i(z) \exp\left[\hat{p}^*_{it}(z)\right] - (1+\tau)\frac{\bar{w}}{\bar{q}} \exp\left(\hat{w}_{t+k} - \hat{q}_{t+k} + \sum_{l=1}^{k} \pi^*_{t+l} - \hat{a}_{it+k} \right) \right\} c^*_{it,t+k}(z),$$

where

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

Here, $\pi^*_t = \ln(P^*_t/P^*_{t-1})$ and $p^*_{it}(z) = P^*_{it}(z)/P^*_t$. 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 τ. 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 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 augment (6) with the degree of attention $m_H \in [0, 1]$ and introduce the "attention-

[4]The derivation is again in Appendix A.1.

augmented" objective function:

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

where $\hat{\boldsymbol{\mu}}_{Ht} = (\hat{\mu}_{Ht}, \hat{\mu}_{Ht+1}, ...)'$ and $\hat{\mu}_{Ht+k} = \hat{w}_{t+k} + \sum_{l=1}^{k} \pi_{t+l}$.[5] 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 $\hat{a}_{it+k}$.[6]

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

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

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 quadratic cost function,

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

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

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

where $\mathbb{E}(\cdot)$ represents the unconditional expectations. In (12), we evaluate $v_{Hi}(\cdot)$ at $\hat{p}_{it}(z) = \hat{p}_{Hi}\left(\hat{\boldsymbol{\mu}}_{Ht}, m_H\right)$ 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

[5] 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.

[6] 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.

inattentive firm's action because m_H in $\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H)$ 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):

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

where

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

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.[7]

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.[8] 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 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, Λ_H for the idiosyncratic productivity would be much larger than Λ_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, κ for their productivity may be much lower than κ for the aggregate shock, such as monetary shocks. Thus, the degree of attention to the idiosyncratic variable is

[7] 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.

[8] 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.

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:

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

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

where ε_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 ε_t^n the nominal shock.

For simplicity, we assume that the log labor productivity also follows a zero-mean i.i.d. process:[9]

$$\ln a_{it} = \varepsilon_{it}, \tag{17}$$

$$\ln a_{it}^* = \varepsilon_{it}^*. \tag{18}$$

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

[9]Later we consider an alternative stochastic process for productivity, but the empirical results from the test of behavioral inattention are unaffected.

$n_t = \int_i \int_{z=0}^{\frac{1}{2}} n_{it}(z)dzdi$ and $n_t^* = \int_i \int_{z=1/2}^{1} n_{it}^*(z)dzdi$.

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 λ affects the extent to which firms place weights on the expected marginal cost. Derivation of the 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:

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

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

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

where

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

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 $q_{it} = S_t P_{it}^{*}/P_{it}$. Using p_{it} and p_{it}^{*}, $\hat{q}_{it}$ is expressed as

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

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:*

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

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

$$m = \omega m_H + (1-\omega) m_F. \tag{30}$$

and $\psi = 2\omega - 1$. *The two random shocks* ε^r_{it} *and* ε^n_t *are given by* $\varepsilon^r_{it} = \varepsilon_{it} - \varepsilon^*_{it} \sim i.i.d.(0, \sigma^2_r)$ *and* $\varepsilon^n_t = \varepsilon^M_t - \varepsilon^{M*}_t \sim i.i.d.(0, \sigma^2_n)$, *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 ε^n_t:

$$\ln q_{it} = \lambda \ln q_{it-1} + \lambda \varepsilon^n_t. \tag{31}$$

This equation is a special case of (29) with $m = 1$ and $\psi = 0$.[10] 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 λ. If $\lambda \to 0$, 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 $\lambda \to 1$, P_{it} never moves, meaning that the good-level real exchange rate keeps track of the nominal exchange rate, which follows a random walk.

Let us compare the stochastic processes for the good-level real exchange rates between the cases $m = 1$ and $0 < m < 1$.[11] 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:

$$\ln q_{it} = \lambda \ln q_{it-1} + (1-m)(1-\lambda) \ln q_t + \lambda \varepsilon^n_t. \tag{32}$$

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

[11] 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.

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[12]

$$\ln q_t = \frac{\lambda}{1-(1-m)(1-\lambda)} \ln q_{t-1} + \frac{\lambda}{1-(1-m)(1-\lambda)} \varepsilon^n_t. \tag{33}$$

Using (33), we can see that the impact multiplier of nominal shocks on the good-level real exchange rate increases from λ 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 ε^r_{it} 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 shocks. Under the fully attentive rational expectations, their model implies:

$$\ln q_{it} = \lambda \ln q_{it-1} + \lambda \varepsilon^n_t + (1-\lambda)(1-\lambda\delta)\psi \varepsilon^r_{it}. \tag{34}$$

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

[12] 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 $\ln q_t = \lambda \ln q_{t-1} + (1-\lambda)(1-m) \ln q_t + \lambda \varepsilon^n_t$. Simplifying the above equation, we obtain (33).

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 $\ln \tilde{q}_{it} = \ln \left[q_{it} / \left(q_{it-1} S_t / S_{t-1} \right)^{\lambda} \right] = \ln q_{it} - \lambda \ln q_{it-1} - \lambda \Delta \ln S_t$ and $\ln \tilde{q}_t = \ln(q_t / q_t^{\lambda}) = (1 - \lambda) \ln q_t$. We estimate the following panel regression model:

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

where α, β, and γ 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

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

where the nominal shock ε_t^n in (29) is replaced by $\Delta \ln S_t$ because the (log) nominal exchange rate follows a random walk with an increment ε_t^n. To implement the regression, we construct $\ln \tilde{q}_{it}$ and $\ln \tilde{q}_t$ using the micro evidence of λ obtained by Nakamura and Steinsson (2008).[13] The error term u_{it} represents an i.i.d. real shock ε_{it}^r and is uncorrelated with the regressor $\ln \tilde{q}_t = (1 - \lambda) \ln q_t$ because ε_{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

[13] Alternatively, we can use the estimates of λ obtained by Bils and Klenow (2004) and Klenow and Kryvtsov (2008). However, our main findings are robust, even if our λ is replaced by their estimates.

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 β.

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:

$$\ln a_{it} = \xi_t + \eta_t + \varepsilon_{it}, \tag{37}$$

$$\ln a^*_{it} = \xi_t + \eta^*_t + \varepsilon^*_{it}. \tag{38}$$

Here, the labor productivity consists of three components: a global component ξ_t, a country-specific component η_t (or η^*_t), and a good-specific component ε_{it} (or ε^*_{it}). In this generalized setting, global and country-specific components follow $\xi_t - \xi_{t-1} = \sum_{j=0}^{\infty} b_j \varepsilon^{\xi}_{t-j}$, $\eta_t = \sum_{j=0}^{\infty} d_j \varepsilon^{\eta}_{t-j}$, and $\eta^*_t = \sum_{j=0}^{\infty} d_j \varepsilon^{\eta^*}_{t-j}$, respectively, where ε^{ξ}_t, ε^{η}_t, and $\varepsilon^{\eta^*}_t$ 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 η_t and η^*_t each follow an AR(1) process with AR coefficient ρ_η and firms are fully attentive to η_t and η^*_t, (36) should be modified to

$$\ln \tilde{q}_{it} = (1 - m) \ln \tilde{q}_t + \frac{(1-\lambda)(1-\lambda\delta)}{1-\lambda\delta\rho_\eta} \psi \eta^r_t + (1-\lambda)(1-\lambda\delta)\psi \varepsilon^r_{it}, \tag{39}$$

where $\eta_t^r = \eta_t - \eta_t^*$. The only difference between (36) and (39) is that the latter includes the new control variable η_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 ε_{it}^r. A smaller coefficient on ε_{it}^r in (29), however, does not change our regression equation (35).

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:

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

where β 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. Unlike the case of (35), however, the presence of a lagged dependent variable in the right-hand side of (40) implies a dynamic panel structure. Therefore, dynamic panel regression estimators, such as the generalized method of moments (GMM) 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).[14]

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. 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.[15] In our analysis, the number of goods and services is 274. Recent studies using these data include Engel and Rogers (2004), Crucini and Shintani (2008), Bergin et al. (2013), Crucini and Yilmazkuday (2014), Andrade and

[14]In Appendix A.5, we report the empirical results based on this alternative strategy.

[15]See Rogers (2007) for details on the comparison between EIU data and the consumer price index data from national statistical agencies.

Zachariadis (2016), and Crucini and Landry (2019).

The data contain prices in multiple cities over 1990–2015. Our analysis focuses on US–Canadian city pairs. We use these city pairs from the two countries. In our data, there are 16 US and 4 Canadian cities.[16] This results in 64 unique cross-country city pairs. However, because some of the US cities have many missing values in the early 90s, our data are an unbalanced panel.[17] Nevertheless, the total number of observations available for our regressions exceeds 350,000.

We compute the log of q_{ijt} for each year ($t = 1990, ..., 2015$), each good ($i = 1, ..., 274$), and each international city pair ($j = 1, ..., 64$). The prices used to construct the good-level real exchange rates are the prices in a US city (expressed in US dollars) and the prices in a Canadian city (expressed in Canadian dollars). We use the spot USD/CAD dollar exchange rates taken from the EIU data to convert prices to common currency units. The EIU records the nominal exchange rate at the end of the week of the price survey. This means that the nominal exchange rate may differ across Canadian cities if the timing of the price survey differs between the two cities. We confirm that the timings of the price survey in Calgary and in the remaining Canadian cities were different from 2003 to 2014. Across other Canadian cities, the USD/CAD exchange rate is common in all periods.[18] 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). For our regression and empirical tests that follow, we augment the micro price data with the aggregate bilateral real exchange rate computed from the official CPI 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 η_t^r, we utilize the difference in real GDP per hour worked between the two countries taken from *OECD.Stat*.

[16] 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.

[17] In particular, the survey only includes the data for Honolulu in 1992 and Lexington and Minneapolis have only been included in the list of cities since 1998.

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

4.3 Estimation results

Table 1 provides the estimation results for (35). In computing $\ln \tilde{q}_{ijt}$ and $\ln \tilde{q}_t$, we calibrate λ at 0.34. This value of λ is obtained by transforming the median monthly frequencies of price changes for the US economy reported in Nakamura and Steinsson (2008) into the *in*frequencies of price changes at an annual rate.[19]

The table reports the estimates of β from regression (35) with the standard errors. To control for unobserved heterogeneity in the LOP deviations, we include two types of fixed effects. One is the good-specific fixed effects and the other is the city-pair-specific fixed effects. Thus, columns (1)–(4) differ from each other in terms of alternative combinations of these fixed effects.[20] In addition, we control for the country-specific component of labor productivity η_t^r in columns (5)–(8) of the table, motivated by (39). In all regressions, we 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.[21]

Overall, $\hat{\beta}$ is around 0.8, and much larger than zero, which is the value under the full attention hypothesis. Because the standard error of the coefficient is around 0.03, the null hypothesis of full attention ($\beta = 0$) is significantly rejected against the alternative hypothesis of partial attention ($\beta > 0$).[22] Our results for the test of behavioral inattention are robust to the presence of fixed effects (see columns (2)–(4)), and to the inclusion of the log-difference in labor productivity as a control variable (see columns (5)–(8)). 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 table also reports the estimated degree of attention $\hat{m}$ $(= 1 - \hat{\beta})$. The point estimates of m range from 0.15 to 0.20.

[19] We transform the monthly median frequency of price changes in Nakamura and Steinsson (2008) into the annual *in*frequency of price changes λ as follows. Let f be the (monthly) median frequency of price changes calculated in Nakamura and Steinsson (2008). 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 $(1 - f)^{12}$. Nakamura and Steinsson (2008) report that the median frequency of price changes in all sectors during 1998–2005 is 8.7 percent. Using the above formula, the annual infrequency of price changes λ is calculated as $\lambda = (1 - f)^{12} = (1 - 0.087)^{12} = 0.34$.

[20] While we do not report the result, we also allow for a fixed effect that is specific to both good i and city pair j. We find that the estimated β is not substantially different.

[21] 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 during 2003–2014. 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 $\ln S_t^k - \ln S_t$ in our samples.

[22] 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.

Table 2 points to the estimation results when the calibrated λ differs across goods. While our model assumes the degree of price stickiness common across goods, the previous empirical studies on sticky prices report the heterogeneity in the frequency of price changes. Thus, we transform each good-specific monthly frequency of price changes reported in Nakamura and Steinsson (2008) into the good-specific infrequency of price changes and construct $\ln \tilde{q}_{ijt} = \ln q_{ijt} - \lambda_i \ln q_{ijt-1} - \lambda_i \Delta \ln S_t$ and $\ln \tilde{q}_t^i = (1 - \lambda_i) \ln q_t$ for good i.[23] Even when we allow for the good-specific degree of price stickiness, the null hypothesis of full attention is again significantly rejected and $\hat{m}$ ranges between 0.11 and 0.24.

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.

Let ρ_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:

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

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,*

$$\rho_q \geq \lambda, \tag{42}$$

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 $m = 1$. ■

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., $m = 1$), ρ_q is equal to λ. However, if firms are inattentive (i.e., $m < 1$), ρ_q becomes strictly greater than

[23] 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.

λ. In the extreme case of $m \to 0$, the aggregate real exchange rate can even follow a random walk, as $\rho_q \to 1$. 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.

The left panel of Figure 2 plots ρ_q against $m \in (0, 1]$. We report the cases of $\lambda = 0.34$ (the solid line) and 0.68 (the dashed line) to assess the impact of changes in λ. Starting from $\rho_q = \lambda$ when $m = 1$, ρ_q increases monotonically as m increases. The persistence becomes closer to unity as m approaches zero. For example, the curve for $\lambda = 0.34$ shows that $\rho_q = 0.34$ when $m = 1$. However, if we use $m = 0.17$, the mean of the estimated degrees of attention in Tables 1 and 2, the same line indicates that $\rho_q = 0.74$.

The right panel of the same figure illustrates the ρ_q to λ ratio, which is defined as:

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

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 ρ_q to λ ratio can be quite large depending on m. When $m = 0.17$, the PPP deviations are more than twice as persistent as what is predicted only by the degree of price stickiness ($\rho_q/\lambda = 2.20$).

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 half-lives and the first-order autocorrelations of $\ln q_t$ between the models with and without full attention.[24] We also report the half-lives and the first-order autocorrelations reported in previous empirical studies in the column headed "Data." In particular, Rogoff (1996) concluded that the half-life

[24]The half-lives are calculated using the formula for the AR(1) process given by $-\ln(2)/\ln \rho_q$.

estimated by previous studies is 3–5 years. This range of half-life is roughly equivalent to 0.79–0.87 in terms of the first-order autocorrelation.

We reconfirm that the model with full attention performs poorly. 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 first-order autocorrelation translates into a very short half-life of 0.64 years. These predictions are inconsistent with the estimated persistence of the aggregate real exchange rates as the half-life of 0.64 years is far below the half-life of 3–5 years in the aggregate real exchange rate.

By contrast, the model with a relatively low degree of attention explains the persistence of the aggregate real exchange rate quite well. For example, when we use the lowest point estimate (specification (2) in Table 2), $m = 0.11$, the model predicts that the half-life of the aggregate real exchange rate is 3.7 years, which falls in the range of 3–5 years.

We next turn to the good-level real exchange rate. We let ρ_{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 behavioral model.

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

$$\rho_q \geq \rho_{qi}, \tag{44}$$

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.6. ■

Proposition 3 explains the stylized fact that good-level real exchange rates are much less persistent than the aggregate real exchange rate. (Imbs et al. 2005; Crucini and Shintani 2008). 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 multisector 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 ρ_q and ρ_{qi}. This point can be further investigated from the ρ_q to ρ_{qi} ratio defined by:

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

where

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

The derivation is in Appendix A.6. Similar to the ρ_q to λ ratio in (43), the ρ_q to ρ_{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: $\rho_q = \rho_{qi} = \lambda$. If firms are inattentive (i.e., $m < 1$), ρ_{qi} is strictly less than ρ_q.

What is necessary for explaining the gap between ρ_q and ρ_{qi} is real friction. More specifically, trade cost (τ) needs to be strictly positive and the elasticity of substitution across brands (ε) needs to be larger than one for the ρ_q to ρ_{qi} ratio (45) to be strictly greater than one. If $\tau = 0$ or $\varepsilon \to 1$, there is no home bias ($\omega = 1/\left(1+(1+\tau)^{1-\varepsilon}\right) = 1/2$) so that $\psi = 2\omega - 1 = 0$. According to (46), either $\tau = 0$ or $\varepsilon \to 1$ makes A zero and thus (45) becomes one. Likewise, σ_r/σ_n, namely the standard deviation ratio of real shocks (ε_{it}^r) to nominal shocks (ε_t^n), in (46) needs to be strictly positive. If the nominal shock fully dominates the real shock such that $\sigma_r/\sigma_n \to 0$, A is again zero, such that the model fails to generate the gap between ρ_q and ρ_{qi}.

To assess the effect of m on the gap between ρ_q and ρ_{qi}, we calibrate the parameters in (45) and (46). For the parameters of real frictions, we set τ to 74 percent from Anderson and van Wincoop (2004) and ε to 4 from Broda and Weinstein (2006).[25,26] Using these values, we obtain the degree of home bias ω of 0.84, which is roughly consistent with the parameter for home bias used in the literature.[27] The resulting calibrated value of ψ 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 δ

[25]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 $0.74(= 1.21 \times 1.44 - 1)$.

[26]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.

[27]For example, Chari et al. (2002) calibrate the degree of home bias as 0.76, while Steinsson (2008) uses 0.94.

is set at 0.98 and the degree of price stickiness λ is set to 0.34.

The left panel of Figure 3 plots ρ_{qi} against $m \in (0, 1]$ in the dashed line. It also includes the curve for ρ_q with $\lambda = 0.34$ taken from the solid line in Figure 2. As suggested by Proposition 3, the curve for ρ_{qi} is always located below the curve for ρ_q. Recall that the lower bound of ρ_q is $\lambda(= 0.34)$ at $m = 1$. This property is preserved for ρ_{qi} because $\rho_q = \rho_{qi} = \lambda$ hold at $m = 1$.

The right panel of Figure 3 indicates that the ρ_q to ρ_{qi} ratio is hump shaped against $m \in (0, 1]$. The ρ_q to ρ_{qi} ratio is one, when $m \to 0$ 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. However, when $0 < m < 1$, the ρ_q to ρ_{qi} ratio exceeds unity. Indeed, when m is set to 0.17, the mean of $\hat{m}$, the ρ_q to ρ_{qi} ratio amounts to 1.57, indicating that the LOP deviations are 57 percent less persistent than the PPP deviations.

Using the above-calibrated parameters and the estimated degree of attention, we assess how much the model can explain the gap in persistence between the aggregate and the good-level real exchange rates. The lower panel of Table 3 presents the predicted half-lives and the first-order autocorrelations of the good-level real exchange rate. In the rightmost column, we report the half-lives and the first-order autocorrelations of the good-level real exchange rates taken from Crucini and Shintani (2008). The observed half-lives range between 1.0 and 1.6 years and the corresponding first-order autocorrelations range between 0.51 and 0.65.

We can see from the table that the model with behavioral inattention ($m < 1$) performs much better than the model with full attention in explaining the half-life of the good-level real exchange rates. In contrast to 0.64 years under $m = 1$, the half-life is 0.93 years under $m = 0.17$. If we set m at 0.11, the half-life is 1.22 years. Thus, the model well explains the gap in persistence between the aggregate and the good-level real exchange rates. In the data, the half-life of the LOP deviations is at most 1.6 years, but that in the PPP deviations is at least 3 years so that there seems to be a gap of around 2 years between them. Based on the estimated values of m, the half-lives predicted from our model are between 0.82 and 1.22 years for the former and between 1.83 and 3.70 years for the latter, suggesting the gap of 1.01 to 2.48 years between them.

Before closing this section, two remarks are in order. First, it is straightforward to combine Propositions 2 and 3 to obtain the ρ_{qi} to λ ratio that measures the amplification from λ to

ρ_{qi}. In particular, using (43) and (45), we have

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

Given that the equality is excluded as long as $A > 0$ and $m < 1$, the persistence of the good-level real exchange rate exceeds λ. We can reconfirm this inequality from the left panel of Figure 3. The dashed line for ρ_{qi} is always located above the dotted line for λ. The result is consistent with Kehoe and Midrigan's (2007) finding that the observed persistence of the good-level real exchange rate often exceeds the degree of price stickiness.

Second, our model assumes quasi-linear preferences $U(c,n) = \ln c - \chi n$, but the estimates of m are robust even when we replace these with the more general constant-relative-risk-aversion (CRRA) form. In this case, firms expect a dynamic path for the labor supply from the time of price setting to the infinite future, and the good-level real exchange rate does not have a simple reduced-form solution. However, as shown in Appendix A.7, we can still conduct a test for behavioral inattention and obtain the degree of attention by employing the instrumental variables estimator under CRRA preferences. Our test rejects the null hypothesis of full attention and the estimated values of degree of attention are very close to the results shown in Tables 1 and 2.

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 adapt the model of behavioral inattention in Gabaix (2014) to a simple two-country, sticky-price model. We show 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 for US and Canadian cities, we implement a test of behavioral inattention and quantify its importance. We find strong evidence consistent with behavioral inattention. Our model produces an aggregate real exchange rate that is more than twice as persistent as the real exchange rate explained only by sticky prices under the estimated degree of attention. With additional calibrated parameters, our model also predicts that the persistence of the LOP deviations is less than two-thirds of the persistence of the PPP deviations.

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

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

[3] 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.

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

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

[6] 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.

[7] 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.

[8] 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.

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

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

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

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

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

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

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

[25] Galí, J. (2015). *Monetary policy, Inflation, and the Business Cycle: an Introduction to the New Keynesian Framework and Its Applications*, Princeton University Press.

[26] 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.

[27] 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.

[28] Kehoe, P. J. and V. Midrigan (2007). "Sticky prices and sectoral real exchange rates," Federal Reserve Bank of Minneapolis, Research Department, Working paper No. 656.

[29] 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.

[30] 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.

[31] 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.

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

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

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

[35] 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.[28] The objective function of US firms that sell their brand in US markets is given by:

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

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:

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

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

Using the definitions of $p_{it}(z)$, w_t, and p_{it}, we rewrite (48) as:

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

Next, 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):

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

where $P_{t+k}/P_t = \prod_{l=1}^{k} P_{t+l}/P_{t+l-1} = \exp\left[\sum_{l=1}^{k} \ln\left(P_{t+l}/P_{t+l-1}\right)\right] = \exp\left[\sum_{l=1}^{k} \pi_{t+l}\right]$. 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):

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

[28] For example, see Galí (2015).

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 objective function of these firms is

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

subject to the demand function by Canadian consumers:

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

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 (48) 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):

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

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

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

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

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

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 $\hat{\boldsymbol{\mu}}_{Ht} = 0$, 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 $\mathbb{E}v_{Hi}[p_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H), \hat{\boldsymbol{\mu}}_{Ht}, 1]$. Here, the profit of the firm is evaluated at $m_H = 1$ (which appears in the last augment in $v_{Hi}(\cdot)$, but the price is distorted by $m_H \neq 1$. We then evaluate 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

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

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 $\partial^2 v_{Hi}^0/\partial \hat{\mu}_{t+k}^2 = \partial^2 v_{Hi}\left(\hat{p}_{Hi}(\mathbf{0}, m_H), \mathbf{0}, 1\right)/\partial \hat{\mu}_{t+k}^2$.

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 $\hat{\mu}_{Ht+k}$ for $k = 0, 1, 2, ...$ and evaluating them at $\hat{\boldsymbol{\mu}}_{Ht} = \mathbf{0}$:

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

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

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

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:

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

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

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

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

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

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. Formally, the choice of attention for US firms that sell their goods in US markets is characterized by:

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

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

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

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

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

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

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

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

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

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

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

Next, the remaining terms are

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

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

(60) becomes

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

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

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

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

$$\hat{q}_t = \hat{c}_t - \hat{c}^*_t. \tag{63}$$

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

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

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

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

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

$$\hat{p}_{it} - \hat{c}_t = \lambda(\hat{p}_{it-1} - \hat{c}_{t-1}) - \lambda\varepsilon^M_t + (1-\lambda)(\hat{p}^{opt}_{it} - \hat{c}_t). \tag{66}$$

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

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

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

$$\hat{q}_{it} = \lambda\hat{q}_{it-1} + \lambda\varepsilon^n_t + (1-\lambda)\left[\left(\hat{p}^{opt*}_{it} - \hat{c}^*_t\right) - \left(\hat{p}^{opt}_{it} - \hat{c}_t\right)\right], \tag{68}$$

where $\varepsilon^n_t = \varepsilon^M_t - \varepsilon^{M^*}_t$.

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

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

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

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

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

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

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

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

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 Estimation results for (40)

Table A.1 reports the estimation results for (40) by GMM. We report the results of four specifications. 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 η_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 β are all positive, consistent with the theory. Therefore, even if we directly regress $\ln q_{ijt}$ on $\ln q_t$, the data are consistent with the partial attention of $0 < m < 1$.

A.6 Persistence of the good-level real exchange rate

A.6.1 Proof of Proposition 3

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

$$\hat{q}_t = \rho_q \hat{q}_{t-1} + \rho_q \varepsilon_t^n. \tag{73}$$

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

$$\sigma_n^2 = \frac{1-\rho_q^2}{\rho_q^2}\sigma_q^2. \tag{74}$$

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

$$\hat{q}_{it} = \lambda \hat{q}_{it-1} + \theta \hat{q}_t + \lambda \varepsilon_t^n + \tilde{\psi}\varepsilon_{it}^r, \tag{75}$$

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

$$R_0 = \lambda R_1 + \theta\sigma_q^2 + \lambda\rho_q\sigma_n^2, \tag{76}$$

$$R_1 = \rho_q R_0, \tag{77}$$

which can be derived from (73) and (75).

We further simplify (76) and (77). Substitute (74) and (77) into (76) to get

$$R_0 = \lambda\rho_q R_0 + \left[\theta + \frac{\lambda(1-\rho_q^2)}{\rho_q}\right]\sigma_q^2. \tag{78}$$

Note that, using the definition of ρ_q, the expression inside the brackets can be simplified as[29]

$$\theta + \frac{\lambda(1-\rho_q^2)}{\rho_q} = 1 - \lambda\rho_q. \tag{79}$$

[29] To see this, $\theta + \lambda(1-\rho_q^2)/\rho_q = \theta + \lambda(1-\rho_q^2)/(\lambda/(1-\theta)) = \theta + (1-\rho_q^2)(1-\theta) = 1 - \rho_q^2(1-\theta)$. Applying the definition of ρ_q to this equation again, we obtain (79).

Using (79), (78) and (77) become

$$R_0 = \sigma_q^2, \tag{80}$$

$$R_1 = \rho_q \sigma_q^2, \tag{81}$$

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 (75), (80), and (81), we have

$$\sigma_{qi}^2 = \lambda\gamma_1 + \theta\sigma_q^2 + \lambda\rho_q\sigma_n^2 + \tilde{\psi}^2\sigma_r^2, \tag{82}$$

$$\gamma_1 = \lambda\sigma_{qi}^2 + \theta\rho_q\sigma_q^2. \tag{83}$$

We obtain (82) and (83) from tedious algebra. Regarding (82), we use (75) and (80)

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

Simplifying the last two terms on the right-hand side of the above equation yields (82):

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

The third equality results from (73), and the fourth equality is from the definition of ρ_q. Regarding (83), use (75) and (81) to get

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

We further simplify σ_{qi}^2 and γ_1. Using (74) and (83), (82) becomes

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

Recall that, from (79), 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, σ_{qi}^2 and γ_1 are

$$\sigma_{qi}^2 = \sigma_q^2 + \frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2. \tag{84}$$

$$\begin{aligned}\gamma_1 &= \lambda\left[\sigma_q^2 + \frac{\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2\right] + \theta\rho_q\sigma_q^2 \\ &= (\lambda+\theta\rho_q)\sigma_q^2 + \frac{\lambda\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2 \\ &= \rho_q\sigma_q^2 + \frac{\lambda\tilde{\psi}^2}{1-\lambda^2}\sigma_r^2.\end{aligned} \tag{85}$$

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

$$\rho_{qi} = \gamma_1/\sigma_{qi}^2 = \omega_\rho\rho_q + (1-\omega_\rho)\,\lambda, \tag{86}$$

where ω_ρ 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

$$A = \frac{\tilde{\psi}^2}{1-\lambda^2}\frac{\sigma_r^2}{\sigma_q^2} \geq 0. \tag{87}$$

Equation (86) means that ρ_{qi} is the weighted average of ρ_q and λ. When we combine Proposition 2, namely $\rho_q \geq \lambda$, with (86), it immediately follows that $\rho_q \geq \rho_{qi} \geq \lambda$.

A.6.2 Derivation of (45) and (46)

Using $\rho_q = \lambda/(1-\theta)$, eliminate λ from (86):

$$\rho_{qi} = \omega_\rho \rho_q + (1-\omega_\rho)(1-\theta)\rho_q = \rho_q \left[1 - \theta(1-\omega_\rho)\right]. \tag{88}$$

Recall (74) and the definition of $\tilde{\psi}$. Then, (87) becomes (46):

$$A = (1-\lambda)^2 (1-\lambda\delta)^2 \psi^2 \frac{1-\rho_q^2}{\rho_q^2 (1-\lambda^2)} \left(\frac{\sigma_r}{\sigma_n}\right)^2. \tag{89}$$

From $\omega_\rho = 1/(1+A)$, $1-\omega_\rho = A/(1+A)$. In addition, recall that $\theta = (1-\lambda)(1-m)$. Therefore, (88) implies

$$\frac{\rho_q}{\rho_{qi}} = \frac{1}{1-(1-\lambda)(1-m)\frac{A}{1+A}}. \tag{90}$$

A.7 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.[30] Because this is inconsistent with the exchange-rate disconnect puzzle, we replace this assumption by the new assumption on the money growth rate:

$$\Delta \ln M_t = \frac{\sigma-1}{\sigma}\pi_t + \frac{1}{\sigma}\varepsilon_t^M, \tag{91}$$

$$\Delta \ln M_t^* = \frac{\sigma-1}{\sigma}\pi_t^* + \frac{1}{\sigma}\varepsilon_t^{M^*}. \tag{92}$$

Under (91) and (92), the nominal exchange rate continues to follow a random walk.[31]

[30] 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 Δs_t.

[31] 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 (91) and (92) into the above equation yields $\Delta s_t = (\sigma-1)(\pi_t - \pi_t^*) + (1-\sigma)(\pi_t - \pi_t^*) + \varepsilon_t^M - \varepsilon_t^{M^*} = \varepsilon_t^n$.

Using the CIA constraints, we can rewrite (91) and (92) as:

$$\sigma \Delta \hat{c}_{t+k} + \pi_{t+k} = \varepsilon^{M}_{t+k}, \tag{93}$$

$$\sigma \Delta \hat{c}^{*}_{t+k} + \pi^{*}_{t+k} = \varepsilon^{M^*}_{t+k}, \tag{94}$$

for $k > 0$. Later, we utilize (93) and (94) for deriving the estimation equation.

A.7.1 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

$$\hat{q}_{it} = (\hat{p}^{*}_{it} - \sigma \hat{c}^{*}_{t}) - (\hat{p}_{it} - \sigma \hat{c}_t). \tag{95}$$

We focus on $\hat{p}_{it} - \sigma\hat{c}_t$ and $\hat{p}^*_t - \sigma\hat{c}^*_t$ and obtain the expression for $\hat{q}_{it}$ using (95). 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

$$\hat{p}_{it} - \sigma \hat{c}_t = \lambda \left(\hat{p}_{it-1} - \sigma \hat{c}_{t-1}\right) - \lambda \varepsilon^{M}_{t} + (1-\lambda)\left(\hat{p}^{opt}_{it} - \sigma \hat{c}_t\right), \tag{96}$$

where we replace $\sigma \Delta \hat{c}_t + \pi_t$ by ε^M_t using (93). Analogously, (26) remains valid under the CRRA preferences. Using (26), we have

$$\hat{p}^{*}_{it} - \sigma \hat{c}^{*}_{t} = \lambda(\hat{p}^{*}_{it-1} - \sigma \hat{c}^{*}_{t-1}) - \lambda \varepsilon^{M^*}_{t} + (1-\lambda)(\hat{p}^{opt*}_{it} - \hat{c}^{*}_{t}). \tag{97}$$

Therefore, the good-level real exchange rate is

$$\hat{q}_{it} = \lambda \hat{q}_{it-1} + \lambda \varepsilon^{n}_{t} + (1-\lambda)[(\hat{p}^{opt*}_{it} - \sigma \hat{c}^{*}_{t}) - (\hat{p}^{opt}_{it} - \sigma \hat{c}_t)]. \tag{98}$$

Equations (96)–(98) generalize (66)–(68), respectively.

We next focus on the expression inside the brackets on the right-hand side of (98). For the case of $\sigma \neq 1$, we recalculate the log optimal prices: $\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H)$, $\hat{p}^*_{Hi}(\hat{\boldsymbol{\mu}}^*_{Ht}, m^*_H)$, $\hat{p}^*_{Fi}(\hat{\boldsymbol{\mu}}^*_{Ft}, m^*_F)$, 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

$$\begin{aligned}\hat{p}_{Hi}(\hat{\boldsymbol{\mu}}_{Ht}, m_H) &= m_H(\sigma\hat{c}_t + \varphi\hat{n}_t) - \hat{a}_{it} \\ &\quad + \mathbb{E}_t \sum_{k=1}^{\infty} (\lambda\delta)^k \left[m_H \left(\sigma\Delta\hat{c}_{t+k} + \pi_{t+k} + \varphi\Delta\hat{n}_{t+k} \right) - \Delta\hat{a}_{it+k} \right]. \\ &= m_H\sigma\hat{c}_t - (1-\lambda\delta)\hat{a}_{it} + m_H\varphi \left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}} \right) \hat{n}_t. \end{aligned} \tag{99}$$

where L is the lag operator. In the second equality, we used (93) and replaced $\sigma\Delta\hat{c}_{t+k} + \pi_{t+k}$ by ε^M_{t+k}, which greatly simplifies the equation.

Equation (99) 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

$$\hat{p}^*_{Hi}(\hat{\boldsymbol{\mu}}^*_{Ht}, m^*_H) = m^*_H\sigma\hat{c}^*_t - (1-\lambda\delta)\hat{a}_{it} + m^*_H\varphi \left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}} \right) \hat{n}_t, \tag{100}$$

$$\hat{p}^*_{Fi}(\hat{\boldsymbol{\mu}}^*_{Ft}, m^*_F) = m^*_F\sigma\hat{c}^*_t - (1-\lambda\delta)\hat{a}^*_{it} + m^*_H\varphi \left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}} \right) \hat{n}^*_t, \tag{101}$$

$$\hat{p}_{Fi}(\hat{\boldsymbol{\mu}}_{Ft}, m_F) = m_F\sigma\hat{c}_t - (1-\lambda\delta)\hat{a}^*_{it} + m_F\varphi \left(\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}} \right) \hat{n}^*_t, \tag{102}$$

respectively.

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

$$\begin{aligned}\hat{p}^{opt}_{it} - \sigma\hat{c}_t = -(1-m)\sigma\hat{c}_t &- (1-\lambda\delta)[\omega\hat{a}_{it} + (1-\omega)\hat{a}^*_{it}] \\ &+ \varphi\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}} \left[\omega m_H\hat{n}_t + (1-\omega)m_F\hat{n}^*_t \right], \end{aligned} \tag{103}$$

$$\begin{aligned}\hat{p}^{opt*}_{it} - \sigma\hat{c}^*_t = -(1-m)\sigma\hat{c}^*_t &- (1-\lambda\delta)\left[\omega\hat{a}^*_{it} + (1-\omega)\hat{a}_{it}\right] \\ &+ \varphi\frac{1-\lambda\delta}{1-\lambda\delta L^{-1}} \left[\omega m_H\hat{n}^*_t + (1-\omega)m_F\hat{n}_t \right], \end{aligned} \tag{104}$$

respectively. In (104), we assumed that $m^*_F = m_H$ and $m^*_H = m_F$.

Plugging (103) and (104) into (98) yields

$$\begin{aligned}\hat{q}_{it} = \lambda\hat{q}_{it-1} + (1-\lambda)(1-m)\hat{q}_t + \lambda\varepsilon^n_t + (1-\lambda)(1-\lambda\delta)\psi\varepsilon^r_{it} \\ - \varphi\psi_m\frac{(1-\lambda)(1-\lambda\delta)}{1-\lambda\delta L^{-1}}(\hat{n}_t - \hat{n}^*_t), \end{aligned} \tag{105}$$

where $\psi_m = \omega m_H - (1-\omega)\, m_F$. Equation (105) 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, σ does not appear in (105). More importantly, σ 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

$$\begin{aligned}\ln q_{it} = \lambda \ln q_{it-1} \quad + & \quad (1-\lambda)(1-m)\ln q_t + \lambda \varepsilon^n_t + (1-\lambda)(1-\lambda\delta)\psi\varepsilon^r_{it} \\ & -\varphi\psi_m \frac{(1-\lambda)(1-\lambda\delta)}{1-\lambda\delta L^{-1}}\left(\ln n_t - \ln n^*_t\right),\end{aligned} \tag{106}$$

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 (106) as

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

or equivalently,

$$\begin{aligned}\ln \tilde{q}_{it} - \lambda\delta \mathbb{E}_t \ln \tilde{q}_{it+1} \quad = & \quad (1-m)(\ln \tilde{q}_t - \lambda\delta \mathbb{E}_t \ln \tilde{q}_{t+1}) \\ & -(1-\lambda)(1-\lambda\delta)\varphi\psi_m(\ln n_t - \ln n^*_t) + (1-\lambda)(1-\lambda\delta)\psi\varepsilon^r_{it},\end{aligned} \tag{107}$$

where $\mathbb{E}_t\varepsilon^r_{it+1} = 0$.

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

$$\ln \tilde{\tilde{q}}_{it} = \alpha + \beta \ln \tilde{\tilde{q}}_t + \gamma' X_{it} + u_{it}, \tag{108}$$

where X_{it} includes the log-difference in labor supply $\ln n_t - \ln n^*_t$ and γ 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.2 reports the estimation results based on (108). We use a common λ in specifications (1) and (2) and the good-specific λ in specifications (3) and (4). In addition, we instrument $\ln \tilde{\tilde{q}}_t$ by $\ln \tilde{q}_{t-1}$ in specifications (1) and (3) and $\ln \tilde{\tilde{q}}_{t-1}$ in specifications (2) and

(4). In all cases, the null hypothesis of full attention, namely $\beta = 0$, is significantly rejected. The estimated values of m are also very close to each other, suggesting robustness to changes in the assumption of preferences.

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

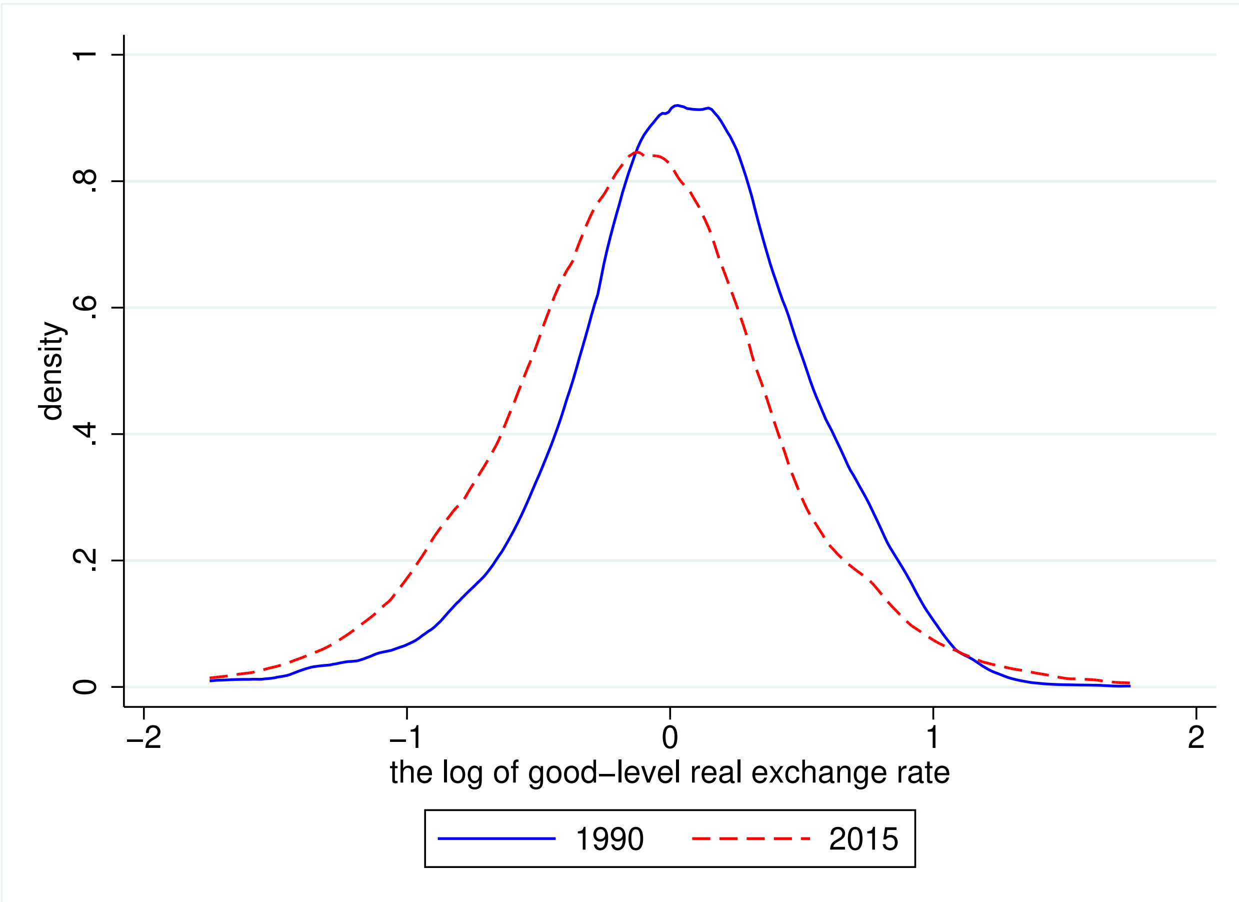

Figure 2: Persistence of the aggregate real exchange rate and the ρ_q to λ ratio

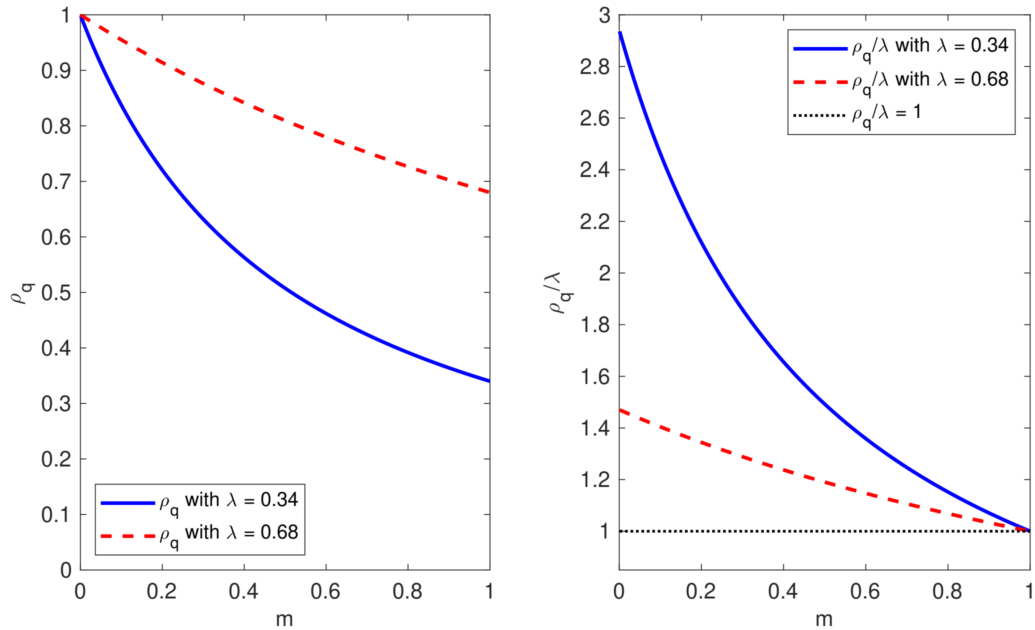

Figure 3: Persistence of the aggregate and the good-level real exchange rates and the ρ_q to ρ_{qi} ratio

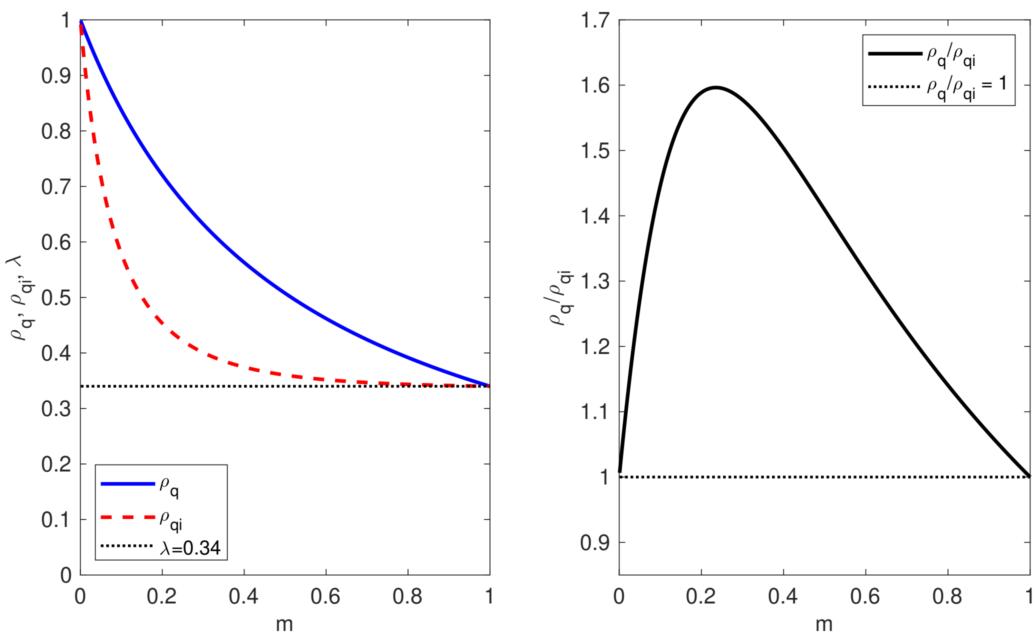

Table 1: Regressions with common λ

	Dependent variable: $\ln \tilde{q}_{ijt}$ with $\lambda = 0.34$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln \tilde{q}_t$	0.843	0.844	0.806	0.802	0.848	0.812	0.844	0.806
	(0.030)	(0.030)	(0.028)	(0.028)	(0.031)	(0.029)	(0.031)	(0.029)
Observations	389,500	389,500	389,500	389,500	389,500	389,500	389,500	389,500
Adj. R^2	0.049	0.225	0.087	0.262	0.049	0.225	0.087	0.262
good FE	N	Y	N	Y	N	Y	N	Y
city-pair FE	N	N	Y	Y	N	N	Y	Y
Control for η_t^r	N	N	N	N	Y	Y	Y	Y
The implied degree of attention from the regression								
$\hat{m}$	0.157	0.156	0.194	0.198	0.152	0.188	0.156	0.194

NOTES: Regressions estimated using panel data for 274 items and 64 US and Canadian city-pairs over 26 years. Standard errors are clustered by goods in parentheses below the coefficients. The dependent variable is $\ln q_{ijt} - \lambda(\ln q_{ijt-1} - \Delta \ln S_t)$, where λ is calibrated at 0.34. The table reports the regression coefficients on $(1-\lambda)\ln q_t$. Each specification includes dummy variables that control for the difference in timing of the price survey in Calgary over 2003–2014. In columns (5)–(8), we control for the difference in the log of real GDP per hour worked. Each column reports the estimate of m along with the p-value for the null of $m = 1$. The "Adj. R^2" denotes the adjusted R-squared.

Table 2: Regressions with good-specific λ

Dependent variable: $\ln \tilde{q}_{ijt}$ with good-specific λ								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln \tilde{q}_t$	0.793	0.894	0.766	0.862	0.765	0.883	0.763	0.88
	(0.045)	(0.029)	(0.045)	(0.028)	(0.053)	(0.032)	(0.053)	(0.033)
Observations	389,500	389,500	389,500	389,500	389,500	389,500	389,500	389,500
Adj. R^2	0.062	0.256	0.100	0.294	0.062	0.256	0.100	0.294
good FE	N	Y	N	Y	N	Y	N	Y
city-pair FE	N	N	Y	Y	N	N	Y	Y
Control for η_t^r	N	N	N	N	Y	Y	Y	Y
The implied degree of attention from the regression								
m	0.207	0.106	0.234	0.138	0.235	0.117	0.237	0.120

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 λ_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: Persistence implied by the estimated degree of attention

	Model with full attention	Model with behavioral inattention $m = 0.237$	$m = 0.174$	$m = 0.106$	Data
The aggregate real exchange rate					
HL	0.643	1.831	2.382	3.704	3-5
ρ_q	0.340	0.685	0.748	0.829	0.79-0.87
The good-level real exchange rate					
HL	0.643	0.819	0.933	1.223	1.0-1.6
ρ_{qi}	0.340	0.429	0.476	0.567	0.51-0.65

NOTES: The table reports the half-lives (denoted "HL") and the first-order autocorrelations (denoted ρ_q and ρ_{qi}) of the real exchange rates predicted by the model. The unit of half-lives is a year. To calculate the predicted values in the table, we use the calibrated values of $\lambda = 0.34$, $\tau = 0.74$, $\varepsilon = 4$, $\sigma_r/\sigma_{\Delta s} = 5$, and $\delta = 0.98$. The computed persistence measures are based on $\hat{m} = 0.237$ (the largest estimate in Tables 1 and 2), $\hat{m} = 0.174$ (the mean of the estimates in Tables 1 and 2), and $\hat{m} = 0.106$ (the smallest estimate in Tables 1 and 2), together with the model with full attention ($m = 1$) for reference. The rightmost column reports estimates from previous studies for comparison. For the aggregate real exchange rate, we take the range of the half-lives from Rogoff (1996) and compute the implied first-order autocorrelations from the half-lives. For the good-level real exchange rate, we take the median estimate of the half-lives and the first-order autocorrelations of all goods provided in Table 4 of Crucini and Shintani (2008).

Table A.1: Estimation results from the GMM

Dependent variable: $\ln q_{ijt}$				
	(1)	(2)	(3)	(4)
$\ln q_t$	0.302	0.300	0.264	0.261
	(0.021)	(0.020)	(0.025)	(0.022)
Observations	371,347	371,347	371,347	371,347
P-value for Hansen's J-test	0.932	0.925	0.922	0.926
Control for η_t^r	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 $\Delta \ln q_{ijt-1}$, $\Delta \ln q_t$, $\Delta^2 \ln S_t$ along with the 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. The test for the over-identifying restrictions is also reported in which 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.

Table A.2: Estimation results under CRRA preferences

Dependent variable: $\ln \tilde{\tilde{q}}_{ijt}$				
	(1)	(2)	(3)	(4)
λ	0.34	0.34	good-specific	good-specific
$\ln \tilde{\tilde{q}}_t$	0.804	0.800	0.870	0.861
	(0.034)	(0.037)	(0.041)	(0.045)
Observations	371,347	371,347	371,347	371,347
good FE	Y	Y	Y	Y
city-pair FE	Y	Y	Y	Y
Instruments	$\ln \tilde{q}_{t-1}$	$\ln \tilde{\tilde{q}}_{t-1}$	$\ln \tilde{q}_{t-1}$	$\ln \tilde{\tilde{q}}_{t-1}$
The implied degree of attention from the regression				
m	0.196	0.200	0.130	0.139

NOTES: The dependent variable is $\ln \tilde{\tilde{q}}_{ijt} = \ln \tilde{q}_{ijt} - \lambda\delta \ln \tilde{q}_{ijt+1}$, where λ is calibrated at the values specified in the first row of the table. The table reports the regression coefficients on $\ln \tilde{\tilde{q}}_t = \ln \tilde{q}_t - \lambda\delta\tilde{q}_{t+1}$ and the difference in the log of hours worked. In specifications (1) and (3), the instrument is $\ln \tilde{q}_{t-1}$. In specifications (2) and (4), the instrument is $\ln \tilde{\tilde{q}}_{t-1}$. See the notes to Table 1 for the remaining details.