

# Notes on Selected Papers in International Macroeconomics

Mikkel Plagborg-Møller\*  
Harvard University  
plagborg@fas.harvard.edu

February 22, 2014

## DISCLAIMER:

These notes were written in preparation for a second-year PhD exam. They are only meant as rough summaries and can't substitute for actually reading the papers. I would be very happy to correct any errors and misinterpretations as well as entirely removing references to papers. Please see the appendix for a list of abbreviations.

## Contents

<b>1</b>	<b>Real Business Cycles in Open Economies</b>	<b>4</b>
1.1	Aguiar and Gopinath (2007): “Emerging Market Business Cycles” . . . . .	4
1.2	Backus, Kehoe and Kydland (1992): “International Real Business Cycles” . . . . .	5
<b>2</b>	<b>International Risk Sharing and Co-Movements</b>	<b>6</b>
2.1	Coeurdacier and Rey (2011): “Home bias in open economy macroeconomics” . . . . .	6
2.2	Corsetti, Dedola and Leduc (2008): “International Risk Sharing and the Transmission of Productivity Shocks” . . . . .	7
2.3	Heathcote and Perri (2009): “The international diversification puzzle is not as bad as you think” . . . . .	8
2.4	Kehoe and Perri (2002): “International business cycles with endogenous incomplete markets” . . . . .	10
<b>3</b>	<b>Global Imbalances</b>	<b>10</b>
3.1	Caballero, Farhi and Gourinchas (2008): “An Equilibrium Model of ‘Global Imbalances’ and Low Interest Rates” . . . . .	10
3.2	Mendoza, Quadrini and Ríos-Rull (2009): “Financial Integration, Financial Development, and Global Imbalances” . . . . .	12

---

\*I thank Gita Gopinath for helpful explanations and for devising the reading list. I am solely responsible for any errors.

<b>4</b>	<b>Valuation Effects</b>	<b>14</b>
4.1	Curcucu, Dvorak and Warnock (2008): “Cross-Border Return Differentials” . . . . .	14
4.2	Gourinchas and Rey (2007): “International Financial Adjustment” . . . . .	15
<b>5</b>	<b>Financial Crises, International Recessions</b>	<b>16</b>
5.1	Mendoza (2010): “Sudden Stops, Financial Crises, and Leverage” . . . . .	16
5.2	Neumeyer and Perri (2005): “Business Cycles in Open Economics: The Role of Interest Rates” . . . . .	18
5.3	Perri and Quadrini (2011): “International Recessions” . . . . .	19
<b>6</b>	<b>Capital Controls</b>	<b>21</b>
6.1	Bianchi (2011): “Overborrowing and Systemic Externalities in the Business Cycle” .	21
6.2	Korinek (2011): “The New Economics of Prudential Capital Controls: A Research Agenda” . . . . .	22
<b>7</b>	<b>Sovereign Debt, Debt Crises</b>	<b>24</b>
7.1	Aguiar and Gopinath (2006): “Defaultable debt, interest rates and the current account”	24
7.2	Broner, Martin and Ventura (2010): “Sovereign Risk and Secondary Markets” . . . .	27
7.3	Bulow and Rogoff (1989): “Sovereign Debt: Is to Forgive to Forget?” . . . . .	29
7.4	Calvo (1988): “Servicing the Public Debt: The Role of Expectations” . . . . .	30
7.5	Cole and Kehoe (2000): “Self-Fulfilling Debt Crises” . . . . .	31
7.6	Eaton and Gersovitz (1981): “Debt with Potential Repudiation” . . . . .	33
7.7	Reinhart and Rogoff (2011): “From Financial Crisis to Debt Crash” . . . . .	34
7.8	Yue (2010): “Sovereign default and debt renegotiation” . . . . .	35
<b>8</b>	<b>Real Exchange Rates and Purchasing Power Parity</b>	<b>37</b>
8.1	Alessandria, Kaboski and Midrigan (2010): “Inventories, Lumpy Trade and Large Devaluations” . . . . .	37
8.2	Burstein, Eichenbaum and Rebelo (2005): “Large Devaluations and the Real Exchange Rate” . . . . .	40
8.3	Carvalho and Nechio (2011): “Aggregation and the PPP Puzzle in a Sticky-Price Model” . . . . .	42
8.4	Engel (1999): “Accounting for U.S. Real Exchange Rate Changes” . . . . .	45
8.5	Gopinath, Gourinchas, Hsieh and Li (2009): “Estimating the Border Effect: Some New Evidence” . . . . .	47
8.6	Imbs, Mumtaz, Ravn and Rey (2005): “PPP Strikes Back: Aggregation and the Real Exchange Rate” . . . . .	50
<b>9</b>	<b>LCP vs PCP</b>	<b>51</b>
9.1	Atkeson and Burstein (2008): “Pricing-to-Market, Trade Costs, and International Relative Prices” . . . . .	51
9.2	Gopinath and Itskhoki (2010): “Frequency of Price Adjustment and Pass-Through”	54
9.3	Gopinath, Itskhoki and Rigobon (2010): “Currency Choice and Exchange Rate Pass-Through” . . . . .	57
9.4	Gopinath and Rigobon (2008): “Sticky Borders” . . . . .	60

<b>10 Open Economy Models with Nominal Rigidity</b>	<b>62</b>
10.1 Chari, Kehoe and McGrattan (2002): “Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?”	62
10.2 Devereux and Engel (2003): “Monetary Policy in the Open Economy Revisited: Price Setting and Exchange-Rate Flexibility”	65
10.3 Engel (2011): “Currency Misalignments and Optimal Monetary Policy: A Reexamination”	68
10.4 Farhi, Gopinath and Itskhoki (2011): “Fiscal Devaluations”	70
10.5 Galí and Monacelli (2005): “Monetary Policy and Exchange Rate Volatility in a Small Open Economy”	75
<b>11 Currency Unions</b>	<b>78</b>
11.1 Chari and Kehoe (2007): “On the need for fiscal constraints in a monetary union”	78
11.2 Galí and Monacelli (2008): “Optimal monetary and fiscal policy in a currency union”	80
11.3 Mundell (1961): “A Theory of Optimum Currency Areas”	83
11.4 Santos Silva and Tenreyro (2010): “Currency Unions in Prospect and Retrospect”	84
<b>12 Exchange Rate Disconnect</b>	<b>86</b>
12.1 Bachetta and van Wincoop (2006): “Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?”	86
12.2 Engel, Mark and West (2008): “Exchange Rate Models Are Not as Bad as You Think”	88
12.3 Engel and West (2005): “Exchange Rates and Fundamentals”	90
12.4 Rogoff (2008): “Comment on ‘Exchange Rate Models Are Not as Bad as You Think’”	93
<b>13 Forward Premium, Time-Varying Risk</b>	<b>93</b>
13.1 Alvarez, Atkeson and Kehoe (2009): “Time-Varying Risk, Interest Rates, and Exchange Rates in General Equilibrium”	93
13.2 Bachetta and van Wincoop (2010): “Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle”	95
13.3 Brunnermeier, Nagel and Pedersen (2009): “Carry Trades and Currency Crashes”	95
13.4 Burnside (2009): “Comment on ‘Carry Trades and Currency Crashes’”	96
13.5 Burnside (2011): “Carry Trades and Risk”	97
13.6 Engel (2011): “The Real Exchange Rate, Real Interest Rates, and the Risk Premium”	97
13.7 Farhi, Fraiberger, Gabaix, Ranciere and Verdelhan (2009): “Crash Risk in Currency Markets”	99
13.8 Lustig and Verdelhan (2007): “The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk”	99
<b>A Appendix: Abbreviations</b>	<b>101</b>

# 1 Real Business Cycles in Open Economies

## 1.1 Aguiar and Gopinath (2007): “Emerging Market Business Cycles”

**Summary** Emerging markets have strongly countercyclical CA, their ratio of consumption volatility to output volatility exceeds 1 and they sometimes face “sudden stops” in capital inflows. A&G show that these facts are consistent with a standard NGM in which trend growth is stochastic and subject to large shocks.

**Intuition** Transitory shocks to productivity have a hard time generating a very countercyclical CA since C is smoothed and so tends to rise less than Y. After a shock to *trend* productivity, however, I increases much more and life-time wealth rises more relative to current Y. Thus, the initial C response may be greater than that for Y. This gives relatively volatile C and a countercyclical CA.

### Model

$$Y_t = e^{z_t} K_t^{1-\alpha} (\Gamma_t L_t)^\alpha,$$
$$z_t = \rho_z z_{t-1} + \varepsilon_t^z,$$
$$\Gamma_t = e^{g_t} \Gamma_{t-1}, \quad g_t = (1 - \rho_g) \mu_g + \rho_g g_{t-1} + \varepsilon_t^g.$$

CD utility in consumption and leisure. Quadratic capital adjustment costs. Upward-sloping supply of loans (Schmitt-Grohe and Uribe, 2003). Constant world interest rate.

**Results** Can Beveridge-Nelson decompose log Solow residual

$$sr_t = z_t + \alpha \log \Gamma_t = \tau_t + s_t.$$

Since  $\lim_{K \rightarrow \infty} K^{-1} \text{Var}(sr_t - sr_{t-K}) = \sigma_{\Delta\tau}^2$ , can estimate ratio  $\sigma_{\Delta\tau}^2 / \sigma_{\Delta sr}^2$  using empirical Solow residuals. Treat ratio as measure of importance of trend shocks. Results are sensitive to choice of  $K$ .

Opt instead for structural estimation. Map model-implied second moments to HP filtered data for Mexico and Canada. Impulse responses (and intuition above) suggest that relative volatility of C and I as well as behavior of TB should be informative about extent of trend shocks. On the other hand, serial correlation of C as well as correlation of C and Y are not very sensitive to  $\sigma_g / \sigma_z$ . Method of Moments estimates give  $\sigma_g / \sigma_z \approx 5$  for Mexico and about 1 for Canada. Qualitatively same for different matched moments and calibrated parameters. Hansen  $J$  test does not reject overidentifying restrictions. Fitted moments correspond well to data.

Finally use Kalman smoother to decompose observed Solow residual for Mexico into  $g_t$  and  $z_t$  components. Feed the estimated shocks into NGM to calculate implied path of TB/Y ratio. Get sudden stop during Tequila crisis of 1994–1995: The observed drop in Solow residual attributed mainly to trend shock, so Y, C and I contract and CA goes sharply positive.

Working paper estimates model with process for world interest rate, find similar results.

## 1.2 Backus, Kehoe and Kydland (1992): “International Real Business Cycles”

**Summary** One of the original attempts to fit an OE RBC model. Two countries. Complete markets. Allow correlation of technology shocks and spill-over. Get *comovement puzzle*: Cross-country correlations of Y and I strongly positive in data but shouldn’t be in theory. C more correlated in theory than in data. I and TB way too volatile in theory relative to data.

**Intuition** Complete markets provide risk sharing so consumption should be highly correlated across countries (MUs are equated but different responses of labor make the C-correlation less than 1). Output and investment should be allocated to the most productive country. Theory thus predicts that the Y-correlation lies between 0 and the C-correlation. But this isn’t the case for developed countries.

**Model** Two countries, one good, complete markets. Utility  $U(c, l) = [c^\mu(1-l)^{1-\mu}]^\gamma/\gamma$ .<sup>1</sup> CD production function  $F(k, l) = \lambda k^\theta l^{1-\theta}$ .<sup>2</sup> Capital accumulation  $k_{t+1}^i = (1-\delta)k_t^i + i_t$ .<sup>3</sup> Technology process  $\lambda_t = (\lambda_t^h, \lambda_t^f)'$  follows VAR

$$\lambda_{t+1} = A\lambda_t + \varepsilon_{t+1}.$$

May be contemporaneous correlation of shocks. Off-diagonal elements of  $A$ : spillover effect. Due to market completeness, can solve planner’s problem

$$\psi E \sum_{t=0}^{\infty} \beta^t U(c_t^h, l_t^h) + (1-\psi) E \sum_{t=0}^{\infty} \beta^t U(c_t^f, l_t^f)$$

subject to previous constraints as well as aggregate resource constraint

$$\sum_i (c_t^i + i_t) = \sum_i F(k_t^i, l_t^i).$$

Let  $\psi = 1/2$ . Planner equates MU. If  $\gamma = 0$  (log utility), C is equalized across countries. Otherwise, labor effort dilutes cross-correlation of C a bit.

**Results** Most parameters calibrated as in Kydland and Prescott (1982). Estimate bivariate technology VAR using empirical Solow residuals from U.S. and EU. Impose symmetry for simplicity. Spillover coefficient 0.088, correlation of innovations 0.258. Linearize model and compare simulated moments to HP filtered data. Get relative volatility of C slightly less than in data, way too volatile I and TB. TB acyclical in model (countercyclical in data). S and I are weakly positively correlated in model (inconclusive in data). Cross-country correlation for C much larger than for Y, in contrast to data. Impulse responses show that this is due to efficient resource reallocation and risk sharing.

Try out different modeling choices. Correlation of S and I sensitive to choice of  $A$ . Raising risk aversion (lowering  $\gamma$ ) dampens volatility slightly and lowers cross-country correlation of C, but not by much. Time-to-build influences volatility of investment a lot.

Try adding trading frictions, e.g., transport costs, in the form of a quadratic cost of net exports in resource constraint. Reduces variability of TB a lot but slightly increases cross-country correlation

<sup>1</sup>More generally, they allow disutility of labor to depend on lags of leisure.

<sup>2</sup>More generally, allow for inventories.

<sup>3</sup>More generally, allow for time-to-build of more than 1 period.

of C. Large effect of small transport costs seems to be due to small welfare gain of international trade.

Try getting rid of all trade and risk-sharing, so that only difference from complete autarky is correlation of technology shocks. Gives results remarkably close to model with trading frictions: Due to spillovers, after positive shock to Home, Foreign consumer knows he will be richer in future and so consumes now and invests later.

Kehoe and Perri (2002): If bonds are only financial instrument, still get comovement puzzle. Bond Euler equation is

$$U_c(c_t^i, 1 - l_t^i)p_t = \beta E_t[U_c(c_{t+1}^i, 1 - l_{t+1}^i)],$$

where  $p_t$  is time- $t$  price of ZCB that pays off at  $t + 1$ . Yields

$$\frac{U_c(c_t^h, 1 - l_t^h)}{E_t[U_c(c_{t+1}^h, 1 - l_{t+1}^h)]} = \frac{U_c(c_t^f, 1 - l_t^f)}{E_t[U_c(c_{t+1}^f, 1 - l_{t+1}^f)]},$$

so tight link between MU is broken compared to complete markets. But quantitatively impact is very small. Due to shocks being purely transitory, the reduced insurance of bonds doesn't hinder risk sharing much.

## 2 International Risk Sharing and Co-Movements

### 2.1 Coeurdacier and Rey (2011): “Home bias in open economy macroeconomics”

**Summary** Literature review of explanations for home bias in international portfolio holdings. Develop approximation to optimal steady state portfolio in a model with equities, non-tradable labor income risk and RER (purchasing power of income) risk. Expand model to include investment and trade in bonds.

**Model** Two countries, two goods. CRRA utility (coefficient  $\sigma$ ), separable in consumption and leisure. Consumption preferences are biased in favor of home good. Show that markets are “locally complete:” For a certain choice of zero-order portfolio, the Backus-Smith condition can be made to hold up to first order. The zero-order portfolio satisfies

$$S = \frac{1}{2} - A(1 - \alpha) \frac{\text{Cov}(\widehat{w_t l_t}, \widehat{d_t})}{\text{Var}(\widehat{d_t})} + B \left(1 - \frac{1}{\sigma}\right) \frac{\text{Cov}(\widehat{RER_t}, \widehat{d_t})}{\text{Var}(\widehat{d_t})},$$

where hats denote Home variables relative to Foreign. The above is a partial equilibrium relation: to achieve local completeness of markets in the basic model's general equilibrium, the two covariances must be restricted. The first term above is the baseline Lucas (1982) diversification term. The second term represents hedging of nontradable labor income. The third term represents RER hedging: If  $\sigma > 1$ , households want to insure against RER appreciation that causes fall in purchasing power (due to home bias in preferences); if  $\sigma < 1$ , would rather have equity payout when consumption bundle is relatively cheap so as to buy more stuff. For log investor (Heathcote and Perri, 2009), RER hedging term drops out.

In general equilibrium, the RER hedging term is proportional to  $(1 - 1/\sigma)/(\lambda - 1)$ , where  $\lambda$  is approximately equal to the elasticity of substitution between Home and Foreign goods. For

$\lambda = 1$ , increases in  $Y$  are perfectly offset by ToT, so equities are perfect substitutes and portfolio is indeterminate (Cole and Obstfeld, 1991). For  $\lambda > 1$ , price responses are small, so  $\widehat{d}_t$  increases with  $\widehat{y}_t$ . For  $\lambda < 1$ , price responses are large (e.g., a bad Home productivity shock causes a large appreciation of the ToT, thus increasing Home's relative equity returns and leading to an appreciation of the RER), so local dividends provide a hedge against RER risk.

The authors extend the model to allow for trade in two bonds that pay off each country's consumption bundle. Also add (home biased) investment. Add shocks to disutility of labor so as to make the portfolio position determinate. Get equity share expression similar to the above, but now all second moments are *conditional* on bond returns. Intuitively, nominal bonds provide a perfect hedge (in this model) against RER movements, so they are used to hedge RER risk in equilibrium. That only leaves the conditional labor income risk, which unambiguously generates home bias (keeping relative prices fixed, a shock that raises  $I$  also raises domestic labor compensation due to home bias in  $I$ , while lowering dividend payout). In the data, the unconditional moments fail to generate home bias according to the formula above but the conditional moments do go the right way. Hence, the presence of bonds partially resolves the puzzle.

The authors show empirically that the time series and cross-country characteristics for home bias in bonds and bank loans are similar to those for equities.

**Other papers discussed** Van Wincoop and Warnock (2008) find that the empirical correlation between excess equity returns and the RER is low, and most of the movements in RER are driven by the NER, which can be hedged in forward markets.

Baxter and Jermann (1997) found that labor and capital returns are highly correlated but this has been challenged by other papers.

Sections 5 and 6 review literature on trade costs and informational frictions.

## 2.2 Corsetti, Dedola and Leduc (2008): “International Risk Sharing and the Transmission of Productivity Shocks”

**Summary** The Backus-Smith condition under complete markets implies that  $C$  in Home should be high when it is relatively cheap, i.e., relative  $C$  should increase when the RER depreciates. The correlation goes the other way in the data. In bonds-only economies, the Backus-Smith result holds in ex ante expectation, but under two different scenarios (both with strong wealth effects of shocks) the ex post correlation could have the other sign: (1) if the price elasticity of tradables is low, or (2) if shocks are very persistent, the output response is hump-shaped and the price elasticity of tradables is high. In both cases, relative price movements are in fact counterproductive for cross-country insurance.

**Model** First illustrate intuition in endowment economy. Two countries, two goods. Home bias in the  $C$  CES aggregator with elasticity of substitution  $\omega$ .

- Under complete markets, an exogenous rise in domestic output unambiguously causes the ToT to depreciate to facilitate risk sharing.
- Under financial autarky (and more generally it turns out), the response of Home's consumption of the Home good to a ToT appreciation depends on the relative magnitudes of the substitution effect (higher relative price) and income effect (higher wealth = value of endowment). The substitution effect dominates iff.  $\omega > 1$ . Income and substitution effects go

the same way for Foreign’s consumption of the Home good. Assume now that  $\omega$  is low. All the above implies that when Home’s  $Y$  is shocked upwards, the ToT must appreciate, since a depreciation would cause too large of a negative income effect for Home to make total demand for Home’s good match the higher supply.

- With international trade in bonds, and when  $\omega$  is high, a positive shock to Home’s  $Y$  endowment is met with a muted long-run depreciation in the ToT, since prices don’t need to adjust much to clear markets. The wealth effect for Home is thus substantial, generating a short-run  $C$  boom. Due to home bias, if the output profile is hump-shaped (as it will be in a bond economy with investment), there is excess demand for Home’s tradable good in the short-run, causing the ToT to appreciate on impact.

After laying out the basic intuition, the authors calibrate a full-fledged model with production. The two countries each produce a tradable and a non-tradable good. To get deviations from PPP, assume that distribution of tradable good requires  $\eta$  units of non-tradable good. This gives wedge between producer and consumer prices, and so LOP holds at wholesale but not consumer level. An international bond is traded, and there’s  $I$  with partial depreciation. Tradable and non-tradable technology shocks for both countries follow VAR.

**Results** Calibration is fairly standard, but focus is on the elasticity  $\omega$ . Have two calibration/estimation strategies. First is to use moment matching between model and data; gives  $\omega$  below 1/2. Second is to draw on empirical trade literature, where  $\omega \approx 4$ .

When  $\omega$  is set to the low estimate, the Backus-Smith correlation produced by the model is indeed negative, although not as much as in the data. Technology spillovers and the bonds-only negative transmission mechanism (whereby Foreign’s wealth effect is negative and so raises labor abroad) generate positive co-movement in  $Y$ ,  $I$  and  $N$ .

When  $\omega$  is high, it is found that highly persistent TFP shocks are necessary to resolve the Backus-Smith puzzle. In this case, the model has a moderately negative correlation. Output is hump-shaped due to endogenous capital accumulation. Due to high  $\omega$ , international prices are much less volatile than in the data.

### 2.3 Heathcote and Perri (2009): “The international diversification puzzle is not as bad as you think”

**Summary** Extend Backus, Kehoe and Kydland (1992) to a model in which the only assets traded are equities in Home and Foreign firms. Make assumptions so that asset markets are endogenously complete, but an incentive emerges for investors to hedge their non-tradable labor income risk. Investment breaks the tight link between labor income and equity dividends so that domestic equity is a good hedge, leading to portfolio home bias. Derive closed-form expression for optimal portfolios.

**Model** Two countries, two non-traded final goods produced with two traded intermediate goods (domestic and foreign). Log utility, separable in leisure. The final good aggregator is CD, meaning unitary elasticity of substitution between Home and Foreign intermediate goods. CD production function implies constant labor share. Households invest in domestic and/or foreign equity, receive dividends each period equal to value of production minus labor compensation and investment. Final

good preferences are home biased, and  $I$  is carried out with domestic final good, making  $I$  home biased as well.

**Theoretical results** Proposition 1 shows that full risk-sharing obtains in equilibrium, and this is achieved with constant portfolio shares. Intuitively, if full risk-sharing across time and states is possible with some portfolio, then there's no incentive to actively retrade. Full risk-sharing is possible, since one can derive two linear relationships between relative  $Y$ ,  $I$  and  $C$ : one from the budget constraint, and one from market clearing and the constant expenditure shares on Home and Foreign intermediate goods implied by CD. For a unique choice of the portfolio share, it is then possible to satisfy these two linear equations together with the Backus-Smith condition (the latter is also linear due to log utility).

The optimal portfolio share depends on the covariance of (relative) labor income with equity dividends. A persistent positive productivity shock to Home raises Home's relative labor income. (In principle, the rise in  $I$  could crowd out labor income if the investment good were very expensive, but due to unitary EoS, the RER depreciates moderately and the crowding out doesn't happen.) It has three effects on Home's relative dividend:  $Y$  is increased (positive), ToT depreciate (negative) and  $I$  increases (negative). Under the maintained assumptions, the two latter effects dominate. Hence, Home's equity is a good hedge for non-tradable labor income risk.

*Comparative statics:* As the trade share increases, portfolio home bias decreases, since the investment good is comprised of a more equal mix of the two intermediate goods, leading to larger ToT depreciation following a positive TFP shock (need larger price adjustment to clear market for domestic goods), i.e., larger offsetting price effects, less variable relative labor income and thus less demand for a hedge. As labor's share increases, portfolio home bias increases, since non-traded income risk is greater so the demand for a hedge rises.

Cole and Obstfeld (1991) have similar model but with full depreciation and transitory productivity shocks. These assumptions make  $I$  a constant proportion of  $Y$ . In Heathcote and Perri's model, however, persistent productivity shocks generate surges in  $I$ , breaking the link between Home and Foreign dividends, so that the optimal portfolio isn't indeterminate.

Baxter and Jermann (1997) develop a model with a single consumption good and production. Due to CD production function and no investment, labor and dividend income in a country are perfectly correlated, so investors should aggressively short domestic equity, leading to stark negative home bias. Heathcote and Perri argue that the single-good and no-investment assumptions are crucial here.

Using numerical methods, the authors carry out a sensitivity analysis. They relax the assumptions of logarithmic utility and CD consumption aggregator. Home bias still obtains for a wide range of parameterizations. However, if the elasticity of substitution is very high, Home equity returns exceed Foreign following a positive Home TFP shock, which kills home bias.

**Empirical results** Using a cross section of OECD data, the authors show that diversification increases in trade openness, as predicted. Country size and GDP per capita do not offer additional predictive power. The "general equilibrium channel" that connects trade openness to the covariance of earnings and dividends holds in the data. So does the "partial equilibrium channel" that connects said covariance to diversification (although the relationship is not as steep as predicted). Changes in trade share over time predict changes in diversification.

## 2.4 Kehoe and Perri (2002): “International business cycles with endogenous incomplete markets”

**Summary** To get cross-country correlations of consumption to be lower than those for output as well as positive co-movement of employment and investment, need more than just assume bond-only credit markets. Can get the right kind of co-movement by assuming that contracts must be enforceable by the threat of exclusion from future trade. Such contracts severely limit risk sharing.

**Model** Two countries, one good, production only uses domestic labor, country-specific technology shocks. For the economy with enforcement constraints, require that the social planner’s solution at every point in time gives both countries a larger continuation utility than the value of financial autarky.

The enforcement constraint involves future consumption and employment and so the problem seems to not be stationary. However, the authors show that the Lagrangian for the social planner’s problem with enforcement constraints can be written in a form where essentially the discount factor is a function of past Lagrange multipliers. It is then shown that the problem is stationary when the state space is augmented by the relative value (for the two countries) of the generalized discount factor.

Also consider a complete markets economy and a bonds-only economy.

**Results** Choosing CD utility and production and standard parameter values, they solve the various models using a clever value function iteration algorithm. The complete markets and bond economies give similar results for co-movement, à la Backus, Kehoe and Kydland (1992). The enforcement economy, however, gets about the same cross-country correlation for C and Y, while I and N have positive cross-correlation like in the data. TB remains procyclical, unlike in the data. The volatility of TB and I is also drastically reduced even without resorting to adjustment costs.

Intuitively, a positive TFP shock to Home raises its value of autarky, so planner must restrict investment flows to satisfy enforcement constraint. In fact, capital is built up in Foreign to discourage Home from reverting to autarky (leads to procyclical TB). The planner must promise persistently high C for Home relative to Foreign, at least until the TFP shock decays.

## 3 Global Imbalances

### 3.1 Caballero, Farhi and Gourinchas (2008): “An Equilibrium Model of ‘Global Imbalances’ and Low Interest Rates”

**Summary** The recent couple of decades have seen the U.S.’s current account deficit explode, the real interest rate decline and the foreign portfolio holdings of U.S. assets increase. This can be explained by a negative shock to the pledgeability of assets in the rest of the world, since this reduces asset supply, causing the interest rate to drop and the U.S. CA to turn negative, even in the long run.

**Model** Time is continuous. Infinitesimal agents die and are born at rate  $\theta$ . When they enter, they are endowed with  $(1 - \delta)X_t$  units of output (in total, not each). They consume all their savings when they die. The only savings vehicle is a tree with dividend rate  $\delta X_t$ . Let  $V_t$  be its price. Then

its return is

$$r_t = \frac{\dot{V}_t}{V_t} + \frac{\delta X_t}{V_t}.$$

By the assumptions, the total stock of savings  $W_t$  evolves according to

$$\dot{W}_t = -\theta W_t + (1 - \delta)X_t + r_t W_t.$$

In equilibrium, the supply and demand for savings are equal:  $V_t = W_t$ . Output is exogenous and grows at rate  $g$ . The above conditions can be combined to yield

$$r_{aut} = g + \delta\theta.$$

Note that  $\delta$  can be interpreted as the fraction of output that is pledgeable, i.e., can be capitalized in traded form today. Assuming a constant interest rate, we have

$$V_t = \delta \int_t^\infty e^{-r(s-t)} X_s ds = \delta P V_t,$$

$$W_t = W_0 e^{(r-\theta)t} + (1 - \delta) \int_0^t e^{(r-\theta)(t-s)} X_s ds,$$

so the supply of assets is increasing in  $\delta$  while the demand is decreasing (with more available assets for capitalization, the non-financial part of output is reduced). If  $\theta = 0$ , a Ricardian-type equivalence holds: Any change in  $\delta$  has offsetting effects on supply and demand for assets. If  $\theta > 0$ , asset demand has a component that is determined by entering and exiting agents. Consequently,  $\delta$  affects the total resources perceived by agents (as in Blanchard's OLG model). Higher  $g$  increases expected capital gains and thus the return. Given a non-Ricardian model  $\theta > 0$ , the higher  $\delta$ , the higher is savings supply relative to demand, so the asset price is lower and the return is higher.

Define

$$TB_t = X_t - \theta W_t, \quad CA_t = \dot{W}_t - \dot{V}_t.$$

Note that  $\theta W_t$  is consumption at time  $t$ . Can get explicit formulas for  $V_t$  (Gordon growth model) and  $W_t$  for given world interest rate  $r$ , as  $t \rightarrow \infty$ . Gives long-run Metzler diagram.

The main intuitions from the model come in the two-country case where  $\delta^i$  may differ. Assume  $r_t$  is the same across countries. Equations for each country are as above. The implied equilibrium interest rate is

$$r_t = g + (\delta^U - x^R(\delta^U - \delta^R))\theta,$$

where  $x^R$  is  $R$ 's share of world output.  $U$  denotes the U.S.,  $R$  the rest of the world.

Consider a symmetric initial position. Suppose that  $R$ 's pledgeability drops permanently. The interest rate drops by the equation above.  $V^U/V^R$  then increases. If there is home bias in portfolio choice, relative savings  $W^U/W^R$  will also increase. As  $R$ 's savings are below the steady state balanced growth path, its savings increase at a rate faster than  $g$ , and many of these funds flow to  $U$  since capitalization in  $R$  has decreased. Hence,  $U$  runs a CA deficit and  $U$ 's share in  $R$ 's portfolio increases. The CA remains in deficit even asymptotically since  $R$ 's supply and demand curves have shifted, cf. Metzler diagram (Figure 3). Intuitively,  $R$ 's saving need grows exogenously with output. It is also clear that the larger  $R$  is compared to  $U$ , the larger must the above-mentioned CA deficit be. The authors formalize this intuition by allowing for different growth rates of output.

The authors note that part of their story could also have been told using an decline in  $\theta^R$ , i.e., exogenous increase in savings (the “global savings glut” story). The only substantial difference is that  $R$ ’s asset price would rise rather than fall on impact, which is counterfactual for the 1990s.

The model is extended in three directions.

- Suppose new agents have an option to plant a certain number of trees at an investment cost. If  $\delta^R$  drops, cost of investment may outweigh the gain in  $R$ , leading to an investment slump. This shifts out the demand schedule in the Metzler diagram further, exacerbating  $U$ ’s CA deficit and further lowering the interest rate.
- Suppose  $R$  residents can sell their trees to  $U$  residents at an exogenously given price (FDI) and that  $U$  residents can capitalize the dividend from the purchased trees at the rate  $\delta^U$ . Then there are gains from trade.  $U$  residents earn intermediation rents, which may be thought of as nontraditional exports. The consequence is that after a negative shock to  $\delta^R$ , the supply of  $U$ -like assets increases which eventually offsets the shock. The intermediation rents finance a permanent trade deficit.
- The authors finally extend the model to have multiple goods. Consumers have CES preferences with home bias. After a drop in  $\delta^R$ ,  $U$  residents are richer than  $R$  residents, so due to home bias the relative demand for  $U$  goods increases, causing an appreciation of  $U$ ’s RER.

### 3.2 Mendoza, Quadrini and Ríos-Rull (2009): “Financial Integration, Financial Development, and Global Imbalances”

**Summary** Countries differ in their financial development. Less financially developed countries may be more restricted by enforcement constraints that prevent them from diverting funds. This restricts the menu of contingent assets available to them (in the limit only non-contingent trade is possible) and thus risk sharing capabilities, depressing the interest rate due to precautionary saving. When a less financially developed country integrates with a more developed one, the former’s demand for riskless bonds causes the developed country to maintain a large negative position in riskless assets accompanied with a positive position in risky, productive assets (these can be more efficiently insured by the developed country).

**Motivation** There is a high degree of heterogeneity in financial markets development. The decline in the U.S.’s NFA position started as the financial globalization process got underway in the 1980s and saw the U.S. take positive net positions in risky assets (portfolio equity and FDI) accompanied with a very large negative position in riskless assets.

**Model** In the simple version, there are two countries  $i = 1, 2$  each with measure 1 of agents. No aggregate uncertainty but agents face idiosyncratic endowment  $w_t$  and investment  $z_t$  shocks. Utility is concave with positive third derivative. Each country is endowed with one productive, non-tradable asset with price  $P_t^i$ . Each agent in the country can use this asset to produce with a one-period lag. Individual production function is  $y_{t+1} = z_{t+1}k_t^\nu$  (DRS due to managerial input).

The individual state  $s_t = (w_t, z_t)$  evolves according to transition probabilities  $g(s_t, s_{t+1})$ . Agents can trade contingent Arrow securities  $b(s_{t+1})$  with price  $q_t^i(s_t, s_{t+1}) = g(s_t, s_{t+1})/(1+r_t)$ . Net worth is

$$a(s_{t+1}) = w_{t+1} + k_t P_{t+1}^i + z_{t+1} k_t^\nu + b(s_{t+1}),$$

and the budget constraint is

$$a_t = c_t + k_t P_t^i + \sum_{s_{t+1}} b(s_{t+1}) q_t^i(s_t, s_{t+1}).$$

Number states  $n = 1, \dots, N$ , where 1 is the worst. Agents are subject to a limited liability constraint

$$a(s_n) \geq 0$$

and an enforcement constraint

$$a(s_n) - a(s_1) \geq (1 - \phi^i)[(w_n + z_n k_t^v) - (w_1 + z_1 k_t^v)],$$

where  $\phi^i$  is the fraction of diverted income that is lost when lying about the state being  $n = 1$ . Key assumption:  $\phi^i$  is determined by country of residence.

Consider a situation with *only endowment shocks* and *autarky*. A country with high  $\phi$ , such that the enforcement constraints don't bind, can perfectly insure idiosyncratic shocks. The return  $R_{t+1}(k, \bar{z})$  must also equal  $1 + r_t$ . Because this return is decreasing in  $k$ , all agents choose  $k = 1$  and output is constant over time. The interest rate must then satisfy  $(1 + r_t)\beta = 1$  for consumption not to outgrow output. A country with  $\phi = 0$  can only trade in riskless bonds. Aggregate output is still constant over time, but individual consumption can't be insured fully. The Euler equation

$$U'(c) = \beta(1 + r_t)E[U'(c(w'))]$$

implies  $\beta(1 + r_t) < 1$  for aggregate consumption not to grow.

If there are *only endowment shocks* but *financial markets integrate*, prices are equalized. By the same argument as above, all agents hold  $k = 1$ . Also,  $\beta(1 + r_t) < 1$  for aggregate consumption not to grow. The high- $\phi$  agents will then gradually reduce consumption over time until they hit the limited liability constraint. For this to bind,  $\sum b(w_{t+1})q(w_t, w_{t+1}) < 0$ , so the NFA position of the developed country is negative.

With *investment shocks only* and financial integration, the productive asset earns a risk premium in the less developed country, since

$$U'(c) = \beta E[U'(c(z'))R_{t+1}(k, z')]$$

and the riskless bond Euler equation imply

$$ER_{t+1} - (1 + r_t) = -\frac{\text{Cov}[R_{t+1}, U'(c(z'))]}{EU'(c(z'))}.$$

Agents in the low- $\phi$  country thus sell some of their productive asset in equilibrium to the high- $\phi$  country. We again get  $\beta(1 + r_t) < 1$ . High- $\phi$  agents again decrease consumption until the limited liability constraint binds and the NFA position in bonds is negative. The average return on the developed country's asset (the productive one) is higher than  $1 + r_t$  due to concavity of the production function (only expected *marginal* returns are equalized), and  $1 + r_t$  is the expected return on the liabilities.

**Quantitative results** The authors extend the model to allow for cross-country investment, i.e., agents can distribute their managerial capital across different countries’ productive assets (in the previous simple model, all managerial capital went into the domestic production technology). The relative economic size of countries is also allowed to vary. Financial integration is introduced in simulations as an unexpected event.

In a two-country scenario calibrated to the U.S. ( $\phi = 0.35$ ) and rest of the world ( $\phi = 0$ ), the U.S. very gradually builds up a positive net position in productive assets and a large negative position in bonds. The interest rate is lower than in U.S. autarky.

There are three possible welfare effects of integration: diversification of investment risk, risk specialization and changes in asset prices. Since the poorer residents in the rest of the world are hurt substantially by a higher interest rate, not all agents are made better off by the integration process.

The authors show that the model’s assumption that  $\phi$  is residence based can be relaxed somewhat without changing the results much.

Finally, they consider a three-country scenario where countries also differ in the lower bound  $\underline{a}^i$  on assets in the limited liability constraint. If countries differ in their credit capacity  $\underline{a}^i$ , the country with the lower  $\underline{a}^i$  will tend to generate a negative NFA position. However, differences in  $\underline{a}$  can’t explain differences in portfolio composition of risky assets and bonds. A simulation is carried out with the U.S. (high  $\phi$ , low  $\underline{a}$ ), other developed countries (high  $\phi$ , high  $\underline{a}$ ) and emerging markets (low  $\phi$ , high  $\underline{a}$ ). The steady state has other developed countries maintaining a positive position in productive assets but a significantly smaller absolute NFA position than the U.S., like in the data.

## 4 Valuation Effects

### 4.1 Curcuru, Dvorak and Warnock (2008): “Cross-Border Return Differentials”

**Summary** The “exorbitant privilege” of the U.S., i.e., the supposedly large returns differential on its assets and liabilities, has been mismeasured by most researchers. The oft-quoted BEA data is internally inconsistent as positions are revised frequently but flows are only partially revised. Better data sources indicate that the returns differential is 1%, almost all of which stems from FDI rather than financial investment.

**Motivation** Gourinchas and Rey (2007, chapter in “G7 Current Account Imbalances,” henceforth GR) found a returns differential of more than 3% per year for the U.S., mostly due to a large differential for equities as the U.S.’s net portfolio is skewed towards these.

**Theory** The return can be constructed as

$$r_t = \frac{A_t - A_{t-1} - FLOW_t}{A_{t-1}} + \frac{INC_t}{A_{t-1}}.$$

The first term represents capital gains, the second interest and dividend income. The BEA heavily revises positions  $A_t$  when new survey evidence infrequently comes in, but flows are only partially revised as firms find it hard to restate past flows.

**Results** Using an international portfolio data set from Bertaut and Tryon (2007), find essentially zero returns differential within each of equity and bond categories. This is consistent with the observation that U.S. asset returns (primarily equities) in GR are much higher than world stock index returns, while the liability returns (primarily bonds) are much lower than corresponding index returns.

The original BEA data give vastly different results than the revised data; average returns differential is only 1%. Almost all of this stems from FDI rather than bonds and equities. The original data is more in line with market index returns. The authors think the original data is better as it is at least internally consistent due to capital gains being imputed with the same vintage data.

The cumulative CA deficit is much more negative than the movement in the net international investment position, suggesting that there should be a high capital gains differential. If this is not the case, some account in the balance of payments must be mismeasured. The authors contend that it may be the CA and cite a U.S. Census Bureau study (1998), which found that exports may be understated by as much as 10%.

While the returns differential is small on average, it varies a lot over time, so the timing of leverage and returns may combine to give large short-run valuation effects.

## 4.2 Gourinchas and Rey (2007): “International Financial Adjustment”

**Summary** The CA can adjust in two ways: the trade channel and the valuation channel. Derive log-linear approximation to the external budget constraint around a deterministic trend. Use it to derive a relation between net external assets, TB and return differentials. Use Campbell-Shiller method to quantify effects of the two channels; valuation channel contributes about 25% of shocks to net external assets. Conversely, net external assets predict returns at short horizons, TB at long horizons and ER at all horizons.

**Motivation** U.S. has run persistently high CA and TB deficits since late 1980s. Almost all liabilities are denominated in dollars, and 70% of assets are denominated in other currencies. Return differentials may mediate part of the future CA adjustment, particularly through ER movements.

**Theory** Accumulation identity

$$NA_{t+1} = R_{t+1}(NA_t + NX_t),$$

where  $R_{t+1}$  is a weighted average of gross returns on assets and liabilities. Along a balanced growth path, ratios of X, M, assets and liabilities to wealth should be stationary, but strong upward trend in postwar data. Perhaps due to long-run structural change in financial markets. Assume instead that budget constraint holds for a fictitious *deterministic* economy with same growth rate for all four quantities. By subtracting stochastic b.c. from deterministic one, get b.c. in deviation from trend. Log-linearize this to get

$$nxa_{t+1} = \frac{1}{\rho} nxa_t + r_{t+1} + \Delta nx_{t+1},$$

where cyclical external imbalances  $nxa$  linearly combines stationary components of exports, imports, assets and liabilities, such that the 1st and 3rd contribute positively and the 2nd and 4th negatively. Weights depend on trend relationships. Return differential  $r$  increases with return on

assets and decreases with return on liabilities. If long-run return exceeds trend growth, can shift forward to get

$$nxa_t = - \sum_{j=1}^{\infty} \rho^j (r_{t+j} + \Delta nxa_{t+j}).$$

Also holds in expectation; can decompose  $nxa_t = nxa_t^r + nxa_t^{\Delta nx}$ .

**Empirics** Use data on U.S. net and gross foreign asset positions, capital gains and total returns between 1952 and 2004. Filter data with HP filter that filters out very long-run cycles. Argue that if there is extra information in long-run movements, the exercise is biased against them finding strong relationships. Construct cyclical external imbalances  $nxa$  with time-constant weights.

First exercise uses Campbell-Shiller VAR method to evaluate the above  $nxa$  decomposition. The two components are positively correlated: valuation and trade effects are mutually reinforcing. A variance decomposition (splitting the covariance term 50/50),

$$1 = \frac{\text{Cov}(nxa_t^r, nxa_t)}{\text{Var}(nxa_t)} + \frac{\text{Cov}(nxa_t^{\Delta nx}, nxa_t)}{\text{Var}(nxa_t)},$$

attributes 27% to returns and 64% to TB (9% unaccounted for).

The  $nxa$  relation also states that cyclical imbalances should predict returns and/or TB. There is significant forecasting power (in sample) for returns at short horizons and for TB at medium and long horizons. Returns are driven partially by ER movements. Find that ER can be predicted at all horizons; argue that this is because ER plays a role in both short-run valuation channel and longer-run trade channel. For the U.S., the ER effect has the same sign on both channels: An appreciation of the dollar causes the return differential to decrease (since U.S. liabilities are denominated in dollars and assets in foreign currency) and the TB to deteriorate through expenditure switching.

Also conduct out-of-sample ER forecasting exercise. Find that the cyclical imbalance model significantly outperforms a random walk model in terms of MSE, overturning Meese and Rogoff (1983).

## 5 Financial Crises, International Recessions

### 5.1 Mendoza (2010): “Sudden Stops, Financial Crises, and Leverage”

**Summary** If an SOE is subject to a leverage constraint on its access to working capital financing and one-period loans, Sudden Stops arise (as in the data) when bad shocks hit after a prolonged expansion that has led to high leverage. Due to precautionary saving, these Sudden Stops are sufficiently rare that the leverage constraint doesn’t much affect the long-run business cycle moments. The model has two credit channels: When the collateral constraint binds, the effective cost of borrowing rises. Furthermore, the resulting fire sale drives down the price of capital which amplifies the initial shock in a Fisherian way.

**Motivation** In the data, Sudden Stops (i.e., a sudden upward jump in the CA combined with a contraction of Y and C and a fall in asset prices) occur after expansions. They are relatively infrequent and asymmetric, in the sense that we do not observe sudden drops in the CA combined with booms in Y and C.

**Model** SOE model with a representative self-employed household. To get stationary NFA, preferences are of the Stationary Cardinal Utility kind with endogenous discount factor. Period utility is quasilinear in C and L à la GHH to eliminate the wealth effect on labor supply. Production requires an imported input  $\nu_t$  sold at an exogenous, stochastic world price  $p_t$  (drives a wedge between TFP and the measured Solow residual). There is an exogenous world interest rate  $R_t$  on non-contingent one-period bond loans. Working capital loans pay for a fraction  $\phi$  of the cost of labor and imported inputs in advance of sales. The budget constraint is then

$$c_t + i_t = e^{\varepsilon_t^A} F(k_t, L_t, \nu_t) - p_t \nu_t - \phi(R_t - 1)(w_t L_t + p_t \nu_t) - R_t^{-1} b_{t+1} + b_t.$$

There are also adjustment costs to net investment. The economy is subject to the collateral constraint

$$R_t^{-1} b_{t+1} - \phi R_t (w_t L_t + p_t \nu_t) \geq -\kappa q_t k_{t+1},$$

where  $q_t = \partial i_t / \partial k_{t+1}$  is the price of capital. Hence, total within-period debt (both working capital and bond loans) can't exceed a fraction of the marked-to-market value of capital.

The FOC for bonds is

$$\lambda_t - \mu_t = R_t E_t[\lambda_{t+1}],$$

where  $\lambda_t$  and  $\mu_t$  are the nonnegative multipliers on the budget and collateral constraints. Rearrange to obtain the external financing premium

$$R_t^h - R_t = \frac{\mu_t}{E_t[\lambda_{t+1}]}, \quad R_t^h := \frac{\lambda_t}{E_t[\lambda_{t+1}]}.$$

$R_t^h$  is the effective real interest rate (ratio of MUs). We see that the external financing premium is positive when the collateral constraint binds. The collateral constraint also positively affects the equity premium (the expected excess total return on capital, including dividends and capital gains), both directly when it binds and indirectly through its effect on the consumption smoothing and thus the covariance of MU and returns. Mendoza derives an iterated-forward expression for  $q_t$  as a PDV of future dividends (MPK minus MC of investment). Because a potentially binding collateral constraint raises the equity premium, it increases discounting and thus lowers the asset price.

There is also an external financing premium on working capital financing due to  $\mu_t$  entering into FOCs for labor and the imported input.

Apart from the financing premium, there is a debt-deflation mechanism. When the collateral constraint binds, agents demand less equity, which lowers  $q_t$  due to upward-sloping supply of equity (because of adjustment costs). Since the value of capital has declined, the collateral constraint then binds even more, leading to further fire sales, etc.

**Simulation** The model is calibrated to Mexican data and simulated. The three shocks (TFP, interest rate and imported input price) follow a  $2^3$ -point symmetric Markov process. The working capital share is  $\phi = 0.26$ . Two different values of the leverage constraint parameter  $\kappa$  is attempted, along with a model without the constraint ( $\kappa = \infty$ ).

Long-run business cycles moments are largely unaffected by the constraint. The constrained economy matches the data better in terms of relative volatility of C and Y, and countercyclicality of R and TB. Due to precautionary saving, agents go less in debt on average.

Mendoza identifies simulated events that look like Sudden Stops. They happen after expansions, like in the data, when the leverage ratio has been driven up. It is found that the constraint greatly amplifies the response of the key variables relative to an unconstrained economy in the same state. However, there is almost no amplification in non-Sudden-Stop states. Asset prices don't decline enough in the model relative to the data. Since imported input price shocks contribute to Sudden Stops, the decline in the measured Solow residual doesn't stem entirely from TFP shocks, although the latter are needed to get the right magnitudes.

It is found to be crucial to have working capital in the model, even though it doesn't do much difference in non-Sudden-Stop states. The reason is that capital is predetermined, so the immediate drop in  $Y$  and  $I$  after a bad shock must come from the financing premium's contemporaneous effect on working capital financing.

## 5.2 Neumeyer and Perri (2005): "Business Cycles in Open Economics: The Role of Interest Rates"

**Summary** In the data, real interest rates lead the business cycle in emerging SOEs.  $Y$ ,  $H$  and  $C$  are more volatile than in developed SOEs, and  $TB$  is more countercyclical. These facts can be matched by a standard NGM with a real interest rate that is determined by an international rate and a country-specific premium. Two important modifications are introduced to the model: Firms have a need for working capital, so labor demand responds to the interest rate, and consumer preferences are GHH, so labor supply does not respond to the interest rate. The business cycle characteristics of the model match the data best if the country-specific interest rate premium depends endogenously on its business cycle, as this creates extra amplification.

**Model** The model is standard NGM with GHH preferences. Firms have to pay a share  $\theta$  of the wage bill in advance, so they must borrow  $\theta w_t l_t$  goods at the start of the period at real rate  $R_{t-1}$ . At the end of each period, they can sell their output and pay off the rest of the wage bill plus the rental rate of capital. Total NFA are thus given by the households' NFA position  $b_t$  minus the firm's debt  $\theta w_t l_t$ .

The interest rate is given by  $R_t = R_t^* D_t$ , where  $R^*$  is the international rate and  $D$  the country-specific premium due to default risk (the government has a time-varying probability of expropriating the interest payments made by domestic firms to foreign lenders).  $D_t$  is modeled in one of two reduced form ways: Either it is driven by exogenous factors (in the calibrations it is given by an AR(1) process) or it is a function of the endogenous business cycle (in the calibrations it is a negatively linear function of expected next-period productivity, as motivated by models of endogenous default).

The combination of firms' and households' FOCs gives

$$\frac{1}{1 + \theta(R_t - 1)} A_{t+1} F_{l,t+1} = w_{t+1} = -\frac{u_{l,t+1}}{u_{c,t+1}}.$$

Hence, working capital creates a labor wedge, since  $\theta(R_{t-1} - 1)w_t$  is the net interest per hour on the fraction of the wage bill that is paid with borrowed funds. Consider an upward shock to the interest rate. The labor demand shifts inward after the initial period. For GHH preferences, labor supply is independent of the interest rate (since it is independent of consumption), so equilibrium labor (and thus  $Y$ ) falls.  $C$  falls due to both the substitution effect and the substitutability of  $C$  and leisure. Since a higher interest rate induces savings,  $TB$  is countercyclical.

For C-D preferences, on the other hand, labor supply shifts out with the initial drop in consumption (substitution effect), while labor demand doesn't respond initially due to the interest payments already being locked in. A boom in H and Y follows on impact. In future periods, labor demand shifts in, so equilibrium H and Y can go either way. Hence, for C-D preferences to fit the facts, the IES has to be extremely small.

**Simulations** A calibration is performed to simulate the model, setting  $\theta = 1$ . It is found that movement in  $D_t$  is necessary to generate countercyclical interest rates and TB. With exogenous country risk, these comovements can attain the right sign but not magnitude. With feedback from the business cycle into  $D_t$ , however, the main comovements match the data well, although H, C and TB are too volatile (households are too willing to substitute). Shutting down movements in  $R^*$  only marginally reduces Y volatility. Shutting down feedback from productivity shocks to country risk  $D_t$  reduces volatility of Y by 27%.

The qualitative conclusions hold even with  $\theta = 0.5$ . The volatilities are very sensitive to the elasticity of labor supply.

### 5.3 Perri and Quadrini (2011): “International Recessions”

**Summary** Two-country model. Firms are subject to a no-default enforcement constraint. Upon default, a firm's assets can be sold to other firms (efficient) or turned into consumption goods (inefficient). Multiple self-fulfilling equilibria arise since the liquidation value of capital depends on firms' ability to buy up defaulting competitors, which depends on whether they are constrained, which again depends on the liquidation value of capital. Due to risk-sharing across countries, enforcement constraints bind simultaneously in the two countries. Hence, endogenous belief-driven credit shocks lead to international comovement of real and financial variables.

**Motivation** During the 2007-2009 crisis, developed countries exhibited a striking degree of comovement in real and financial variables. The contraction in Y was preceded by a contraction of credit. Leading up to the crisis, credit had expanded more rapidly than real variables.

**Model** Consider first a model without capital accumulation. Two countries, two types of agents in each: investors, who own firms, and workers. Investors derive income only from firm dividends. Workers are more patient, so firms borrow from workers. Firms face a cashflow mismatch: wage, dividend and current debt payments (net of new issue) are made before the realization of revenue. To cover the mismatch, they contract an intra-period loan

$$l_t = w_t h_t + d_t + (b_t - R_t^{-1} b_{t+1}) = F(h_t),$$

where the last equality follows from the firm's budget constraint. They can default, leading to an enforcement constraint

$$\xi_t \bar{k} \geq l_t + \frac{b_{t+1}}{R_t}.$$

Here  $\xi_t$  is the liquidation value of capital. The RHS are the total liabilities of the firm. The partial equilibrium implication is that when the constraint tightens, firms must cut back on dividends. However, since their owners have concave utility, they only cut back moderately and instead effectuate the remaining adjustment by reducing labor demand, leading to a contraction.

There is perfect capital mobility, and since investors only derive utility from consumption, they choose to perfectly diversify their portfolios. There is then a representative international investor, and firms discount profits with the same SDF  $m_t$ . Let  $\mu_t$  be Lagrange multiplier on the domestic firm's enforcement constraint. Then the FOC wrt. debt  $b_{t+1}$  is

$$E_t[m_{t+1}] = \frac{1 - \mu_t}{R_t}.$$

Since  $m$  and  $R$  are equalized across countries, so must  $\mu$  be. This implies international comovement of credit conditions. The labor wedge is then also equalized across countries.

To reduce the number of state variables, domestic workers are assumed to trade Arrow-Debreu securities with foreign workers (firms or investors can't participate). This leads to a constant consumption ratio between workers in the two countries. Due to this along with the fact that there is a representative international investor, the only state variable is worldwide debt. Also because of the constant consumption ratio and same labor demand in the two countries, the wage ratio is equalized, so real variables comove perfectly in the model without capital accumulation.

If shocks to  $\xi_t$  and  $\xi_t^*$  (foreign liquidation value) were exogenous, debt flows between the two countries would tend to counteract the real effects of the shocks (in order to smooth dividends). The authors now introduce an endogenous credit shock channel. Suppose a unit of capital can either be turned into  $\underline{\xi}$  units of consumption, or it can be sold to another firm and transformed into  $\bar{\xi}$  units of reinstalled capital,  $\underline{\xi} < \bar{\xi} < 1$ . Non-defaulting firms are only able to buy up defaulting firms' assets if they are unconstrained, which depends on the value of  $\xi_t$ . Multiple self-fulfilling equilibria arise. The authors argue that since  $\mu_t$  is equalized across countries, so is  $\xi_t$ . Hence, the model features endogenous international credit shocks.

They extend the model to allow for capital accumulation. To limit the number of state variables, they assume that the C-D production function depends on capital only through the total worldwide stock (positive externalities), and there are CRS in reproducible factors (AK). They can then normalize all equations so that only worldwide debt and the domestic share of aggregate capital are state variables.

**Simulations** They calibrate the model to match the U.S. as country 1 and the other G7 countries as country 2. For the sunspot process, they assume that in states in which multiple equilibria are possible, both countries expect  $\xi_t = \underline{\xi}$  with some constant probability.

A credit expansion is simulated by imposing a long sequence of  $\xi_t = \underline{\xi}$  draws and then switching to  $\xi_t = \bar{\xi}$  going forward (a credit contraction is simulated in the opposite way). Following a credit expansion, the stock of debt increases.  $Y$  and  $H$  expand, but only gradually. The response to a credit contraction is much larger at impact. When experimenting with the length of the expansion, the authors find that a prolonged expansion leads to a more severe subsequent reversal.

The model generates large volatility of labor and stock market valuations. If the model is simulated with TFP shocks instead (calibrated using Solow residuals), the volatility is subdued.

Finally, the authors extend the model to try to explain the more severe response of labor in the U.S. relative to Europe. By adding variable labor utilization (to increase fluctuations in measured labor productivity) and adjustment costs in hours (to create differential cross-country effects from rigidities). If the rigidity in hours is different across the two countries, the response of  $H$  is much larger in the less rigid country, with  $Y$  and  $I$  not differing much. Consequently,  $Y/H$  falls more in the less rigid country.

## 6 Capital Controls

### 6.1 Bianchi (2011): “Overborrowing and Systemic Externalities in the Business Cycle”

**Summary** Calibrates a standard two-good SOE model with a borrowing constraint. Compares with the constrained efficient allocation by a social planner who internalizes the effect of borrowing on the value of pledgeable wealth (which depends on the relative price of nontradables). The decentralized economy has much more frequent financial crises, and conditional on a crisis, the downturn is much worse than for the constrained planner, who does more precautionary saving. The constrained optimum can be implemented through state-contingent taxes or margin requirements of moderate size.

**Model** SO endowment economy with a tradable and nontradable good. Measure 1 of consumers. The consumption aggregator is of CES type. The price of tradables is normalized to 1. Price of nontradables is  $p^N$ . Creditors restrict loans such that the value of debt does not exceed a fraction  $\kappa^T$  of tradable income plus a fraction  $\kappa^N$  of nontradable income,

$$b_{t+1} \geq -(\kappa^N p_t^N y_t^N + \kappa^T y_t^T).$$

If  $\kappa^N$  differs from  $\kappa^T$ , it can be seen as outside creditors having a higher preference for tradable collateral (e.g., due to nontradable collateral being subject to domestic judicial practices).

The intertemporal FOC for consumption allocation gives

$$p_t^N \propto \left( \frac{c_t^T}{c_t^N} \right)^{\eta+1}. \quad (1)$$

Hence, an equilibrium reduction of  $c_t^T$  due to a binding borrowing constraint generates a fall in the price of nontradable consumption, which further leads to debt deflation.

The constrained efficient allocation is one in which the social planner performs the credit operations for the economy (i.e., decides the stock of debt), rebates proceeds lump sum and lets private agents allocate consumption. Due to identical preferences, the wealth distribution doesn't matter, so the social planner effectively carries out the representative agent's optimization problem subject to the additional equation (1), i.e., the planner internalizes the pecuniary externality. The social planner's FOC wrt. tradable consumption is then

$$\lambda_t^{sp} = \frac{\partial u_t}{\partial c_T} + \mu_t^{sp} \Psi_t,$$

where  $\lambda_t^{sp}$  is the multiplier on the budget constraint (shadow value of wealth) and  $\mu_t^{sp}$  the multiplier on the borrowing constraint. If  $\Delta_t$  denotes the collateral value (RHS of the borrowing constraint),  $\Psi_t = (\partial \Delta_t / \partial p_t^N)(\partial p_t^N / \partial c_t^T)$ . The second term in the FOC above is not present in the decentralized economy, i.e., private agents undervalue wealth.

The constrained optimum can be implemented in one of two ways.

- By charging a state-contingent tax  $\tau_t$  on debt issues at time  $t$ , a wedge is introduced into the private Euler equation for bonds. The optimal tax is found by comparing FOCs. Appendix B shows that in a richer model the tax can also be implemented indirectly by regulating reserve or capital requirements for banks that intermediate loans.

- The planner can also impose a margin requirement  $\theta_t$  such that the credit constraint becomes  $b_{t+1} \geq -(1 - \theta_t)(\kappa^N p_t^N y_t^N + \kappa^T y_t^T)$ . The socially optimal amount of borrowing  $b_t^{sp}$  is attained by choosing  $\theta_t$  so that the constraint reduces to  $b_{t+1} \geq b_t^{sp}$ .

**Simulations** The model is calibrated to Argentinian data. The endowment process is an approximate AR(1) process for tradable and nontradable output. The intratemporal elasticity of substitution  $1/(\eta + 1)$  is set to the upper bound used in the literature, for conservativeness (smallest price movements). The baseline has  $\kappa^N = \kappa^T$  and calibrates the value to match the frequency of Sudden Stops in cross-country data.

The borrowing policy functions are plotted in Figure 1 as a function of current debt. There are three main regions: (1) the borrowing constraint binds, (2) the borrowing constraint doesn't bind but may possibly bind in the next period (warranting a positive tax in the second-best), and (3) the borrowing constraint can't bind next period. In the first region, the decentralized policy coincides with the constrained optimal one, as both are constrained. In the second region, the optimal policy takes on less debt due to precautionary saving. In the third region, the difference is small. The economy spends about 80% of the time in the middle region. As a result, the decentralized economy has a much larger left tail of bond holdings, although the average debt/GDP ratio is similar.

The average implied tax on debt in the constrained optimum is 5%. It goes to zero in the third region mentioned above, where there is no need to constrain borrowing. Margins are tightened by 9% on average.

A crisis is defined as an event in which the borrowing constraint binds and net capital outflows exceed one standard deviation in the decentralized ergodic distribution. By calibration, the decentralized crisis probability is 5.5%, while it is only 0.4% for the social planner. The magnitude of crises is also much larger in the decentralized economy: C drops 60% more, the CA/Y ratio increases (unlike for the social planner) and the RER depreciates way more. The decentralized economy is more volatile, and unlike in the social planner economy, the CA is strongly counter-cyclical. However, the welfare gain from internalizing the pecuniary externality is tiny, measured as the constant percentage increase in consumption necessary to compensate for the presence of the externality. This is because the externality has no supply-side effects and risk is only aggregate.

The author also shows that an optimal simple time-invariant tax of 3.6% achieves 62% of the welfare gains of the optimal state-contingent tax. However, a fixed margin constraint hurts welfare since it further tightens the constraint in bad times.

## 6.2 Korinek (2011): “The New Economics of Prudential Capital Controls: A Research Agenda”

**Summary** The paper outlines the main messages of recent research into the effects of prudential capital controls on the macroeconomy through regulation of pecuniary externalities. A simple two-period model is used to illustrate financial amplification and Pareto suboptimality arising from a balance sheet externality. The central result is that when the borrowing constraint binds, the social value of liquidity exceeds the private value. From this insight one can derive that the unregulated economy features overborrowing, excessive risk-taking and excessive use of short-term debt. The author argues that these conclusions call for the use of prudential capital controls in practical policy considerations.

**Intuition** The debt-deflation feedback loop has three components.

1. *Falling prices.* The exchange rate may depreciate due to expenditure switching, reduction in money demand, interest rate reductions, lower tax revenues leading to greater need for seignorage, etc. Asset prices fall either if the constraint reduces the value of future payoffs relative to today's or if agents have to cut back on investment and adjustment is costly (q-theory).
2. *Balance sheet effects.* As prices fall, the value of collateral declines, reducing agents' access to credit.
3. *Falling aggregate demand.* When access to credit tightens, agents are forced to cut back consumption.

The total effect is that capital flows become procyclical, i.e., credit tightens exactly when it is needed the most.

**Model** SO endowment economy, two periods, two goods. In the second period, only the tradable good exists. Utility

$$U = \log c_1 + c_{T,2}, \quad c_1 = (c_{T,1})^\sigma (c_{N,1})^{1-\sigma}.$$

Endowments are normalized to  $y_{T,1} = \sigma, y_{N,1} = 1 - \sigma$ . World gross interest rate is 1. The representative consumer inherits a tradable debt of  $d_0$ , such that the first-period liquidity is  $m = y_{T,1} - d_0$ . The consumer chooses first-period debt  $d_1$  subject to budget constraints and the financial constraint

$$d_1 \leq \kappa(y_{T,1} + py_{N,1}).$$

The intratemporal FOC gives

$$p = \frac{c_{T,1}}{\sigma},$$

where  $p$  is the relative price of nontradables.

If  $d_0 \leq \kappa$ , the equilibrium has  $d_1 = d_0, c_{T,1} = y_{T,1} = \sigma$  and  $p = 1$ . The derivative of the value function with respect to liquidity is  $\partial V / \partial m = 1$ .

If, however,  $d_0 > \kappa$ , the financial constraint binds. The level of borrowing is determined by the constraint, and  $\partial V / \partial m = \sigma / c_{T,1} > 1$ . Intuitively, higher liquid wealth relaxes the borrowing constraint, leading to higher tradable consumption, which leads to higher  $p$ , which further relaxes the borrowing constraint, etc.

The author discusses the nature of the pecuniary externality. With incomplete markets, since ratios of marginal utilities are not equalized, a Pareto improvement can be effectuated by changing resources relative to the private optimum (at a second-order cost) and incurring first-order gains by redistributing to agents with higher marginal valuations. This is shown in the model by evaluating the derivative of aggregate welfare with respect to liquidity  $m$ . Since the social planner internalizes the indirect effect of  $m$  on  $p$ , we get  $\partial V^{sp} / \partial m > \partial V / \partial m$  when the constraint binds. "A healthy balance sheet is a public good."

The above result is used to illustrate three pernicious effects of pecuniary externalities.

- *Overborrowing.* Suppose the representative consumer also makes a choice in period 0 of how much initial debt  $d_0$  to take on. Only tradable consumption is available in period 0, and the

endowment is 0 so all consumption comes through borrowing. The utility function is changed to  $U = \log(c_{T,0}^\sigma) + \log c_1 + c_{T,2}$ . The FOC is

$$\frac{\sigma}{d_0} = \frac{\partial V(y_{T,1} - d_0)}{\partial m},$$

where  $V$  is the starting-in-period-1 value function from above. Since  $V$  is concave, the above result gives that when the first-period constraint binds, the social planner chooses a smaller  $d_0$  than in the decentralized outcome. The planner smooths aggregate consumption more since he internalizes the full effect of liquidity on later borrowing capability. The second-best allocation can be implemented by imposing a tax  $\tau$  on debt inflows  $d_0$ . This multiplies the LHS above by  $(1 - \tau)$ . Alternatively, the optimal allocation can be implemented by requiring borrowers to hold an unremunerated reserve (the opportunity cost is then essentially a Pigouvian tax).

- *Excessive risk-taking.* Suppose there are two states of the world in period 1, and the consumer can trade state-contingent claims with a risk-averse foreign investor in period 0. Because the planner values liquidity more highly in a period-1 state in which the constraint binds, he promises smaller repayments in such states. Hence, the second best features less amplification than in the decentralized outcome. In richer models, a pecking order of financing is introduced. Foreign-denominated debt is most risky due to adverse RER movements during bad times. CPI-indexed debt is less risky. Stock market indexed debt (e.g., FDI debt) is least risky since repayments are procyclical.
- *Excessive short-term debt.* Suppose everything is deterministic, but now there are two financing vehicles available in period 0: a short-term and a long-term bond. The long-term bond carries a premium but pays off in period 2 and does not need to be rolled over at time 1. Because the planner values liquidity in period 1 higher, he takes on more long-term relative to short-term debt than the private outcome. This leads to a smoother consumption profile.

The author briefly discusses how heterogeneity of agents affect the results. If recipients of capital inflows differ mostly in profitability, price controls are preferred to quantity controls since this allows for more flexibility in resource allocation. If recipients of capital inflows differ mostly in riskiness, quantity controls are preferable.

## 7 Sovereign Debt, Debt Crises

### 7.1 Aguiar and Gopinath (2006): “Defaultable debt, interest rates and the current account”

**Summary** In standard SOE models that allow the country to default on its non-contingent bonds, it is hard to generate realistic default rates. With deterministic trend growth, the threat of autarky is minor, since on the one hand unit root  $Y$  shocks limit the usefulness of consumption smoothing, while on the other hand, i.i.d.  $Y$  shocks do not impact life-time wealth much. Consequently, the default decision depends mostly on current debt (not the state), making the supply curve of debt very steep, which causes agents to not take on too much debt. If, however, the trend of  $Y$  growth is stochastic and subject to shocks, the current state influences the default decision to a greater extent, flattening the supply curve and allowing agents to rationally take on more debt. Introducing

trend shocks also helps make both interest rates and CA countercyclical, since (1) the interest rate schedule as a function of debt is flatter and (2) the probability of default drops more markedly in good times, causing a favorable shift in interest rates.

**Motivation** Unlike in developed SOEs, the interest rate and CA is countercyclical in emerging markets. We also observe a fairly high rate of default.

**Model** SO endowment, bonds-only economy. Representative agent has CRRA preferences. Endowment is  $y_t = e^{z_t}\Gamma_t$ , where  $z_t$  is an AR(1) process and  $\Gamma_t = g_t\Gamma_{t-1}$ , with  $\log g_t$  following a different AR(1) process. NFA are denoted  $a_t$ . Upon default,  $a_t$  is set to 0 but the economy faces a proportional output loss of  $1 - \delta$  and is forced into autarky. If in default, the economy is redeemed with probability  $\lambda$ .  $V^G(a_t, z_t, \Gamma_t)$  is the value function if in non-default and  $V^B(z_t, \Gamma_t)$  is in default. The Bellman recursions are then

$$V^B(z_t, \Gamma_t) = u((1 - \delta)y_t) + \lambda\beta E_t V(0, z_{t+1}, \Gamma_{t+1}) + (1 - \lambda)\beta E_t V^B(z_{t+1}, \Gamma_{t+1}),$$

$$V^G(a_t, z_t, \Gamma_t) = \max_{c_t = y_t + a_t - q_t a_{t+1}} u(c_t) + \beta E_t V(a_{t+1}, z_{t+1}, \Gamma_{t+1}),$$

$$V(a_t, z_t, \Gamma_t) = \max\{V^G(a_t, z_t, \Gamma_t), V^B(z_t, \Gamma_t)\}.$$

The economy internalizes the effect of  $a_{t+1}$  on the bond price  $q_t$ . International investors are risk neutral and require expected return  $r^*$ . Hence,

$$q(a_{t+1}, z_t, \Gamma_t) = \frac{E_t(1 - D_{t+1})}{1 + r^*},$$

where the default function  $D(a_t, z_t, \Gamma_t)$  is 1 if  $V_t^B > V_t^G$  and 0 otherwise.

**Results** Two calibrated models are analyzed. The calibrations are meant to match key moments for Argentina. In Model I, the trend is deterministic, so  $g_t \equiv \mu_g$ . In Model II, the trend  $\Gamma_t$  is stochastic and they set  $z_t \equiv 1$  for simplicity. The model is solved numerically by exploiting homogeneity of the value, default and price functions, such that everything can be written in terms of detrended variables.<sup>4</sup>

For given  $z$ , it is clear that there exists a unique cut-off  $a_t$  for which  $V^G$  equals  $V^B$ , since the former depends positively on  $a_t$  while the latter does not depend on  $a_t$ .

It is not so clear how the default decision depends on  $z_t$  for given  $a_t$ . When  $\lambda = 0$ ,  $V^G$  should be at least as steep as  $V^B$  at the indifference point. This is because the continuation value for an agent with good credit standing is always higher than for one with bad standing (since for  $\lambda = 0$ , the good agent can mimic the bad one), so if an agent is indifferent between defaulting or not, it must be that the current consumption absent default is less than that under default, implying higher current MU in the non-default case. Suppose then that the agent is offered an additional unit of endowment. Under autarky this unit must be consumed. If a non-defaulting agent consumed the unit, her utility would increase more due to the higher MU, and in fact she can even choose to save some of the extra endowment if it makes her better off. The conclusion follows. For  $\lambda > 0$ , the argument doesn't apply, however.

---

<sup>4</sup>All the following expressions involve these detrended variables.

For Model I, the default region features a very steep slope along the  $a_t$  axis for the intuitive reasons mentioned earlier. The bond supply curve is therefore also very steep:

$$q_t(a_{t+1}) = \frac{1 - \Pr(z_{t+1} < \bar{z}(a_{t+1})|z_t)}{1 + r^*}.$$

Because the economy internalizes the steepness of the supply schedule, a higher desire to borrow doesn't result in more debt. The default rate is consequently very low. Due to minimal borrowing, NX are very stable. The calibrated income process is very persistent, yielding a countercyclical CA. However, the interest rate is procyclical, counterfactually. The reason is the following. In good times agents therefore want to borrow more, but due to the steepness of the bond supply schedule, interest rates will spike (however, the slope of the interest schedule turns out to be countercyclical, which allows the CA to also be countercyclical). There is also a second effect, namely that in good times default becomes less likely, so the interest rate schedule shifts favorably. However, since shocks are transitory, the latter effect is small compared to the former.

The authors perform a Lucas (1985) calculation showing why it is difficult to sustain a realistic level of debt. They stack the cards against autarky by assuming that shocks are i.i.d., autarky lasts forever and that financial integration implies perfectly smooth consumption at the cost of interest payments  $rB$  per period. The  $Y$  process is  $Y_t = \bar{Y}e^{z_t}e^{-\sigma_z^2/2}$ ,  $z_t \sim \mathcal{N}(0, \sigma_z^2)$ . Then

$$V^B = E \sum_t \beta^t \frac{Y_t^{1-\gamma}}{1-\gamma}$$

and

$$V^G = \sum_t \beta^t \frac{(\bar{Y} - rB)^{1-\gamma}}{1-\gamma}$$

can be calculated and compared. Without default penalty  $\delta$  and for realistic  $\sigma_z^2$  and  $\gamma$ , maximum sustainable debt payments as a percentage of  $Y$  are tiny. Imposing  $\delta = 0.02$  make debt payments of 20% of  $Y$  sustainable.

Model II generates much better results. The reason is that shocks to the growth rate imply permanent effects on  $Y$ , so access to capital markets is very beneficial in good times. Because  $V^G - V^B$  depends more elastically on the state than in Model I, the default region has a less sharp slope, leading to a less steep bond price function, more borrowing in equilibrium and more default. The CA and interest rate are both countercyclical. The latter is obtained since introducing trend shocks makes both the two previously mentioned effects of the state on the interest rate go in the right direction. NX and interest rates are more volatile than in Model I, although the latter is still more muted than in the data.

While Model II performs better, the default rate is still too low. A quick fix is provided by introducing third-party bailouts, i.e., debt guarantees up to some limit  $a^*$  in the form of transfers to the international creditors (from an unmodeled source). The price of debt is then

$$q_t = \frac{\min\{1, a^*/a_t\} + E_t(1 - D_{t+1}) \max\{1 - a^*/a_t, 0\}}{1 + r^*},$$

i.e., the probability of default only discounts the un-guaranteed fraction of the debt. Bailouts essentially subsidize defaults, so it is possible to get the default rate into the right territory with stochastic trend shocks, although interest rate volatility remains subdued.

## 7.2 Broner, Martin and Ventura (2010): “Sovereign Risk and Secondary Markets”

**Summary** The conventional wisdom is that in the absence of default penalties, sovereign risk eliminates the possibility for international trade. The authors show that this is not the case if debt is traded in secondary markets. If the Home country is in debt, the government wishes to not enforce contracts. However, infinitesimal domestic agents privately have an incentive to buy the Home country debt back from foreign creditors at a price below or at face value, while foreign creditors are willing to sell at any positive price if they fear non-enforcement. Because contracts are then sold back to Home agents, the Home government now wishes to enforce the contracts so as to not unequally redistribute wealth. All bonds end up in the hands of those that are capable of redeeming them. The authors show that this intuition holds in a wide class of models, as long as (1) there are no impediments to trade in secondary markets and (2) governments only decide whether to enforce contracts and they can’t commit in advance to do so. If these assumptions are not met, sovereign risk does influence foreign asset trade but in a surprisingly rich way.

**Model** The simplest model is called Debtor-Creditor World (D-C World). Two periods, Today and Tomorrow. There are two countries, Debtor and Creditor, with infinitesimal agents. The utility function is additively separable in Today and Tomorrow consumption with no discounting. All debtors (inhabitants of Debtor) receive endowment  $y - \varepsilon$  Today and  $y + \varepsilon$  Tomorrow. All creditors (inhabitants of Creditor) receive  $y + \varepsilon$  Today and  $y - \varepsilon$  tomorrow. Since endowments differ across regions, there are gains from trade.

The bond market that opens Today is called the primary market, while the one that opens Tomorrow is the secondary market. There is a government in each country. If enforcement of contracts were full, bonds would trade at price 1 and resources would be distributed such that every agent consumed  $y$  in both periods.

Suppose instead governments strategically choose which bond payments to enforce *after* secondary markets have opened, and they care about the average (Tomorrow) utility of their domestic agents. If there were no secondary market, the Debtor government would always choose not to enforce the repayments to the creditors, so in equilibrium there would be no trade. However, with secondary markets operational, it is possible to achieve the same Pareto efficient outcome as in the full-enforcement case, with the same prices: Suppose creditors sell their claims back to debtors at face value in the secondary market, but these claims are distributed unequally among the debtors. If secondary market contracts are enforced, all debtors (and creditors) consume  $y$  Tomorrow. If contracts are not enforced, some debtors consume less than  $y$  (because they were holding more claims) and others more than  $y$ , which is strictly worse for the Debtor government due to Jensen’s inequality. Hence, it chooses to enforce the contracts and the equilibrium outcome is the same as in the full-enforcement case, except that the necessary trading volume is higher (under full enforcement, no secondary market trade is necessary to achieve the optimum).

Another way to think about the above is that secondary markets create a prisoner’s dilemma situation among the debtors. They would collectively like to default on their aggregate debt so as to increase their total consumption, i.e., they would like to all agree on not buying back each other’s debt from the creditors. This is not an equilibrium, however, because creditors would be willing to sell claims at any positive price, so each debtor has an enormous private incentive to buy back claims than can be redeemed at face value.

In the above world, there also exists a pessimistic equilibrium, where the Debtor government

isn't expected to enforce any contracts, so no bonds are issued Today or Tomorrow and the autarky outcome prevails. This equilibrium is eliminated if we assume that there is a tiny welfare cost to not enforcing contracts.

It is possible to slightly alter the assumptions to obtain models in which the strategic enforcement outcome is a Pareto improvement upon the full-enforcement outcome. For example, in a world with more than one state of nature and if insurance markets are missing so only non-contingent bonds are available, introducing the possibility of selective default can help complete markets.

The authors show that the above conclusions also hold in a much richer model with many regions, many periods, many shocks, market incompleteness and extensive heterogeneity. The argument basically only needs two ingredients: (1) if the government is expected to enforce payments within a group of agents, individual maximization leads to agents within this group buying back their own maturing assets, and (2) if each group buys back their own assets (so there are no outstanding claims left across groups), enforcement of contracts within each group is consistent with government maximization, by Jensen's inequality.

The above results did, however, rest on the assumptions that the secondary market is well functioning, i.e., there are no trading frictions and individuals are infinitesimal, and the government only chooses whether to enforce contracts or not, without having the option of committing in advance.

- If transaction costs are introduced, the additional secondary-market trading necessary in the strategic enforcement case entails that consumption smoothing is hindered more than under full enforcement.
- If there is a single large debtor (or debtors can fully coordinate), financial trade completely collapses. The debtor will internalize the prisoner's dilemma and refuse to trade on the secondary market Tomorrow so that default ensues. However, this breaks down trade. If there are multiple debtors but one is large, consumption smoothing partially breaks down. Coordination among *creditors* doesn't help them: they never receive payment directly from debtors and so can't exploit their market power.
- Suppose governments can commit to enforcing or not enforcing certain contracts before the secondary market opens. Suppose also the basic D-C World is extended to have two Tomorrow states, with one group of debtors being lucky (in terms of endowment) in one state and unlucky in the other, and another being unlucky in the first and lucky in the second. Enforcement of Debtor contracts must be nondiscriminatory in equilibrium since creditors can always sell their claim at face value to a debtor and de facto receive their payment. Hence, the Debtor government can either enforce all payments or shut down all. Due to the *ex post* heterogeneity between debtors, there is then a tradeoff after primary market trade: committing not to enforce avoids payments to creditors, while committing to enforce improves the distribution of consumption among debtors.
- If governments have alternative policy instruments, they will use those to fight the reallocation of contracts on the secondary market. Suppose for example there are two types of goods and utility depends on an aggregator of those two goods. Suppose the Debtor government can impose capital controls on transactions but can't discriminate between different types of transactions. Then there's a tradeoff between thwarting payments to creditors, on the one hand, and facilitating trade in the two goods to exploit gains from trade, on the other hand.

### 7.3 Bulow and Rogoff (1989): “Sovereign Debt: Is to Forgive to Forget?”

**Summary** The threat of loss of reputation in international borrowing arrangements is not enough to deter a country from defaulting on its debt. As long as it still has cash-in-advance contracts (i.e., a contract where the country pays a positive amount up front for a state-contingent, non-negative payment next period) available after defaulting, it is always possible, from some time period on, to increase consumption in all subsequent periods by defaulting and initiating a sequence of feasible cash-in-advance contracts. Hence, non-reputational costs of default are needed to justify the existence of sovereign debt.

**Model** The borrower is a small country that can’t affect the world interest rate  $r$  (constant over time). Foreign investors are risk neutral. The country production function is  $Y_t = f(\vec{\theta}_t, \vec{I}_{t-1})$ , where  $\vec{\theta}_t = (\theta_t, \theta_{t-1}, \dots)$  is the history of exogenous shocks and  $\vec{I}_{t-1} = (I_{t-1}, I_{t-2}, \dots)$  is the history of investment decisions. Both these are fully observable by everyone.

A reputation contract (implicitly) specifies state-contingent payments  $P_t(\vec{\theta}_t)$ . The present value of future repayments is

$$D_t(\vec{\theta}_t) = E_t \sum_{s=t}^{\infty} \frac{P_s}{(1+r)^{s-t}}.$$

It is assumed that if a country defaults, it still has access to *cash-in-advance contracts* that are indexed to the same variables as the reputation payments: The country pays  $A_t$  in period  $t$ , and in period  $t+1$  it receives the state-contingent payoff  $G_{t+1}(\vec{\theta}_{t+1})$ . Foreign investors accept such a contract if they break even,

$$(1+r)A_t = E_t[G_{t+1}(\vec{\theta}_{t+1})],$$

and no further future payments are needed from the borrower,

$$G_{t+1}(\vec{\theta}_{t+1}) \geq 0 \quad \forall \vec{\theta}_{t+1}.$$

Under the assumption that the present value of the country’s output (net of investment) is finite, Theorem 1 shows that in any equilibrium,  $D_t \leq 0$  for all  $t$ . Hence, default must occur in at least some state. The proof goes as follows:

- Find a period  $s$  with relatively high  $D_s$ .
- Default on the debt and initiate a series of carefully specified, feasible cash-in-advance contracts.
- The resulting payments are uniformly smaller in every period  $t \geq s$ . Equality only holds if  $D_t \leq 0$ .

Hence, the proof is an arbitrage-type argument and does not rely on the form of the utility function.

The authors then consider a model in which lenders can impose a random output penalty of  $\pi_t(\vec{\theta}_t, \eta_t) < Y_t - I_t$  on a defaulting borrower, where  $\eta_t$  is independent of  $\theta$ . Define

$$\Pi_t = E_t \sum_{s=t}^{\infty} \frac{\pi_s}{(1+r)^{s-t}}.$$

Theorem 2 shows that any equilibrium must have  $D_t - \Pi_t \leq 0$ , i.e., the debt obligations are fully supported by non-reputational sanctions.

In the conclusion, the authors argue that, while the existence of private information invalidates their assumptions and thus their results, it should also limit the extent of reputation contracts available. Hence, research into non-reputational penalties is warranted.

#### 7.4 Calvo (1988): “Servicing the Public Debt: The Role of Expectations”

**Summary** If public debt may be partially repudiated, e.g., by monetization, the temptation to repudiate is increasing in the burden of the debt, while the interest rate is increasing in the likelihood of repudiation. Hence, multiple equilibria naturally emerge: Either the interest rate and repudiation are low, or the interest rate and repudiation are high.

**Model** Two periods, two types of agents: competitive consumers and the government. In period 0, government borrows  $b$  at gross interest rate  $R_b$ . Government can repudiate a fraction  $\theta \in [0, 1]$  of the debt. Consumers have perfect foresight and break even, i.e.,  $(1 - \theta)R_b = R$ , where  $R$  is the return on physical capital. The government’s taxation need  $x$  is

$$x = (1 - \theta)bR_b + g + \alpha\theta bR_b,$$

where  $\alpha \in [0, 1]$  is the proportional cost of repudiation (e.g., transaction costs). Government purchases  $g$  are exogenous. The more the government taxes, the less debt it has to repudiate.

Individuals consume all their resources in period 1, giving consumption

$$c = y - z(x) + kR + (1 - \theta)bR_b - x,$$

with  $z(x)$  representing the deadweight loss from taxation. The government is benevolent, and so in period 1 it chooses  $x$  to maximize  $c$  subject to its budget constraint.  $R_b$  is taken as given in period 1 (no commitment). Because  $\theta \in [0, 1]$ , this is equivalent to minimizing

$$z(x) - \frac{\alpha}{1 - \alpha}x$$

subject to

$$g + \alpha bR_b \leq x \leq g + bR_b.$$

Let  $x^*$  be the unconstrained optimum. The government’s reaction function (i.e., the optimal taxation  $x$  as a function of  $R_b$ ) is depicted in Figure 1. For low values of  $R_b$ , it is optimal not to repudiate  $\theta = 0$ , so  $x = g + bR_b$ . For high values of  $R_b$ , full repudiation is optimal and  $x = g + \alpha bR_b$ . For intermediate values of  $R_b$ , partial repudiation and  $x = x^*$  are optimal.

Substituting the perfect-foresight break-even condition into the government’s budget constraint, we get the consistency condition

$$x = g + (1 - \alpha)bR + \alpha bR_b.$$

Equilibria are now determined as the intersections of the government’s reaction curve and the above consistency condition in  $(R_b, x)$  space. See Figure 2.

There are three cases.

1. If  $x^* > g + bR$ , there are two equilibria. One has no repudiation,  $R_b = R$  and  $x = g + bR$ . The other has partial repudiation,  $R_b > R$  and  $x = x^*$ . Substituting the perfect-foresight break-even condition into the expression for period-1 consumption, we get

$$c = y - z(x) + (k + b)R - x,$$

so the equilibria can be Pareto-ranked based on  $x$ . The first equilibrium is thus better than the second.

2. If  $x^* < g + bR$ , no equilibrium exists. The intuition is clearest for  $\alpha = 0$ : A benevolent government would repudiate all debt, but this can't be an equilibrium in which consumers break even.
3. If  $x^* = g + bR$ , there is a unique equilibrium with  $R_b = R$  and no repudiation.

Comparative statics:

- The higher outstanding debt  $b$  is, the *smaller* is the interest rate in the bad equilibrium.
- The higher repudiation costs  $\alpha$  are, the *larger* is the consumption cost in the bad equilibrium.

Calvo also develops an alternative model in which the government chooses the price level  $P_t$  (and thus inflation  $\pi$ ). The share of debt that is not repudiated can then be thought of as  $1 - \theta = P_0/P_1$ . Hence,  $\theta = \pi/(1 + \pi)$ . Inflation is assumed not to be costly for the government but costly to consumers. Furthermore, the government collects an inflation tax. Even though  $\theta$  is now only constrained to lie in  $(-\infty, 1]$ , the previous conclusion goes through: Under the right conditions, multiple equilibria exist.

## 7.5 Cole and Kehoe (2000): “Self-Fulfilling Debt Crises”

**Summary** Analytical model in which fundamentals determine whether a crisis *can* occur, while market expectations (sunspot shocks) then determine whether the crisis actually occurs. Causality runs from expectations to interest rates. The government can't commit *ex ante* to repay its debt, so investors' beliefs about repayment matter for the government's ability to service the debt. In the equilibrium class that the authors study, it is often optimal for the government to forego some consumption smoothing and gradually run down its debt so as to exit the fundamental-determined “crisis zone.” Exiting the crisis zone causes interest rates to drop and private investment to boom. Extending the maturity of outstanding debt reduces the crisis zone, since smaller per-period debt payments limit the incentive of the government to default.

**Motivation** During the Mexican crisis of 1994–1995, fear of default led to the inability of the government to service its debt. This was exacerbated by the short maturity structure of the outstanding debt.

**Model** There is a single good, which can either be consumed or saved as capital. Three types of agents: consumers, international investors and the government. The first two are risk neutral in private consumption with discount factor  $\beta$ . Consumers also derive additive concave utility from government consumption and their budget constraint is

$$c_t + k_{t+1} \leq (1 - \theta)a_t f(k_t).$$

The government's objective is to choose borrowing  $B_{t+1}$ , the repayment decision  $z_t \in \{0, 1\}$  and public consumption  $g_t$  to maximize consumers' utility. Its budget constraint is

$$g_t + z_t B_t \leq a_t \theta f(k_t) + q_t B_{t+1}.$$

If the government chooses to default,  $z_t = 0$ , productivity drops from  $a_t = 1$  to  $a_t = \alpha < 1$  forever, and the government loses all access to international borrowing and lending.

The timing is such that first a sunspot variable  $\zeta_t$  is realized and becomes part of the state vector  $s_t$ , then the government chooses  $B_{t+1}$  given the price schedule  $q_t = q(s_t, B_{t+1})$ , then international investors choose how much debt to supply given price  $q_t$ , then government chooses whether to default, and finally consumers choose consumption and investment given  $a_t$ . Hence, the timing is such that the government must roll over old debt into new debt, which opens the possibility for a self-fulfilling crisis.

A crisis is one in which the government is unable to sell new debt at a positive price  $q_t$ . The authors focus on a class of equilibria that are Markov in the state vector (i.e., not time-dependent) and for which the probability  $\pi_t$  of a crisis (given that the fundamentals allow for a crisis occurring) is constant over time,  $\pi_t \equiv \pi$ . The region for fundamentals (debt and capital levels) in which a crisis occurs with probability  $\pi \in (0, 1)$  is called the *crisis zone*. The sunspot shock  $\zeta_t$ , which is uniform on  $(0, 1)$ , determines whether a crisis actually occurs in the crisis zone. The *no-crisis* zone is the region in which it never pays for the government to default, regardless of fundamentals or sunspots. In the *default zone*, the government always defaults.

Given risk neutrality, the international investors demand a price  $q_t = \beta(1 - \pi)$  in the crisis zone,  $q_t = \beta$  in the no-crisis zone and  $q_t = 0$  in the default zone.

Consumers are risk neutral and so behave in a rather mechanical way. Their optimal investment plan depends solely on the value of the productivity parameter in the next period. Depending on whether they ascribe probability 0,  $\pi$  or 1 to it being  $\alpha$ , they choose one of three (stationary) capital levels.

The three zones can now be characterized in terms of the government's value function. Let  $V_g^n(s, B', q)$  and  $V_g^d(s, B', q)$  be the values of not defaulting and defaulting, resp., given state  $s = (B, K, a = 1, \zeta)$ , after it has sold new debt  $B'$  at price  $q$ . In any equilibrium with positive lending, the participation constraint

$$V_g^n(s, B', q) \geq V_g^d(s, B', q)$$

must hold. If we set  $q = \beta(1 - \pi)$ , the constraint determines the upper bound  $\bar{B}(K, \pi)$  on debt for which positive lending can occur in a sunspot equilibrium of the above-mentioned type. Furthermore, for a crisis to be able to occur, the government must satisfy a no-lending condition

$$V_g^n(s, 0, 0) < V_g^d(s, 0, 0),$$

i.e., that if  $q = 0$ , then the government prefers to default. This condition determines the upper bound  $\bar{b}(K)$  on debt at which the government prefers to repay even if  $q = 0$  and a crisis can't occur. The crisis zone is then given by  $(\bar{b}(K), \bar{B}(K, \pi)]$ .

It is possible to show that in the crisis zone, the government either chooses to default now, to run the debt down to  $\bar{b}(k^n)$  over  $T$  periods, or to keep the debt constant and not run it down. Here  $k^n$  is the level of capital if consumers expect default with probability 0. The value function for each of these cases (and for each  $T$ ) can be calculated, leading to a characterization of optimal policy. Intuitively, the government is motivated to forego some consumption smoothing and instead

gradually run down the level of debt, because as it exits the crisis zone, the interest rate  $1/q_t$  on debt decreases, making it easier to finance future government consumption. Furthermore, as the country exits the crisis zone, private investment booms since the risk of a productivity decline is eliminated. As the equilibrium parameter  $\pi$  increases, the motivation to exit the crisis zone becomes greater.

The authors draw a number of conclusions from their model.

- A policy that enhances the credibility of the government by increasing the cost of default may be interpreted as a drop in  $\alpha$ . This indeed raises the level of debt  $\bar{B}$  that can be sustained if a crisis doesn't occur, and it also raises the level of debt  $\bar{b}$  at which a crisis is possible. The gap between these two need not be removed, though, so the crisis zone may remain.
- If the maturity of the bonds is increased, the amount of debt that needs to be rolled over every period is smaller, decreasing the incentive to default. For sufficiently high maturities, the crisis zone is eliminated.
- The more the government cares about private consumption relative to government consumption, the higher is the cost of default, since private consumption really suffers. The government can then satisfy the participation constraint with a higher level of debt. But this high level of debt makes a crisis possible, widening the crisis zone.

## 7.6 Eaton and Gersovitz (1981): “Debt with Potential Repudiation”

**Summary** Sovereign debt is characterized by the absence of an explicit mechanism to enforce contracts. A dynamic model is laid out in which the borrower may choose to default at the (endogenous) cost of being excluded from future borrowing. Only non-state-contingent, one-period debt is considered. Lenders are risk neutral and competitive. Under general conditions, it is shown that the credit supply curve is upward-sloping and the equilibrium level of credit is the minimum of the demand for credit and a credit ceiling imposed by the lenders. Hence, credit rationing may occur. The model is specialized to one in which national income can take one of two values. While the demand for credit is always increasing in the growth rate and variability of output, the borrowing limit depends ambiguously on these (if output is stochastic). The implications of the model are tested with an econometric model that exploits the minimum condition on realized borrowing.

**Model** A country produces output  $y_t$ . Consumption is  $c_t = y_t + b_t - p_t$ , where  $b_t \in B_t$  (the feasible set of debts) and  $p_t$  are debt service payments. The borrowing country maximizes the sum of future discounted utility. Debt matures in one period and the repayment due is denoted  $d_{t+1} = R(b_t)$ . If  $d_t < p_t$ , the country is cut off from future borrowing, i.e.,  $B_\tau = \{0\}$  for  $\tau \geq t$ . The probability of default is  $\lambda_t = \Pr(V_t^D > V_t^R)$ , where  $V_t^D$  is the continuation value of autarky and  $V_t^R$  is the value function given a decision not to default,

$$V^R(y_t, d_t) = \sup_{b_t \in B_t} U(y_t + b_t - d_t) + \beta E[\max\{V^D(y_{t+1}), V^R(y_{t+1}, d_{t+1})\}].$$

The probability of default is shown to be increasing in the debt service obligations  $d_t$ .

Lenders are competitive and risk neutral, so they break even in expectation:

$$[1 - \lambda(R(b_t))]R(b_t) = (1 + \bar{r})b_t,$$

where  $\bar{r}$  is the safe world interest rate. From this it can be shown that the set of feasible loans has the form  $B_t = [0, \bar{b}_t]$  for some credit ceiling, and  $R(b_t)$  is increasing and convex, i.e., the supply of credit slopes upwards. If the ceiling is exceeded by the amount  $b_t^*$  demanded by the country, it can be shown that there is credit rationing and actual borrowing is  $b_t = \min\{b_t^*, \bar{b}_t\}$ . In credit markets, prices can't adjust to clear markets since the price won't be paid if the borrower defaults. Hence, some other margin must adjust and credit rationing occurs.

The model is specialized in two ways. First, a deterministic model is considered in which output alternates between a high and a low value. The country borrows when output is low and saves when output is high (at potentially different interest rates). Trend output grows at gross rate  $G$ . Utility is CRRA. The amount of credit demanded is increasing in growth and volatility (i.e., the amount by which output deterministically changes from period to period) for the usual reasons. The authors also show that the credit ceiling is increasing in volatility, while  $G$  has an ambiguous effect on the ceiling. A higher interest rate on saving makes smoothing via borrowing less attractive, reducing the credit ceiling. Finally, if the lenders can impose an exogenous penalty on the country if it defaults, the credit ceiling rises.

The second specialized model is stochastic. For tractability, the authors assume no growth and no saving. In each period, output can take on one of two values with equal probability. Debt is non-contingent and so default is more attractive in bad times. In contrast with the deterministic case, the debt ceiling may either rise or fall with the volatility of output. Intuitively, higher volatility increases the value of future consumption smoothing but also raises the cost of debt service if output is low. If  $\beta$  is low, the latter effect may dominate.

The various conclusions from the theoretical models are summarized in Section 2.4.

**Empirics** The model is taken to the data by postulating

$$b_i^* = X_i' \alpha_1 + u_i, \quad \bar{b}_i = X_i' \alpha_2 + v_i,$$

with  $u_i$  and  $v_i$  normally distributed and independent. Observed debt is then given by  $b_i = \min\{b_i^*, \bar{b}_i\}$ , as in the model. The data is a cross section of developing countries. The model is estimated by ML. It is found that the growth rate of GNP does indeed raise desired debt but has a near-zero effect on the ceiling.

If an additional penalty of default were exclusion from international trade, this penalty should be increasing in the volatility of exports. Both  $b^*$  and  $\bar{b}$  rise with the volatility of exports, as predicted by the model.

While the credit ceiling depends significantly on GNP, as expected, the desired debt does not, which is surprising.

The probability model can be used to determine the number of countries in the sample that are more likely to be credit rationed than not. By this measure, 65 of 85 countries were constrained in 1970 and 1974.

## 7.7 Reinhart and Rogoff (2011): “From Financial Crisis to Debt Crash”

**Summary** Using an extensive cross-country data set that stretches back almost 200 years, the authors study correlations between debt, default and crises of various kinds. The main conclusions are:

- Debt (both public and private) surges non-linearly before a debt crisis hits.

- Banking crises tend to precede debt crises.
- Debt becomes more short-term before a debt crisis.
- Banking crises in international financial centers predict banking crises in other countries.
- International defaults come in waves and are often separated by many years.
- Currency and inflation crises go hand in hand.
- Periods of high indebtedness are associated with higher incidence of inflation crises.
- The median duration of default spells has been cut in half post-WW2. The frequency of banking crises has also dropped, but they occur as often in developed as in emerging markets.

## 7.8 Yue (2010): “Sovereign default and debt renegotiation”

**Summary** Following sovereign defaults creditors typically are able to recover a fraction of the value of the loans. Yue develops an SOE model with endogenous default decisions and Nash bargaining over debt restructuring following default. Recovery rates are higher in relatively good states, yielding useful insurance for the SOE. Interest rates increase with the level of debt because of higher default probability and lower recovery rates. In simulations, due to countercyclical interest rates (and trend shocks), the TB is also countercyclical. The model generates volatile consumption and about the right amount of interest spread volatility. The average spread is too low since international investors are risk neutral.

**Model** SOE with identical households that receive exogenous endowment  $y_t$ . International investors are risk neutral and require expected return  $r$ . Borrowing happens only through one-period debt with face value  $b'$  (amount to be repaid next period). Positive  $b$  signifies positive NFA. The bond price function is denoted  $q(b', y)$ .

A country’s credit history is denoted by a binary variable  $h$ . If  $h = 0$ , the country is in good standing,  $h = 1$  otherwise. A country with  $h = 0$  can choose to default, in which case it repays nothing in the current period but goes into bad credit standing. The debt is immediately restructured through Nash bargaining (see below) to a new level that is a fraction  $\alpha(b, y)$  of the original debt  $b$ . If  $h = 1$  and  $b < 0$  (unpaid debt arrears), the country has an unresolved default and suffers a proportional output loss  $\lambda y_t$ ,  $\lambda \in (0, 1)$ , as well as exclusion from capital markets.

The value function  $v(b, h, y)$  is characterized as follows. If  $h = 0$ , we have

$$v(b, 0, y) = \max\{v^r(b, 0, y), v^d(b, 0, y)\},$$

where the repayment value function is

$$v^r(b, 0, y) = \max_{b'} u(y + b - q(b', y)b') + \beta E[v(b', 0, y') | y],$$

and the default value function is

$$v^d(b, 0, y) = u(y) + \beta E[v(\alpha(b, y)b, 1, y') | y].$$

If  $h = 1$  and the country has only partially repaid its debt arrears,  $b < 0$ , the debt is rolled over at interest rate  $r$ . Hence, the value function is

$$v(b, 1, y) = \max_{b' \in [b, 0]} u \left( (1 - \lambda)y + b - \frac{b'}{1 + r} \right) + \beta E[v(b', 1, y') \mid y].$$

The model therefore allows for an endogenous time of exclusion from debt markets, despite the immediate restructuring. When all debt arrears are paid in full, the country regains good credit standing:

$$v(0, 1, y) = v(0, 0, y).$$

The debt renegotiation proceeds as follows. The threat point for the debtor country is eternal autarky. The threat point for international investors is no recovery. Let  $a$  denote the hypothetical recovery rate. The surplus for the country is then

$$\Delta^B(a; b, y) = u(y) + \beta E[v(ab, 1, y') \mid y] - v^{\text{aut}}(y),$$

where the last term is the continuation value of autarky. The surplus for the lenders is

$$\Delta^L(a; b, y) = -\frac{ab}{1 + r},$$

the present value of recovered debt. Let  $\theta$  be the bargaining power of the debtor country. The debt recovery rate  $\alpha(b, y)$  solves the Nash bargaining program

$$\max_{a \in [0, 1]} (\Delta^B)^{\theta} (\Delta^L)^{1 - \theta}$$

subject to  $\Delta^B, \Delta^L \geq 0$ . Since the recovery rate is state-contingent, it provides additional insurance for the defaulting country and helps complete markets.

Foreign lenders are competitive so the bond price schedule is  $q(b', y) = 1/(1 + r)$  if  $b' \geq 0$  and

$$q(b', y) = \frac{1 - p(b', y)}{1 + r} + \frac{p(b', y)\gamma(b', y)}{1 + r}$$

if  $b' < 0$ . Here  $p(b', y)$  is the expected probability of default, which is  $E[\mathbf{1}_{\{V^d(b', y') > V^r(b', y')\}} \mid y]$  in equilibrium.  $\gamma(b', y)$  is the expected discounted recovery rate conditional on default, which is

$$E[\alpha(b', y')/(1 + r) \mid y, V^d(b', y') > V^r(b', y')]$$

in equilibrium.

**Results** Yue first characterizes the model solution theoretically. Because  $a$  only enters into the bargaining problem through the product  $ab$  (i.e., given default, only the absolute amount of debt reduction matters to creditors and debtor), we can write  $\alpha(b, y) = \min\{\bar{b}(y)/b, 1\}$ . The recovery rate decreases with the amount of defaulted debt. The default region is increasing in the level of debt, and so is the default probability. Because default probability rises with debt and the recovery rate decreases with debt, the interest rate increases with debt.

The model is calibrated to the most recent Argentinian default. Output is subject to trend shocks as in Aguiar and Gopinath (2006).  $\beta$  and  $\theta$  are calibrated to match the average default

frequency since 1824 and a 27% recovery rate during the 2005 restructuring. The resulting discount factor and bargaining power parameter are both 0.72.

Quantitatively, the option to restructure is not so beneficial that countries always default when debt is positive. When a country is hit by a bad shock, the higher debt reduction implies a higher incentive to default, leading to higher interest rate spreads (also through the direct channel of shocks on default probability). Thus, the correlation between output and interest rates is negative as in the data. Consumption is more volatile than output due to countercyclical interest rates. These also imply a countercyclical TB (although trend shocks contribute here).

The model supports a 10% debt/GDP ratio, lower than in the data but better than Arellano (2008). Bond spreads are in the right ballpark in terms of volatility but are too low on average due to risk neutrality among investors. Debt recovery rates are negatively correlated with debt and thus with default probability. This is consistent with cross-country data. The average length of exclusion from financial markets following default is only 2 quarters, perhaps because debt restructuring is immediate.

The effect of the debtor's bargaining power  $\theta$  on the recovery rate is negative, as expected. If the recovery rate decreases, *ceteris paribus*, the interest rate spread increases, but this deters borrowing, lowering the default probability and thus lowering interest rates. The effect of  $\theta$  on the interest rate spread and the default probability is therefore non-monotone, so higher bargaining power does not necessarily translate into higher borrowing costs.

## 8 Real Exchange Rates and Purchasing Power Parity

### 8.1 Alessandria, Kaboski and Midrigan (2010): “Inventories, Lumpy Trade and Large Devaluations”

**Summary** At least in emerging markets, importers face non-negligible fixed transaction costs and delivery lags, creating economies of scale in the transaction technology. This leads importers to transact infrequently and hold substantial inventories. A model of the importer's inventory management problem is constructed and calibrated. Quantity adjustment frictions break the link between a good's replacement cost and its marginal valuation, and the relationship is non-monotonic in the size of current inventories. Realistic transaction frictions have a tariff equivalent of 20%, much larger than the physical cost of trade. Inventories affect the dynamics of adjustment following a large unanticipated devaluation: The resulting increase in the wholesale price of imports and the increase in the interest rate cause importers to want to reduce their inventories rapidly, leading to a drop in imported goods (mainly driven by the extensive margin due to lumpy decisions). Because of the desire to reduce inventories, retail prices are raised more slowly than the wholesale price.

**Motivation** The authors present a number of stylized facts gathered from several micro- and macro-level data sets.

- Time lags in international trade are considerable. Importing time lags are about three weeks on average.
- Importing and exporting costs as a fraction of shipment are around 5–10%.

- In a panel of Chilean manufacturing plants, importers hold about 30% of its annual purchases in inventories, while non-importers hold about 20%. Plants tend to have 2.1 months worth of domestic goods on hand and 4.3 months of imported goods.
- In transactions level data from a U.S. steel wholesaler, international orders tend to be 50% larger and half as frequent as domestic orders.
- Comprehensive monthly data on U.S. exports and imports show that for most goods, exports are concentrated in just a few months of the year, rather than being sold continuously. Import values are higher for firms that have not recently imported.
- In Argentinian CPI data, nonstorable (no inventory) goods respond more to the NER. The short-run pass-through is smaller for storable goods that are held in inventories.
- As in Burstein et al. (2005), in six large devaluation episodes, the retail price rose more slowly than the wholesale price (import price at the dock). Imports collapsed in all episodes, driven mostly by the extensive margin (note that this is opposite of the traditional J-curve view of sluggish trade adjustment for developed economies). The higher the initial increase in the ToT, the larger the immediate drop in imports.

**Model** Partial equilibrium model of the importer's inventory problem. There's a continuum of importers. The state is denoted  $\eta_t$ . Importer  $j$  is monopolistically competitive with demand curve

$$y_j(\eta_t) = e^{\nu_j(\eta_t)} [p_j(\eta_t)]^{-\theta},$$

where  $\nu_j(\eta_t)$  is a demand shock with i.i.d. distribution  $\mathcal{N}(0, \sigma^2)$ . The wholesale per-unit cost of imported goods is  $\omega$  for all importers. Let  $i_j(\eta_t)$  denote the quantity imported. If  $i_j(\eta_t) > 0$ , the importer pays a fixed cost  $f$ . Re-exporting goods,  $i_j < 0$ , is not allowed. There is a one-period lag between ordering imports and delivery, i.e., sales  $q_j(\eta_t)$  are constrained not to exceed the firm's beginning-of-period inventory  $s_j(\eta_t)$ :

$$q_j(\eta_t) = \min\{y_j(\eta_t, p_j(\eta_t)), s_j(\eta_t)\}.$$

The law of motion for inventories is

$$s_j(\eta_{t+1}) = (1 - \delta)[s_j(\eta_t) - q_j(\eta_t) + i_j(\eta_t)],$$

where  $\delta$  is the depreciation rate. Let  $V^a(s, \nu)$  be the value function of a firm conditional on adjustment, and  $V^n(s, \nu)$  be the value function conditional on not adjusting. Then the value function is  $V(s, \nu) = \max\{V^a(s, \nu), V^n(s, \nu)\}$ , where

$$V^a(s, \nu) = \max_{p, i > 0} q(p, s, \nu)p - i\omega - f + \beta E[V(s', \nu')],$$

$$V^n(s, \nu) = \max_p q(p, s, \nu)p + \beta E[V(s', \nu')],$$

subject to the sales function  $q$  and law of motion for  $s'$ .

The optimal level of imports conditional on adjustment satisfies the FOC

$$\omega = \beta(1 - \delta)E[V_s(s', \nu')],$$

and since  $\nu_j$  is i.i.d., the optimal next-period inventories  $s'$  conditional on adjustment are constant. The adjustment threshold  $s^*(\nu)$  increases with  $\nu$ , as firms with high demand deplete their current inventories rapidly and thus import more readily.

Due to isoelastic demand, price is set at a constant markup  $\theta/(\theta-1)$  over marginal cost  $V_s(s, \nu)$ . However, marginal cost does not always equal the replacement cost  $\omega$  (see their Figure 3). For low  $s$ , the firm will stock out (i.e., fully deplete inventories), and so the price is set so high that inventories are just depleted. For slightly higher  $s$ , current inventory holdings do not constrain current sales but the firm will import to restock its inventory. Hence,  $\omega = V_s(s, \nu)$  and price equals a constant markup over the replacement cost. There is a kink in the price function at the adjustment threshold  $s^*(\nu)$ . For slightly higher  $s$ , the firm doesn't adjust and the value of a marginal good in inventory exceeds the replacement cost since it allows the firm to delay paying the fixed cost. For very high  $s$ , however, extra goods in inventory become a nuisance due to the carrying costs and the inability to re-export; hence, price is set below the markup times replacement cost to deplete the inventory.

**Simulations** The model is calibrated using some of the above-mentioned data sets. A period is one month. The resulting fixed cost is about 3.6% of the average value of an import shipment. The authors calculate the *tariff equivalent* of the import costs, i.e., how big of an ad valorem tariff the firm would pay to get rid of the fixed cost and delivery lag. The value to a firm facing no frictions but a tariff  $\tau$  is

$$V^f(\tau) = \max_p E_0 \sum_t (p_t - (1 + \tau)\omega)y(\eta_t, p_j(\eta_t)),$$

and the tariff equivalent then satisfies  $V^f(\tau) = E[V(0, \nu)]$ . It comes out to 20% for the benchmark calibration. This is due to not only the fixed cost but also the carrying costs.

For comparison, the model is calibrated to match domestic retailers. The delivery lag is set to half a month and the fixed cost  $f$  is set to match a twice as high frequency of orders. The difference in cost per shipment between this and the baseline is just 1.9%, but the tariff equivalent of the gap is 11 percentage points.

With  $f = 0$ , inventories are smaller, and the tariff equivalent (solely arising from the delivery lag) is roughly 2/3 of the benchmark. With the benchmark  $f$  but no delivery lag, inventory holdings drop (no reason to hold a buffer stock) and the tariff equivalent is 40% of the benchmark. Relative to the benchmark, increasing carrying costs through a doubling of  $\delta$  raises the tariff equivalent by 60%.

A simulation experiment is now carried out to match the model to one of the large devaluations. The devaluation is modeled as an unanticipated increase in  $\omega$ ,  $\Delta \log \omega = 0.5$ . Since the interest rate typically rises in these situations,  $\beta$  is decreased following the devaluation. The mean of  $\nu$  is shifted down by 0.15 to encompass the contraction in aggregate demand. Furthermore, it is assumed that importers produce their final output with both labor  $l$  and imported goods  $m$  through the production function  $y = l^\alpha m^{1-\alpha}$ , with  $\alpha = 0.25$ . This breaks the tight link between marginal cost and the wholesale price.

Following the devaluation, the fixed cost is high relative to profits, so firms desire to decrease their overoptimistically large inventories. In the transition from the previous steady state to the new one, firms run down their inventories by (1) importing less on average (primarily a drop in extensive margin) and (2) making the retail price rise less than one-for-one with the wholesale price (since the marginal value of inventories is small). This is consistent with the empirical evidence. In an alternative simulation experiment, the devaluation is assumed to be gradual, and firms learn

the full path of  $\omega$  from the beginning. In this case, firms increase imports initially to stock up on inventories while they are cheaper, but otherwise the conclusions are the same.

## 8.2 Burstein, Eichenbaum and Rebelo (2005): “Large Devaluations and the Real Exchange Rate”

**Summary** Studies the behavior of prices of traded and nontraded goods following five large devaluations. It is found that the ER pass-through to pure traded goods is high and consistent with relative PPP, while the pass-through to nontraded goods is low. This implies that the relative price of nontradables is the main driver of RER movements (which are substantial) in these cases. Using conventional measures of the price of traded goods, based on retail prices, is misleading, as these include a large component of nontradable distribution services as well as local goods that aren’t actually traded. Traded goods prices at the dock are much more consistent with relative PPP. In-depth analysis of the Argentinian devaluation of 2002, using various detailed data sets, confirms that goods prices change much more and more frequently than prices of services. There is also evidence of a “flight from quality,” i.e., consumers substitute to lesser-quality items which may bias downward official measures of CPI inflation. Finally, the authors look at large *appreciations* and medium-size devaluations; the main results are qualitatively the same, although more muted. For small business-cycle frequency fluctuations in the NER, the traded component of the RER is more important than the nontradable, but the latter still contributes significantly.

**Model** The real exchange rate is defined as

$$\text{RER}_t = \frac{P_t}{S_t P_t^*}.$$

Note that an appreciation of the RER corresponds to an increase in  $\text{RER}_t$ , whereas an appreciation of the NER corresponds to a decrease in  $S_t$  (which is in local per foreign units).

For the purposes of price accounting, the CPI is written as

$$P_t = (P_t^T)^{1-\omega} (P_t^N)^\omega,$$

where the price of tradables is

$$P_t^T = (P_t^I)^{1-\theta} (P_t^L)^\theta.$$

Here  $P_t^I$  is the retail price of traded (not the same as tradable) goods and  $P_t^L$  the retail price of local goods. They make the following assumptions.

- The technology used to transform traded goods or local goods into retail tradable goods uses nontradable distribution services and is of Cobb-Douglas form. The C-D exponent on distribution services is  $\phi$ .
- The price of distribution services equals the price of nontradable goods.
- The distribution sector is perfectly competitive.
- The ex-distribution price of local goods equals the price of nontradables.

From the first three assumptions, we can write

$$P_t^I = (\bar{P}_t^I)^{1-\phi}(P_t^N)^\phi, \quad P_t^L = (\bar{P}_t^L)^{1-\phi}(P_t^N)^\phi,$$

where a bar signifies ex-distribution price. From the fourth assumption,  $\bar{P}_t^L = P_t^N$ . Then the CPI can be written

$$P_t = (\bar{P}_t^I)^{1-\alpha}(P_t^N)^\alpha,$$

where  $\alpha = 1 - (1 - \theta)(1 - \phi)(1 - \omega)$ .

**Results** The five large devaluations are Argentina (2002), Brazil (1999), Korea (1997), Mexico (1994) and Thailand (1997). All these are accompanied by a large drop in the (trade-weighted) RER, so inflation is much slower than the NER depreciation. The prices of imports and exports move very much in lockstep with the NER. In contrast, the retail price of tradable goods moves much less and nontradables even less. The latter can't be explained by the behavior of government-controlled prices, as these have approximately the same rate of inflation as other nontradables.

The authors conduct a number of price accounting exercises. First, they argue that classifying goods as tradables and services as nontradables is too simplistic. By this measure, nontradables constitute about 50% of the CPI basket. But the distribution margin, defined as (retail price – producer price)/retail price, is roughly 50% for these countries. Since distribution services are inherently nontradable, this increases nontradables' share in CPI to 75%. Furthermore, some goods, like yogurt, aren't actually traded. Taking local goods into account (based on import statistics), reduces the weight of pure traded goods in CPI to about 14%.

They then use the CPI decomposition described under “Model.” They take the price of nontradables from the data and use two different measures of  $\bar{P}_t^I$ . The first assumes that relative PPP holds for pure traded goods:  $\bar{P}_t^I = kS_t\bar{P}_t^{I*}$ .  $\bar{P}_t^{I*}$  is a trade-weighted average of foreign import and export prices. The second measure is an equally weighted geometric average of import and export prices for the country in question. Conditional upon a choice for  $\alpha$ , the authors can now construct a theoretical CPI series and compare its rate of inflation with the actual rate of inflation. They consider four different assumptions on  $\omega, \theta, \phi$ .

- If all goods are assumed to be traded,  $\alpha = 0$ , the implied rate of inflation is much too high.
- If  $\phi = \theta = 0$ , i.e., no distribution costs or local goods, the rate of inflation is still too high.
- If  $\theta = 0$ , so there are no local goods, but the distribution margin is set to  $\phi = 0.5$ , the implied inflation rate looks better but is still a bit too high for all countries.
- If  $1 - \alpha$  (the share of pure traded goods in CPI) is set to the total import weight of consumption, the result is a lower bound on the importance of pure traded goods since it abstracts from exportables (for which they don't have good data). The implied inflation rate is close to the true one, but undershoots a bit.

It's unrealistic to assume that the distribution sector is perfectly competitive. It's possible to introduce a multiplicative markup factor in the prices  $P_t^I$  and  $P_t^L$ . Doing so, the authors can back out the implied markup that makes the model CPI fit the data. In the latter two cases above, the implied markup moves less than 20% over the devaluation, whereas it moves 40%–60% if no distribution costs or local goods are assumed.

The authors argue that, since all their devaluations are associated with a contraction, one would expect consumers to substitute toward lower-quality items, a “flight from quality.” If CPI price inspectors underestimate the extent of the quality reduction, this induces a downward bias in measured CPI inflation.

The authors decompose the RER into relative tradable prices and relative relative nontradable prices, as in Engel (1999). Using the conventional measure of retail prices for tradables, it would seem that the relative price of tradables is the most significant component of the RER movements. However, if the decomposition is based on at-the-dock prices of pure tradables (equally weighted average of import and export prices), the nontradable component becomes dominant.

The Argentinian experience is examined in detail using better data. Disaggregated CPI data shows that the rate of increase in retail prices is higher for goods that have a higher market share of imported and exported goods. This conclusion is backed by scanner data. A survey on prices of goods and some services in Buenos Aires shows that the frequency of price changes for goods is much higher than that for services. Finally, some evidence on flight from quality is presented: disappearing products tend to be the higher-priced ones, premium brands suffer larger declines in market share and surveys indicate that many consumers switched to cheaper brands.

The authors assess the robustness of their results by comparing with other episodes. First, they consider two large RER appreciations, in Argentina and Mexico. Again, movements in the prices of pure traded goods account for a small fraction of the movement in the RER. They also study medium-size devaluations in 11 OECD countries. There is still substantial ER pass-through to import and export prices, but less so than for the large devaluations. Using retail prices rather than pure traded prices still leads to bias but not as much as before. Finally, the authors consider small ER fluctuations at the business cycle frequency in OECD countries (the data is HP filtered to isolate higher frequencies). They find that the volatility of at-the-dock traded prices is smaller than that of the overall RER. However, at-the-dock traded prices still account for about 2/3 of the RER movements. Thus, the relative price of nontraded goods is a significant contributor, but it is not dominant in this case.

### 8.3 Carvalho and Nechio (2011): “Aggregation and the PPP Puzzle in a Sticky-Price Model”

**Summary** Develop a two-country model with multiple sectors in each country. The degree of price stickiness varies across sectors. The RER is more persistent and volatile in the heterogeneous-stickiness multiple-sector world than in a counterfactual one-sector world with average price stickiness. When the distribution of price stickiness across sectors is calibrated to match U.S. micro data, the model generates a half-life of the aggregate RER of 39 months, consistent with the consensus in the literature. The authors argue that the model helps explain the conflicting findings in the empirical literature on the aggregation bias and the PPP puzzle.

**Model** Two countries, complete markets. Consumers supply labor to intermediate firms that they own. Intermediate firms are monopolistically competitive and operate in one of  $K$  sectors. They can price discriminate across countries. Each sector has a continuum of firms. The relative weight of sector  $k$  in the country’s total mass of firms is  $f_k$ . The final consumption good is nontraded and produced competitively. Per-period utility is  $(C_t^{1-\sigma} - 1)/(1 - \gamma) - N_t^{1+\gamma}/(1 + \gamma)$  and the budget constraint is

$$P_t C_t + E_t[\Phi_{t,t+1} B_{t+1}] \leq W_t N_t + B_t + T_t,$$

where  $\Phi_{t,t+1}$  is the SDF and  $T_t$  are profits received from intermediate firms. Foreign consumers solve a similar problem. State-contingent assets are denominated in Home currency; the NER (price of Foreign currency in terms of Home's) is  $\mathcal{E}_t$ . The aggregate real exchange rate is

$$Q_t = \mathcal{E}_t \frac{P_t^*}{P_t}.$$

Due to complete markets, the Backus-Smith condition

$$Q_t = (C_t^*/C_t)^{-\sigma}$$

holds.

The final good  $Y_t$  is produced by combining aggregate bundles  $Y_{k,t}$  from sectors  $k = 1, \dots, K$ ; the elasticity of substitution across sectors is  $\eta$ . Aggregate sector bundles are produced by combining Home and Foreign sector bundles  $Y_{H,k,t}$  and  $Y_{F,k,t}$  using home-bias weight  $\omega$  and elasticity of substitution across countries  $\rho$ . Finally, Home and Foreign sector bundles are produced by combining goods from individual firms in the relevant sector in the relevant country; the elasticity of substitution within sectors is  $\theta$ .

Home's final good producing firm maximizes

$$P_t Y_t - \sum_k f_k \int_0^1 (P_{H,k,j,t} Y_{H,k,j,t} + P_{F,k,j,t} Y_{F,k,j,t}) dj$$

subject to the production technology for  $Y_t$ . The sectoral real exchange rate is

$$Q_{k,t} = \mathcal{E}_t \frac{P_{k,t}^*}{P_{k,t}},$$

with sectoral price indices appropriately defined.

The Calvo probability of updating in sector  $k$  is  $\alpha_k$ , common across the two countries. Intermediate firm  $j$  from sector  $k$  in Home chooses prices  $X_{H,k,j,t}, X_{H,k,j,t}^*$  to be charged domestically and abroad (LCP), to maximize

$$E_t \sum_s \Phi_{t,t+s} (1 - \alpha_k)^s (X_{H,k,j,t} Y_{H,k,j,t+s} + \mathcal{E}_{t+s} X_{H,k,j,t}^* Y_{H,k,j,t+s}^* - W_{t+s} N_{k,j,t+s})$$

subject to the demand curves and technology constraint

$$Y_{H,k,j,t} + Y_{H,k,j,t}^* = N_{k,j,t}^X.$$

The equilibrium is assumed to be symmetric so that all firms in a given sector within a country choose the same price. The model is closed by specifying a process for nominal aggregate demand  $Z_t = P_t Y_t$  in Home and Foreign (alternatively, an interest rate rule could be defined). The model is then log-linearized around a zero-inflation steady state.

Additionally, the authors construct a *counterfactual one-sector world economy* by setting the number of sectors  $K$  to 1 and letting the probability of price change in this sector equal the average probability in the multi-sector world  $\bar{\alpha} = \sum_k f_k \alpha_k$ .

To show some analytical results, the authors specialize to log utility, linear disutility of labor and linear production function (no strategic complementarity). NGDP is assumed to follow an

AR(1) process in first differences, with coefficient  $\rho_z$ . It is shown that the sectoral RER follows the AR(2) process

$$(1 - \rho_z L)(1 - (1 - \alpha_k)L)q_{k,t} = \varphi_k u_t,$$

where  $\varphi_k$  is a constant that depends on  $\beta, \alpha_k, \rho_z$  and  $u_t$  is white noise (depends linearly on the NGDP shocks in Home and Foreign). If the sectoral RERs are added up, the aggregate RER follows an ARMA( $K + 1, K - 1$ ) process (Granger and Morris, 1976):

$$(1 - \rho_z L) \prod_k (1 - (1 - \alpha_k)L)q_t = \left( \sum_k \prod_{j \neq k} (1 - (1 - \alpha_j)L) f_k \varphi_k \right) u_t.$$

The dynamics of the aggregate RER can therefore be quite different from those of the sectoral ones. However, from the above it follows that the RER in the counterfactual one-sector world economy follows an AR(2) process  $(1 - \rho_z L)(1 - (1 - \bar{\alpha})L)q_t^{\text{1sec}} = \varphi u_t$ . The authors consider three measures of persistence: the cumulated impulse response (CIR), largest autoregressive eigenvalue (LAR) and the sum of the autoregressive coefficients (SAC). For each of these measures, it is shown that the persistence of the multi-sector RER is higher than in the one-sector world. This also holds true for the variance. The reason is that these persistence measures are convex functions of the frequency of price adjustments.

**Simulations and empirics** The full model is then calibrated, assuming that the cross-sectoral distribution of price stickiness is the same in both countries. The data on sectoral frequency of price adjustment from Nakamura and Steinsson (2008) is used. The resulting average monthly frequency of price changes is once per 4.7 months ( $\bar{\alpha} = 0.211$ ). The parameter  $\rho_z$  is set to 0.8 for now, in line with empirical estimates.

The model generates a half-life of the aggregate RER of 39 months, in line with the consensus view of 3–5 years. The counterfactual one-sector world economy produces a half-life of just above 1 year. The volatility of the RER is more than twice as big in the heterogeneous model. The impulse response to a shock to the RER is hump-shaped as in the data, and the sluggishness is both in the up-life and quarter-life. The degree of persistence and volatility depends crucially (and positively) on the exogenous persistence parameter  $\rho_z$ .

To relate to the empirical literature, the authors decompose the persistence of the aggregate RER as follows. Let  $\mathcal{P}$  denote a persistence measure. The *total heterogeneity effect*  $\mathcal{P}(q) - \mathcal{P}(q^{\text{1sec}})$  is decomposed as

$$\mathcal{P}(q) - \mathcal{P}(q^{\text{1sec}}) = \left( \mathcal{P}(q) - \sum_k f_k \mathcal{P}(q_k) \right) + \left( \sum_k f_k \mathcal{P}(q_k) - \mathcal{P}(q^{\text{1sec}}) \right).$$

The first term is the *aggregation effect* (the difference between the persistence of the average and the average persistence) while the second term is the *counterfactuality effect* (the difference between the average persistence and the persistence in the counterfactual one-sector world economy). Using the Eurostat data from Imbs et al. (2005), it is shown that both these terms are positive in the data (for the standard persistence measures), but the counterfactuality effect is much larger. Chen and Engel (2005) and Crucini and Shintani (2008) had concluded that the aggregation effect was small. Imbs et al. (2005), by obtaining a half-life from their Mean Group estimated autoregressive model, had essentially concluded that the total heterogeneity effect is large. The conclusions in

the literature are therefore consistent with each other after all, as they were measuring different objects.

The (simplified) model predicts that the standard persistence measures for sector  $k$  should increase in  $(1 - \alpha_k)$ . The same is true for the sectoral variance. This cross-sectional prediction is tested. First, it is found that an AR(2) specification is a very good fit to almost all of the sectors in the data; the first and second estimated AR coefficients are for almost all sectors positive and negative, respectively, as predicted by the simple model. For most of the countries in the sample, there is a significant positive relationship between sectoral price stickiness and persistence/volatility of the RER.

It is found that a relatively high value of  $\rho_z$  is needed to match the theoretical AR(2) dynamics to the estimated ones. A value of  $\rho_z$  of 0.8, however, implies much too high an autocorrelation of changes in the NER in the model. If  $\rho_z$  were calibrated to match the latter, one would get  $\rho_z \approx 0.35$ . The authors explore whether this shortcoming can be fixed by tampering with the real rigidities, since these should in theory be able to generate additional endogenous persistence (less nominal adjustment per unit shock). An exercise is carried out where the  $\rho_z$  and the C-D exponent  $\chi$  are held fixed at 0.35 and 2/3, while the within-country elasticities of substitution  $\theta$  and  $\eta$  are increased gradually. It is found that if these are increased enough, the necessary persistence of the RER is salvaged. The required elasticities are quite high, but the authors interpret their exercise as showing what could happen if one were to model additional realistic real rigidities. They find it interesting that increasing the degree of real rigidities raises the persistence relatively more in sectors that are relatively flexible, thus perhaps explaining why persistence in the data is somewhat high uniformly across sectors.

#### 8.4 Engel (1999): “Accounting for U.S. Real Exchange Rate Changes”

**Summary** Decomposes the RER into the relative price of traded goods and the relative price of nontraded relative to traded goods. Using five different measures of traded and nontraded goods prices, finds that almost all the MSE of the RER over various horizons can be accounted for by drift and volatility in traded goods. This is true for most U.S. bilateral RERs with other developed economies. The findings seems to be qualitatively robust to misclassification of traded and nontraded goods.

**Model** The geometric domestic price index can be decomposed as

$$p_t = (1 - \alpha)p_t^T + \alpha p_t^N,$$

where  $\alpha$  is the share of nontraded goods. The same holds for the foreign country, whose variables are denoted with stars (and whose share of nontraded goods is  $\beta$ ). Then

$$q_t = s_t + p_t^* - p_t = x_t + y_t,$$

where  $s_t$  is the log NER (price of foreign currency in terms of domestic),

$$x_t = s_t + p_t^{*T} - p_t^T$$

is the relative price of traded goods and

$$y_t = \beta(p_t^{*N} - p_t^{*T}) - \alpha(p_t^N - p_t^T)$$

is the *relative relative* price of nontraded goods.

The paper chooses MSE as a comprehensive measure of both the drift and variance of RERs. Engel looks at two different decompositions of the MSE. The first assumes that  $x_t - x_{t-n}$  is uncorrelated with  $y_t - y_{t-n}$ , so that the share of the MSE of the RER attributed to traded goods is

$$\frac{\text{MSE}(x_t - x_{t-n})}{\text{MSE}(x_t - x_{t-n}) + \text{MSE}(y_t - y_{t-n})},$$

where

$$\text{MSE}(x_t - x_{t-n}) = \text{Var}(x_t - x_{t-n}) + [E(x_t - x_{t-n})]^2.$$

In general

$$\text{MSE}(a + b) = \text{MSE}(a) + \text{MSE}(b) + 2\text{Cov}(a, b) + 2E(a)E(b).$$

The second decomposition arbitrarily attributes half of the latter two terms to  $x_t$ , i.e.

$$\frac{\text{MSE}(x_t - x_{t-n}) + \text{Cov}(x_t - x_{t-n}, y_t - y_{t-n}) + E(x_t - x_{t-n})E(y_t - y_{t-n})}{\text{MSE}(q_t - q_{t-n})}.$$

Standard errors are calculated by bootstrapping: Two independent random walks for  $x_t$  and  $y_t$  are simulated, given the sample mean and variance for the differences.

**Results** Five different data sets are used.

1. *CPI data.* Traded goods are commodities, while nontraded are housing and services. The weight of nontraded goods in the overall index is computed by regression. The author notes that marketing is a component of the traded goods series, although it probably shouldn't be included. On the other hand, financial services are part of the nontraded series. The traded and nontraded indexes are approximately uncorrelated in first differences, so the first MSE decomposition is used. For all countries except Canada, traded goods account for almost 100% of the movement of the RER, over all horizons (from 1 month up to almost 40 years). This also holds true if the decomposition were based on just variance or squared drift. It holds over both the fixed ER period 1962–69 and the subsequent floating period.
2. *Output prices.* Data from OECD's sectoral database. De Gregorio et al. (1994) had classified these sectors as traded or nontraded based on export share in total production. MSE decompositions give the same results as for CPI above. Since marketing should be a much smaller proportion of output prices, this suggests that the CPI results aren't driven by misclassification.
3. *PCE deflators.* Traded goods is the deflator index for PCE on commodities; nontraded are services. Pattern is the same as for CPI and output prices.
4. *PPI.* Overall PPI is used as an index of traded goods, so that  $x_t = s_t + \log \text{PPI}_t^* - \log \text{PPI}_t$  and  $y_t = \log \text{CPI}_t^* - \log \text{PPI}_t^* - [\log \text{CPI}_t - \log \text{PPI}_t]$ . The author notes three shortcomings of this approach: some of the PPI should be attributed to nontraded goods, the data come from different surveys, and the CPI index is not geometric. Furthermore,  $x_t$  and  $y_t$  are no longer uncorrelated in differences, so the second MSE decomposition must be used. The results still indicate, however, that traded goods prices are the dominant driver of movements in the RER, although the share is about 80% rather than close to 100%.

5. *U.S.-Japan RER*. The author takes a closer look at the U.S. vs. Japan RER. The reason why the traded goods component is so dominant is that the upward movement in Japanese relative nontraded to traded prices has been matched very closely by similar movements in the U.S. relative price. Furthermore, the share of nontraded goods in Japanese consumption is only about 1/3. There is no evidence that the relative price of marketing has moved over time.

Even if the relative price of marketing is constant over time, the MSE decomposition could still be biased. Suppose that the price of marketing were perfectly correlated with the measure  $p_t^N$  of nontraded goods prices. Then the measure  $p_t^T$  of traded goods prices could be written  $p_t^T = \gamma p_t^N + (1 - \gamma)p_t^C$ , where  $p_t^C$  is the price of truly traded goods. Then the true weight of nontraded goods in the overall price index is  $\alpha + \gamma(1 - \alpha)$ . Engel redoes some of the calculations assuming values of 1/3 and 2/3 for  $\gamma$ . While a higher value of  $\gamma$  does increase the importance of relative nontraded goods prices for RER movements, the traded goods component continues to be dominant.

Why are traded goods so important and nontraded so unimportant? One explanation is terms of trade movements: If relative prices of various traded goods fluctuate a lot, and these prices receive different weights in countries' price indices, the relative traded goods price index will fluctuate, even if PPP holds good for good. However, when Engel computes foreign price indices based on U.S. weights, the results aren't much different.

The remaining explanation is that there are large deviations from PPP in traded goods (among rich countries). Engel refrains from speculating on why that may be but lists a number of areas that must be better understood. He also cautions against drawing conclusions about the long run from his study, since the stationarity and possible cointegration between  $q_t$ ,  $y_t$  and  $x_t$  can't be discerned based on his limited sample.

## 8.5 Gopinath, Gourinchas, Hsieh and Li (2009): “Estimating the Border Effect: Some New Evidence”

**Summary** Use detailed data from a large grocery chain with stores in the U.S. and Canada to estimate the border effect. The data set has weekly prices and costs for products identified at the barcode level from several stores in both countries, located at various distances from the border. During the 2004–2007 period, the median price gap of Canadian products rose almost in lockstep with the NER. This was almost entirely due to a rise in the relative cost gap, as relative markups stayed virtually constant. A circular city model is presented that explains the relationship between relative prices, distance to the border and border costs. The model predicts that markets become fully segmented when border costs rise above a certain level; in this case, cross-border prices are not further affected by the border cost. Empirically, the authors obtain a more precise estimate of the border effect using a regression discontinuity design. There are large price discontinuities at the border, but these range from positive to negative across goods. Again, the entire discontinuity is driven by wholesale prices, not markups. This is consistent with full market segmentation.

**Model** The authors develop a model with a circular world of unit length. Locations are indexed by  $\omega \in [0, 1]$ . There are  $N_A$  stores in country A and  $N_B$  in country B. The borders are located at locations  $\omega = 0$  and  $\omega = N_A/(N_A + N_B)$  (see their Figure 5). Each store sells a homogeneous good. There is a unit mass of consumers who each buy one unit of the good. The utility of a consumer

located at  $\omega$  from the purchase of a good from a store located at location  $\omega_i$  is

$$u(\omega) = \nu - \theta p - t(\omega_i - \omega) + bI(\omega_i, \omega).$$

Here  $\theta$  is the price elasticity of demand,  $t$  the unit cost of travel,  $b$  is the border cost and  $I(\omega_i, \omega) \in \{0, 1\}$  equals 0 iff.  $\omega$  and  $\omega_i$  are on the same side of the border. The marginal cost of producing a good in location  $\omega_i$  is  $c_i = \min\{\chi_i, \chi_{-i} + b_c\}$  in country  $i \in \{A, B\}$ , where  $\chi_i$  is the wholesale price in country  $i$  and  $b_c$  the border cost to the retailer.

They look for an equilibrium where each consumer purchases from the nearest store. Between any two stores  $i$  and  $i + 1$ , there will be a marginal consumer who is indifferent. It is possible to derive the total demand for products at each interior store  $i$  (i.e., one that is not right next to the border) in terms of the prices of firms  $i - 1$ ,  $i$  and  $i + 1$ . The store then chooses  $p_i$  to maximize its static profits, taking the other prices as given. The FOC gives rise to a difference equation which is solved by imposing an initial condition for the price at the border. Proposition 1 implies that the price  $p_i$  of store  $i$  only depends on the border cost indirectly through its influence on the price of the store at the border, and the indirect effect decreases with the distance from the border.

They then characterize the prices at the border. There are two cases:

- If the border cost  $b$  is high, the marginal consumer is located right on the border. The price difference at the border is too small to warrant crossing the border, so national markets are fully segmented. The observed price difference at the border is thus independent of  $b$ .
- If the border cost  $b$  is small, there is partial segmentation, and the model solution depends on whether the marginal consumer is located in country A or B. The border cost  $b$  enters the expressions for the prices. Intuitively, with integrated markets, stores compete for customers on either side of the border. Hence, if  $c_A > c_B$ , the border store in country A charges a lower price than interior stores in its country, while the border store in country B charges a higher price than interior stores in its country.

The model illustrates that simple regressions using a border dummy and distance from the border may be misleading. If the two countries were completely symmetric, the endogenous price distribution would be identical across countries with no price discontinuity at the border, despite the presence of border costs. On the other hand, if the border cost is so high that markets become fully segmented, the magnitude of the border cost does not further influence prices. Furthermore, idiosyncratic factors in the two countries may influence prices, but a regression may mistakenly attribute this to border effects (this is outside the model). The authors thus argue for an RD design as a better empirical specification.

**Empirics** The data is from a leading grocery store chain that operates in the U.S. and Canada. Data is weekly and across many stores and product codes. It is possible to match goods across stores precisely using unique identifiers. Prices exclude taxes. The data set includes information on the wholesale cost, i.e., the price the retailer pays for the product at a wholesaler. This is not the total cost of the good for the retailer, since transportation and wholesale rebates are not included. To allow for this, the authors compute the net cost as the price minus “adjusted gross profits.” It is argued that at short horizons, where rent, capital and labor are fixed, this measure may be interpreted as the marginal cost of the good to the retailer and adjusted gross profits as the markup. The prices in the data set change every 2.4 weeks in the U.S. (4.5 in Canada).

Preliminary summary evidence on the failure of the LOP is presented. The median cross-border price gap varies across time and rises from 5% in 2004 to 15% in June 2007. The time series profile is almost identical to that of the NER, since the Canadian dollar appreciated markedly in that period. Using the decomposition  $\log(\text{price}) = \log(\text{cost}) + \log(\text{markup})$ , the authors study the two subcomponents. It is found that the entire price gap is accounted for by the relative costs, whereas the relative markup stays basically constant. There is a large dispersion of price gaps across products; some gaps are positive, others negative. Markups are much tighter linked. The median price gaps within countries are much smaller. The preliminary evidence thus indicates high border costs and virtually full market segmentation.

The authors move on to an RD design. The model is

$$\log p_i^k = \alpha^k + \gamma^k C_i + \beta^{k'} X_i + \varepsilon_i^k,$$

where  $k$  indexes products and  $i$  stores.  $C_i$  is a dummy that equals 1 if the store is in Canada.  $X_i$  are controls (demographics, density of retail stores in the area, opening year, household income).  $\gamma^k$  is the border effect, the parameter of interest. The problem is that  $E[\varepsilon_i^k | C_i, X_i] \neq 0$ , i.e., unobserved characteristics may correlate with the location of store  $i$ .

Let  $D_i$  be the distance of store  $i$  to the border ( $D_i > 0$  if located in the U.S.). If

$$\lim_{\delta \uparrow 0} E[\varepsilon_i^k | D_i = \delta] = \lim_{\delta \downarrow 0} E[\varepsilon_i^k | D_i = \delta],$$

i.e., unobserved characteristics don't change discontinuously at the border, then the border effect may be estimated consistently by

$$\gamma^k = \lim_{\delta \uparrow 0} E[\log p_i^k - \beta^{k'} X_i | D_i = \delta] - \lim_{\delta \downarrow 0} E[\log p_i^k - \beta^{k'} X_i | D_i = \delta].$$

Each of the two terms can be estimated by local linear regression. The bandwidth (cut-off distance for  $\delta$ ) may be obtained by cross-validation.

There is no significant discontinuity in store density right at the border, suggesting that stores don't locate strategically. Most of the covariates don't display marked discontinuities at the border either.

The RD estimates tell the same story as the preliminary evidence above. The discontinuities at the border are significant for most products but vary between being positive and negative. The median discontinuity tracks the NER closely. This is found to be almost 100% due to changes in relative wholesale prices, not markups. This indicates full market segmentation, including in wholesaling.

Some product brands are produced by the retail chain itself. It should thus have more control over the supply chain. We would thus expect the market segmentation to be weaker for these products, which is indeed the case in the data.

Finally, the authors investigate whether a border effect is present at the Washington-Oregon border. There is no evidence of price or cost discontinuities.

The authors conclude that their microlevel data don't support the hypothesis that aggregate level border effects are due to compositional effects. Surprisingly, cross border price differences are due to an apparently tradable component of costs, not markup differences.

## 8.6 Imbs, Mumtaz, Ravn and Rey (2005): “PPP Strikes Back: Aggregation and the Real Exchange Rate”

**Summary** The authors argue that time series or panel data models of the persistence of RERs that impose common persistence among sectoral RERs (or use aggregate data) are likely to be biased upwards relative to the true mean sectoral persistence, if these persistence parameters are indeed heterogeneous among sectors (as in a random coefficients setup). Intuitively, sectors with higher persistence have higher variance and so receive larger weight when estimating the supposedly common persistence parameter by standard procedures. Monte Carlo and analytical evidence is presented to back up this claim. The authors instead advocate using estimation methods that allow for dynamic heterogeneity. When using such methods, they estimate a mean half-life of about 1–2 years as opposed to the consensus in the literature of 3–5 years.

**Model** Suppose the true DGP for the RER  $q_{it}$  in sector  $i$  is given by

$$q_{it} = \gamma_i + \rho_i q_{i,t-1} + \varepsilon_{it}, \quad (2)$$

where  $\gamma_i = \gamma + \eta_i^\gamma$ ,  $\rho_i = \rho + \eta_i^\rho$  for some mean-zero variables  $\eta_i^\gamma, \eta_i^\rho$ . The parameter of interest is  $\rho$ , or more precisely the half-life  $\log 2 / \log \rho$ . Then

$$q_{it} = \gamma + \rho q_{i,t-1} + e_{it}, \quad (3)$$

where

$$e_{it} = \varepsilon_{it} + \eta_i^\gamma + q_{i,t-1} \eta_i^\rho.$$

If one were to estimate the common-coefficient model (3) by, say, pooled OLS, the resulting estimate of  $\rho$  would be biased, since the regressor  $q_{i,t-1}$  is correlated with the error term  $e_{it}$  (at least conditional on  $\eta_i^\gamma, \eta_i^\rho$ ). In fact, we can say more.<sup>5</sup> Set  $\gamma_i \equiv 0$  for simplicity. The sign of the bias of the common-coefficients pooled OLS-estimator is given by the sign of

$$E[e_{it} q_{i,t-1}] = E[q_{i,t-1}^2 \eta_i^\rho] = E\{E[q_{i,t-1}^2 | \eta_i^\rho] \eta_i^\rho\} = E\left[\frac{\sigma_i^2}{1 - \rho_i^2} \eta_i^\rho\right].$$

Here I have set  $\sigma_i^2 = E[\varepsilon_{it}^2]$  and assumed  $E[\varepsilon_{it} q_{i,t-1}] = 0$  and that  $\eta_i^\rho$  is independent of  $\{\varepsilon_{it}\}_t$ . If the distribution of  $\rho_i$  is concentrated on  $(0, 1)$ ,  $\eta_i^\rho$  and  $\rho_i^2$  are positively related. It follows that the above bias will be positive.

Instead of panel regressions, some papers perform a time series regression on aggregate data. Let  $\omega_i$  be the weights for aggregating sectoral RERs to the aggregate RER  $q_t = \sum_i \omega_i q_{it}$ . Then under the above true model,

$$q_t = \gamma + \rho q_{t-1} + e_t,$$

where  $e_t = \sum_i \omega_i e_{it}$ . For the same reason as above, OLS estimation will be biased, and the bias may be expected to be positive.

A consistent estimator is Pesaran and Smith’s (1995) Mean Group estimator, which is just the cross-sectional average of the equation-by-equation OLS estimates  $\hat{\rho}_i$  from (2). Imbs et al. also consider a variety of other estimators that differ in their treatment of the error covariance structure.

<sup>5</sup>The following is a simplified version of Imbs et al.’s argument.

**Results** The dataset is from Eurostat and contains monthly cross-country observations on sectoral prices. The RERs are CPI-based and relative to the U.S. Standard common-coefficients estimation procedures such as FE or Arellano-Bond give standard estimates of a half-life between 3 and 5 years. Estimators that allow for dynamic heterogeneity all yield substantively smaller estimates of the “average” half-life (I put “average” in quotation marks due to Jensen’s inequality: the average  $\rho$  is estimated consistently but the resulting half-life is not the average of the sectoral half-lives). The MG estimator gives a point estimate of about 2 years, while other estimators that adjust for cross-sectional correlation in the errors yield even smaller estimates. Standard errors are on the order of two months. The individual sector and country estimates of  $\hat{\rho}_i$  reveal substantial heterogeneity, particularly among sectors within countries, but also within sectors across countries. Generally, traded sectors tend to display lower persistence than nontraded ones.

A Monte Carlo experiment confirms that severe upward bias may result from estimators that don’t take into account the possibility of cross-sectional heterogeneity in persistence. Furthermore, the average persistence implied by their preferred estimates squares well with the calibrations in the theoretical model of Chari, Kehoe and McGrattan (2002), and so is consistent with only a moderate degree of nominal stickiness.

Finally, the authors conduct some robustness checks. First, a Hausman test (performed for each cross section unit) based on the difference between OLS and IV rejects the notion that measurement error of sectoral RERs is an issue. Second, the authors perform some bootstrap-based and analytical bias corrections on their autoregressive estimates. The resulting bias-corrected half-lives are somewhat higher than the uncorrected ones, but still well below the literature’s consensus range.

The authors conclude that it is important to allow for cross-sectional heterogeneity when holding theory up against data. The aggregate RER is still a well-defined object with a well-defined aggregate persistence. However, the high persistence estimates arising from such time series regressions do not imply a failure of theory; in fact, the properly estimated mean sectoral persistence seems to be in line with what is produced by standard OE models with price stickiness.

## 9 LCP vs PCP

### 9.1 Atkeson and Burstein (2008): “Pricing-to-Market, Trade Costs, and International Relative Prices”

**Summary** Relative PPP predicts that the ToT should move one for one with the PPI-based RER. On the other hand, since the CPI is a trade-weighted average, the CPI-based RER should be smoother than the PPI-based RER. However, in U.S. data, the ToT are much less volatile than the PPI-based RER, while the CPI-based RER is about as volatile. This indicates that relative PPP does not hold. The authors develop a flexible price model with nested CES demand functions and trade costs. A reasonable calibration of the model matches the empirical facts. To break the link between the RER and ToT, it is necessary to have pricing-to-market. Two ingredients are key: imperfect competition with variable markups and international trade costs. Without variable markups, firms would set prices at a country-constant markup over costs and relative PPP would hold. Without trade costs, firms would face the same competitors at home and abroad and thus don’t price-to-market. For the CPI movements to be large, the import share of expenditures must be small, which is ensured with trade costs. To get the relative volatility of the ToT to move in the right direction, it is necessary to have a large amount of price dispersion since only large firms

(that have significant market power abroad) price-to-market in the required way.

**Empirics and basic theory** Their Figure 1 shows the evolution of the U.S. PPI-based RER, CPI-based RER and the ToT. Movements in the former two are of approximately the same magnitude, whereas the latter is 1/3 to 2/3 as volatile. The ToT is positive correlated with the PPI-based RER, but less than perfectly.

The log change in the PPI-based RER may be decomposed as

$$\frac{\widehat{PPI}}{\widehat{PPI}^*} = \frac{\widehat{EPI}}{\widehat{IPI}} + \frac{\widehat{PPI}}{\widehat{EPI}} + \frac{\widehat{IPI}}{\widehat{PPI}^*}, \quad (4)$$

where the first term on the RHS is the ToT, the second is the ratio of producer and export prices, and the third is the ratio of import (foreign export) and foreign producer prices. All prices are expressed in a common currency. Under relative PPP, the second and third terms should be zero, and so the PPI-based RER and the ToT should move one for one. In the data all three terms on the RHS are positively correlated with the LHS. This holds both at the aggregate, trade-weighted level as well as for country-of-origin and sectoral measures.

Consider a simple two-country symmetric model with balanced trade. Since there are only two countries,  $\widehat{EPI}_1 = \widehat{IPI}_2$  and vice versa. The log change in the CPI for country 1 is then approximately

$$\widehat{CPI}_1 \approx \widehat{PPI}_1 + s_M(\widehat{IPI}_1 - \widehat{EPI}_1),$$

where  $s_M$  is the share of consumption spent on imports (same for both countries due to symmetry and balanced trade), and for country 2,

$$\widehat{CPI}_2 \approx \widehat{PPI}_2 + s_M(\widehat{EPI}_1 - \widehat{IPI}_1).$$

Then

$$\frac{\widehat{CPI}_1 - \widehat{CPI}_2}{\widehat{PPI}_1 - \widehat{PPI}_2} \approx 1 - 2s_M \frac{\widehat{EPI}_1 - \widehat{IPI}_1}{\widehat{PPI}_1 - \widehat{PPI}_2}.$$

Hence, for the CPI-based RER to move about one for one with the PPI-based RER, it is necessary to have (1) a low import share  $s_M$  and (2) small movements  $\widehat{EPI}_1 - \widehat{IPI}_1$  in the ToT relative to the PPI-based RER.

**Model** Two symmetric countries indexed by  $i = 1, 2$ . Aggregate shocks to productivity are seen as the driving shocks. Consumer preferences are  $u(c, l) = \mu \log c + (1 - \mu) \log(1 - l)$ . Asset markets are complete, so the Backus-Smith condition  $c_2/c_1 = P_1/P_2$  holds. The final nontradable consumption good  $c_i$  in each country is produced by a competitive firm using output from a continuum of sectors indexed by  $j \in [0, 1]$  using a CES production function with elasticity  $\eta$ . This leads to an inverse demand function

$$\frac{P_{ij}}{P_i} = \left( \frac{y_{ij}}{c_i} \right)^{-1/\eta}$$

for goods in sector  $j$ . In each country-sector pair there are  $K$  domestic firms selling distinct goods. The  $K$  firms from the corresponding Foreign sector may sell goods in the Home sector as well (more

on this below). The sectoral output is a CES aggregate of all firm outputs. The inverse demand function for goods within the sector is then

$$\frac{P_{ijk}}{P_{ij}} = \left( \frac{q_{ijk}}{y_{ij}} \right)^{-1/\rho}.$$

The production function is CRS and uses only labor,  $A_i z l$ , where  $A_i$  denotes aggregate productivity and  $z$  is idiosyncratic productivity that differs across firms,  $\log z \sim \mathcal{N}(0, \theta)$ . Marginal costs, excluding trade costs, are then  $W_i/(A_i z)$ . There is a fixed labor cost  $F$  for a firm that wishes to export. Furthermore, it pays a proportional “iceberg” cost  $D \geq 1$  for each good exported, so marginal cost for exported goods is  $DW_i/(A_i z)$ .

Goods within a sector are more substitutable than goods across sectors,  $\rho > \eta > 1$ . Domestic firms in a sector always enter. The entry decision of foreign firms is modeled as follows. All firms that have entered a sector are assumed to play a static game of quantity competition in every period: They choose their quantity taking as given other quantities, the wage rate and the final consumption price and quantity, but they recognize that their decision influences sectoral prices and quantities. I.e., they choose  $P, q$  to maximize  $Pq - qW_i/(A_i z_{ijk})$  subject to the inverse demand curve. The FOCs give, for country 1,

$$P_{1jk} = \frac{\varepsilon(s_{1jk})}{\varepsilon(s_{1jk}) - 1} \frac{W_1}{A_1 z_{1jk}}, \quad \varepsilon(s) = [\rho^{-1}(1-s) + \eta^{-1}s]^{-1},$$

where  $s_{ijk} = P_{ijk}q_{ijk}/\sum_l P_{ijl}q_{ijl}$  is the market share for firm  $k$  in its sector. In terms of prices,  $s_{ijk} = P_{ijk}^{1-\rho}/\sum_l P_{ijl}^{1-\rho}$ .

The number of foreign firms that enter a sector is determined sequentially. Let aggregate variables be given. First the most productive foreign firm tries to enter (this is just one of many possible equilibria). Equilibrium prices and quantities are computed, recognizing that it has a higher marginal cost. If the foreign firm’s resulting profits exceed the fixed cost, it enters, and the next-most productive foreign firm tries to enter. Otherwise, no foreign firms enter. This process is then repeated. The general equilibrium is determined by finding a fixed point in terms of aggregate variables for the sectoral entry decision and price and quantity determination.

The assumption  $\rho > \eta$  implies that markups increase with the market share: a firm with no market share perceives only the sectoral demand elasticity  $\rho$ , whereas a firm with the entire market share perceives the lower elasticity of demand across sectors  $\eta$ . This opens up for the possibility of less-than-full pass-through of costs: If a firm’s marginal costs go up relative to other firms, this firm loses market share and so decreases its markup in equilibrium. If there are trade costs, firms will tend to have a smaller market share abroad than at home. Hence, if firms in country 1 experience an aggregate cost shock, they will price-to-market and raise their export price less than their domestic producer price. Thus, the ToT rises less than the PPI-based RER.

**Simulations** The model is calibrated to match trade volume, the share of firms that export and the median Herfindahl index across sectors. Two alternatives to the benchmark are considered. One sets  $\rho = \eta$  which gives constant markups. The other sets  $D = 1$  and  $F = 0$  so there is frictionless trade.

The benchmark model manages to reproduce all the qualitative features of the relative volatilities and comovement of the PPI- and CPI-based RER and ToT. As in the data, the subcomponents (4) of relative PPI-RER changes are positively correlated. The import share is small at  $s_M = 0.165$

as required. The CPI-based RER doesn't move quite as much as the PPI-based RER, although it's close. The authors show that an extension of the model to include distribution costs increase the relative volatility of the CPI-based RER (they decrease the effective import share, and their price fluctuations exceed those of the PPI-based RER if there is incomplete pass-through).

In the frictionless trade parameterization, there is still incomplete pass-through since  $\eta < \rho$ , but all firms have the same cost for sales in each country, so the market share, price and quantity is the same across countries for each firm. Therefore export prices remain constant relative to domestic producer prices. The CPI-based RER does not move at all because relative PPP holds and the consumption baskets are identical across countries.

For the constant markup parameterization, relative PPP holds good for good.

In the benchmark model there is a great deal of heterogeneity in pricing behaviors. For many firms with small market shares, export prices tend to change *more* than domestic producer prices. However, large firms dominate the index. Intuitively, for a firm that exports, there are two effects of an *idiosyncratic* increase in its marginal costs.

- First, its home market share tends to be larger than that abroad, and so the domestic markup is reduced more; this effect goes the wrong way.
- Second, if the domestic sectoral price level rises relative to that abroad, a domestic firm has a comparatively high price in the foreign market and so wants to reduce its markup abroad; this effect goes the right way.

The latter effect only wins out for large firms with sizeable foreign market shares (note that by assumption, only the most productive firms enter the foreign market).

The sensitivity analysis shows that the only really crucial feature is the relatively high value of the within-sector Herfindahl index. The authors conclude that their results are sensitive to the model.

## 9.2 Gopinath and Itskhoki (2010): “Frequency of Price Adjustment and Pass-Through”

**Summary** In the data, goods whose prices are adjusted frequently also exhibit the largest amount of long-run pass-through (LRPT) following exchange rate movements. The correlation between frequency and size of adjustment is close to zero. The authors develop two models that explain this. First, in a static menu cost set-up, they show that variable markups deliver the desired result. Higher markup elasticity raises the curvature of the profit function with respect to prices, which reduces the region of non-adjustment, but it also reduces the size of the desired adjustment so that the price is more likely to stay within the adjustment bounds. The latter effect dominates, so a higher markup elasticity reduces both frequency and size of adjustment. The authors go on to develop a dynamic version of the model that they calibrate and simulate. Varying the markup elasticity generates a good deal of variation in the frequency of adjustment, although less than in the data. If both the markup elasticity and the menu cost vary independently across firms in the cross-section, the zero correlation between frequency and size of adjustment can be matched. Neither a menu cost model with constant markups nor a Calvo model matches the facts.

**Empirics** The data set from Gopinath and Rigobon (2008) is used. Only dollar-priced imports are considered, restricting attention to market transactions and excluding intrafirm prices. The

authors run two kinds of regressions. Before running the regressions, the products are sorted into two equal-sized bins according to the frequency of price adjustment.

- The first regression is at the micro-level and uses life-long changes, i.e., the cumulative changes over the life of the product in question. The (CPI inflation adjusted) change in log price is regressed on the cumulative change in the log bilateral CPI-based RER (an increase means a real dollar depreciation). The *life-long pass-through* is then the estimated coefficient on the RER.
- The second regression is at the country level. For each month it regresses the monthly change in the log price level (for that country) on 24 distributed lags of the bilateral RER. The *aggregate LRPT* is then the sum of the lag coefficients.

Using the life-long PT regressions, the PT is more than double for the high-frequency goods compared to the low-frequency goods (44% to 21%). This is even starker if attention is restricted to imports from high-income OECD countries. Results are basically the same in the sample of differentiated goods. The difference holds within countries as well as in the aggregate. If the number of frequency bins is increased to 10, LRPT goes from 18% to 75% across these bins. The median absolute size of price adjustment is about the same across the 10 frequency bins. The authors also regress the change in log price on the change in log RER, the frequency of adjustment and the interaction between the latter two. The interaction term comes out significantly positive. Finally, a variance decomposition of price frequency into between- and within-sector components shows that within-sector variation is dominant.

For the aggregate LRPT regressions, the results are qualitatively the same, although somewhat diluted and with larger standard errors.

The authors stress that in this paper they are measuring *long-run* as opposed to medium-run PT. Gopinath and Rigobon (2008) measured the latter. They found that MRPT is decreasing in frequency, but that is purely driven by the non-dollar-denominated prices.

**Simple model** There is a single firm. It sets a price before observing productivity and ER shocks, after which it can pay a menu cost to be allowed to update its price. The demand schedule is  $Q = \varphi(P|\sigma, \varepsilon)$ , where  $P$  is its price and  $\sigma, \varepsilon$  are parameters. The price elasticity of demand is  $\tilde{\sigma} = -\partial \log \varphi / \partial \log P$  and the superelasticity of demand is  $\tilde{\varepsilon} = \partial \log \tilde{\sigma} / \partial \log P$ . A normalization is imposed such that if  $P = 1$ , then  $\tilde{\sigma} = \sigma$  and  $\tilde{\varepsilon} = \varepsilon$ . Furthermore, it is assumed that  $\varphi(1|\sigma, \varepsilon) \equiv 1$ . The constant MC are

$$MC = (1 - a)(1 + \phi e)c,$$

where  $a$  is productivity shock,  $e$  a real ER shock and  $\phi$  is the sensitivity of MC to the ER. The shocks are independent with mean zero. The constant  $c$  is set to  $(\sigma - 1)/\sigma$ , which implies that the optimal flexible price when  $a = e = 0$  is 1. Profits are

$$\Pi(P|a, e) = \varphi(P)(P - MC(a, e)).$$

The desired price  $P(a, e)$  is the argmax of the above;  $\Pi(a, e)$  is the optimal profit  $\Pi(P(a, e)|a, e)$ . The desired price is given implicitly by

$$P(a, e) = \frac{\tilde{\sigma}(P(a, e))}{\tilde{\sigma}(P(a, e)) - 1} (1 - a)(1 + \phi e)c.$$

Let  $\bar{P}_0$  denote the price the firm sets before the shocks hit. The firm chooses to reset its price if

$$L(a, e) \equiv \Pi(a, e) - \Pi(\bar{P}_0|a, e) > \kappa.$$

The non-adjustment region is therefore  $\Delta = \{(a, e) : L(a, e) \leq \kappa\}$ , and the initial price is optimally chosen as

$$\bar{P}_0 = \arg \max_P \int_{(a,e) \in \Delta} \Pi(P|a, e) dF(a, e) = \arg \max_P \varphi(P)(P - E_\Delta[MC]),$$

using linearity of profits in MC. A log-linearization shows that  $\bar{P}_0 \approx P(0, 0) = 1$ . Intuitively, if  $\Delta$  is nearly symmetric around zero and shocks are symmetrically distributed,  $a = e = 0$  is a good starting point for setting the price.

It is shown that

$$\frac{P(a, e) - \bar{P}_0}{\bar{P}_0} \approx \Psi(-a + \phi e), \quad \Psi = \frac{1}{1 + \varepsilon/(\sigma - 1)},$$

and ERPT is  $\Psi_e = \phi\Psi$ . To interpret, note that the markup elasticity is

$$\frac{\partial \log[\tilde{\sigma}/(\tilde{\sigma} - 1)]}{\partial \log P} = -\frac{\tilde{\varepsilon}}{\tilde{\sigma} - 1},$$

which is  $-\varepsilon/(\sigma - 1)$  at  $P = 1$ .  $\Psi$  is a negative transformation of the (absolute value of the) markup elasticity. A higher price increases the elasticity of demand which leads to a lower desired markup.

They also show that

$$L(a, e) \approx \frac{1}{2} \frac{\sigma - 1}{\Psi} \left( \frac{P(a, e) - \bar{P}_0}{\bar{P}_0} \right)^2.$$

Hence, ceteris paribus, a higher (absolute) markup elasticity increases  $L(a, e)$ , the benefit from adjusting. This is because it increases the curvature of the profit function.

Combining the above,

$$L(a, e) \approx \frac{1}{2}(\sigma - 1)\Psi(-a + \phi e)^2.$$

This is decreasing in the markup elasticity. Even though a high markup elasticity increases profit losses conditional on the PT, it also reduces the desired PT, and so the end result is a lower frequency (i.e., probability) of price adjustment. Note also that the frequency of price adjustment increases with the sensitivity of MC to the ER,  $\phi$ . An increase in the menu cost  $\kappa$  clearly reduces the frequency of adjustment.

The model predicts that frequency and size of adjustment are positively correlated (since both depend negatively on the markup elasticity). However, this can be diluted (as in the data) by introducing independent cross-sectional variation in  $\varepsilon$  and  $\kappa$ .

**Dynamic model and simulations** The dynamic model has continuum of goods varieties. There is a unit measure of U.S. varieties and  $\omega < 1$  of foreign varieties. The demand schedule is derived from the Kimball (1995) aggregator, and results in

$$\varphi(P_j, P) = \left[ 1 - \varepsilon \log \frac{P_j}{P} \right]^{\sigma/\varepsilon}.$$

MC are constant and derive from domestic and foreign wages. Profits are

$$\Pi_{jt}(P_{jt}) = \left[ P_{jt} - \frac{W_t^{1-\phi} (W_t^*)^\phi}{A_{jt}} \right] \varphi(P_{jt}, P_t),$$

where  $\phi$  is the share of foreign labor inputs in production.  $A_{jt}$  is idiosyncratic productivity, which follows an AR(1) in logs. The firm is assumed to only base its decisions on the state vector  $S_{jt} = (P_{j,t-1}, A_{jt}, P_t, W_t, W_t^*)$  for numerical convenience. Let  $V^N$  and  $V^A$  be the value functions conditional on not adjusting and conditional on having paid the menu cost. Then  $V = \max\{V^N, V^A - \kappa\}$  is the overall value function, and

$$\begin{aligned} V^N(S_{jt}) &= \Pi_{jt}(P_{j,t-1}) + E[Q(S_{j,t+1})V(S_{j,t+1}|S_{jt})], \\ V^A(S_{jt}) &= \max_{P_{jt}} \Pi_{jt}(P_{jt}) + E[Q(S_{j,t+1})V(S_{j,t+1}|S_{jt})], \end{aligned}$$

where  $Q$  is the one-period SDF, which is set equal to  $\beta$ .

All prices are in units of the domestic currency. The latter is assumed stable, so  $W_t = 1$ . The wage-based RER is  $E_t = W_t^*/W_t = W_t^*$ . A period is one month. The discount rate is 4%.  $\omega$  is calibrated to match the import share.  $\phi = 0$  and  $\phi^* = 0.75$ . The log RER follows a very persistent AR(1).  $\sigma = 5$  so the s.s. markup is 25%. A range of different values for  $\varepsilon$  is tried, with  $\varepsilon = 4$  being the baseline since it matches the LRPT in the data. The menu cost is 2.5% of revenues, conditional on adjustment, in s.s.

The LRPT produced by the model is estimated as in the empirical section. There is a strong positive relationship between frequency and LRPT for importers. The relationship is essentially flat for domestic firms (the only reason why it's not entirely flat is the indirect effect working through the price index). Both regressions used in the empirical section produce estimates that are close to the infeasible counterpart (which uses actual MC).

Varying  $\varepsilon$  generates substantial variation in the frequency and LRPT. The dispersion of frequencies is not as large as in the data, though. The LRPT depends too steeply on the frequency relative to the data. Varying  $\phi$  generates a lot of variation in the LRPT, unsurprisingly, but not much in the frequency. Varying  $\kappa$  generates a ton of variation in the frequency, of course, but almost no change in the LRPT. The authors now try to let both  $\varepsilon$  and  $\kappa$  vary independently across firms according to some distributions. This sufficiently dilutes the steepness in the frequency-LRPT relation, and it ensures that frequency and size of adjustments are uncorrelated.

Finally, the authors find that a Calvo model with large exogenous differences in the probability of adjustment can't induce a sufficiently positive correlation between frequency and measured LRPT. The Calvo model generates much slower adjustment than the menu cost model, so the low-frequency firms don't get to adjust enough following ER movements, and so the measured LRPT at, say, the 24-month horizon is non-zero. However, even this measured difference between high- and low-frequency firms is much too small.

### 9.3 Gopinath, Itskhoki and Rigobon (2010): “Currency Choice and Exchange Rate Pass-Through”

**Summary** In U.S. import data, there is a large difference between the ERPT of goods priced in dollars and of those priced in non-dollars, at both medium and long horizons. This is inconsistent with standard OE models in which the assignment of firms to LCP vs PCP is exogenous: Under this

assumption, in the short run the PT is 100% for PCP firms and 0% for LCP firms, but when prices adjust, ERPT is the same. In the data, non-dollar pricers adjust less frequently than dollar pricers. The authors construct a Calvo pricing model, with the option for firms to choose their currency when resetting prices. First they show analytically that, in a first-order log-linearization, a firm chooses PCP if the desired *medium-run PT* (i.e., the expected desired ERPT until the next price change) is high. This particular MRPT measure may be recovered from a simple regression of the cumulative price change since last adjustment on the increase in the NER, *without* controlling for other variables. Hence, a sufficient statistic for optimal currency choice is available and estimable, regardless of the source of incomplete PT. On the other hand, LRPT is a less useful statistic in this regard. The model is analyzed numerically, assuming variable markups and imported inputs in production. Both Calvo and menu cost pricing is considered. A firm selects PCP when the elasticity of its markup is low and the share of imported inputs is high (i.e., when the sensitivity to the ER is high).

**Empirics** Same data set as in Gopinath and Rigobon (2008). Include only countries that have a significant share of exports to the U.S. priced in non-dollars. This leaves only richer European countries, Japan and Canada.

They first present aggregate evidence. For every country, two price change series are constructed, one for dollar goods and the other for non-dollar goods. Then for each of these series, the price change is regressed on lagged NER changes, lagged foreign inflation and lagged U.S. GDP growth. The PT is defined as the sum of the coefficients on NER lags. First all the countries are pooled. The contemporaneous PT is close to zero for dollar goods and close to 1 for non-dollar goods, consistent with price stickiness. However, the PT only gradually rises up to 0.2 in the dollar subsample as the horizon is increased to 24 months. This pattern is found in all the individual country series, except for Canada where the two PT elasticities intersect. The pattern also holds up using individual good panel regressions.

The authors go on to estimate micro-level PT conditional on a price change, i.e., for every price change, the change in price is regressed on the cumulative NER change over that period and controls. The PT measured in this way is 0.24 for dollar goods and 0.92 for non-dollar goods. The results are similar in the subsample of differentiated goods. When the exercise is carried out on separate sectors, there is substantial variation in the results, but 18 of the 19 sectors that have a mix of dollar and non-dollar items show a larger ERPT for the latter items.

It is also possible to measure the lifelong PT, i.e., using the change in price and ER over the life of each good in the sample. Life-long PT is 0.49 for dollar goods and 0.98 for non-dollar ones. Hence, PT for dollar goods is about twice as high in the long run as for the first round of adjustment.

Non-dollar pricers have lower price adjustment frequency (0.07) than dollar pricers (0.10). This partly reflects that non-dollar pricers are mostly in the differentiated sector. The median size and absolute size of price changes is similar regardless of currency.

**Model** Partial equilibrium environment. State vector  $s_t$ . A single firm that exports into the U.S. decides which price  $p_t$  to set. The *desired* (flexible) log price is

$$\tilde{p}(s_t) = \arg \max_p \Pi(p|s_t).$$

The desired log price in the producer's currency is then  $\tilde{p}_t^* = \tilde{p}_t - e_t$ .  $e_t$  is assumed to follow a random walk for now. Each period, there's a probability  $(1 - \theta)$  that the firm is allowed to adjust its price and its currency of pricing. The value to the firm of LCP (resp., PCP) is  $V_L(p|s_t)$  (resp.,  $V_P(p^*|s_t)$ ), where

$$V_L(p|s_t) = \Pi(p|s_t) + \delta\theta E_t[V_L(p|s_{t+1})] + \delta(1 - \theta)E_t[V(s_{t+1})],$$

$$V_P(p^*|s_t) = \Pi(p^* + e_t|s_t) + \delta\theta E_t[V_P(p^*|s_{t+1})] + \delta(1 - \theta)E_t[V(s_{t+1})].$$

The optimal local (resp., producer) currency price is denoted  $\bar{p}_{L,t}$  (resp.,  $\bar{p}_{P,t}^*$ ). Then

$$V(s_t) = \max\{V_L(\bar{p}_L(s_t)|s_t), V_P(\bar{p}_P^*(s_t)|s_t)\}.$$

Proposition 1 shows that, up to a first-order log-linearization, we have

$$\bar{p}_L(s_t) = (1 - \delta\theta) \sum_{l=0}^{\infty} (\delta\theta)^l E_t \tilde{p}(s_{t+l}), \quad \bar{p}_P^*(s_t) = (1 - \delta\theta) \sum_{l=0}^{\infty} (\delta\theta)^l E_t [\tilde{p}(s_{t+l}) - e_{t+l}],$$

so  $\bar{p}_L(s_t) = \bar{p}_P^*(s_t) + e_t$  (due to the RW assumption). So the reset price will be the same (converted to the same currency), regardless of LCP vs PCP. Hence, if firms with the same underlying characteristics are assigned a currency exogenously, the ERPT should be the same conditional on the first adjustment (the proof does not rely on the currency choice being endogenous).

The difference between the LCP and PCP value functions are

$$\mathcal{L}(s_t) = V_L(\bar{p}_L(s_t)|s_t) - V_P(\bar{p}_P^*(s_t)|s_t) = \sum_{l=0}^{\infty} (\delta\theta)^l E_t [\Pi(\bar{p}_L(s_t)|s_{t+l}) - \Pi(\bar{p}_P^*(s_t) + e_{t+l}|s_{t+l})],$$

which may be seen from the Bellman equations. In Proposition 2, this is approximated to second order to arrive at

$$\mathcal{L}(s_t) \approx \frac{-\partial^2 \Pi(\tilde{p}(s_t)|s_t)}{\partial p^2} \sum_{l=0}^{\infty} (\delta\theta)^l \text{var}_t(e_{t+l}) \left[ \frac{1}{2} - \frac{\text{cov}_t(\tilde{p}(s_{t+l}), e_{t+l})}{\text{var}_t(e_{t+l})} \right].$$

Hence, LCP is chosen iff.

$$\bar{\Psi} \equiv (1 - \delta\theta)^2 \sum_{l=1}^{\infty} (\delta\theta)^{l-1} l \frac{\text{cov}_t(\tilde{p}(s_{t+l}), e_{t+l})}{\text{var}_t(e_{t+l})} < \frac{1}{2}.$$

The sufficient statistic on the LHS is termed the *MRPT*. It is a weighted average of ERPT coefficients into desired prices. We see that currency choice does not just depend on LRPT or the desired PT on impact, but on a weighted average of the entire path of PTs until the next adjustment. LRPT is determined by the characteristics of the profit function, not on price setting dynamics.

The authors put some more structure on the problem. First they write

$$\tilde{p}_t \equiv \tilde{p}(e_t, P_t|z_t) = \mu(\tilde{p}_t - P_t|z_t) + mc^*(e_t|z_t) + e_t,$$

where  $P_t$  is the log sectoral price level. They show in Proposition 3 that if  $\phi_t \equiv \partial(mc_t^* + e_t)/\partial e_t$  and  $\Lambda_t \equiv -\partial\mu_t/\partial(p_t - P_t)$  are constant, and the log sectoral price level follows an AR(1) process, then the covariance terms in  $\bar{\Psi}$  do not depend on  $s_t$ . Hence, currency choice is once-and-for-all.

Corollary 4 shows that if the desired profile of PT to ER shocks is increasing, then the firm is more likely to choose PCP if it has a longer price duration (as in the data) or an everywhere higher desired PT profile.

Proposition 5 shows that under the simplifications above, the population micro-level regression of the price change since last adjustment  $\bar{p}_{L,t} - \bar{p}_{L,t-\tau}$  on  $\Delta e_t$  yields the coefficient  $\bar{\Psi}$  (for PCP the LHS variable should be  $\bar{p}_{P,t}^* + e_t - \bar{p}_{P,t-\tau}^* - e_{t-\tau}$ ). This is exactly the MRPT regressions carried out in the empirical sections, except that there they used the cumulative ER change  $e_t - e_{t-\tau}$  on the RHS. They argue that using the latter should result in less noisy estimates.

**Simulations** The model is now parameterized and solved numerically, without imposing some of the simplifications needed for the analytics. The general message is that the analytical conclusions hold up in the richer set-up.

The model incorporates variable markups and imported intermediate inputs. The demand schedule is the Kimball-type

$$q = [1 - \varepsilon(p - P)]^{\sigma/\varepsilon},$$

such that the elasticity is  $\sigma/[1 - \varepsilon(p - P)]$  and superelasticity is  $\varepsilon/[1 - \varepsilon(p - P)]$ . The steady state ( $p = P$ ) markup elasticity is  $\Gamma = \varepsilon/(\sigma - 1)$ . The log MC are  $mc_t = \phi e_t - a_t$ .  $\phi$  is thought of as the share of domestic inputs in production (for the exporter). In the Calvo version of the model, the firm updates price and currency with probability  $1 - \theta$ . In the menu cost version, the firm can pay cost  $\kappa$  to update.

Both  $a_t$  and  $e_t$  follow persistent AR(1) processes.  $\theta$  and  $\kappa$  are calibrated to match the empirical price durations. The sectoral price level follows an AR(1)

$$(P_t - \bar{P}) = \alpha(P_{t-1} - \bar{P}) + (1 - \alpha)\bar{\phi}e_t, \quad \bar{P} = \log[\sigma/(\sigma - 1)].$$

Hence,  $\bar{\phi} = 0.5$  measures the aggregate LRPT.  $\alpha = 0.95$ . The steady-state elasticity of demand is set to  $\sigma = 5$ .  $\varepsilon$  and  $\phi$  are varied on  $[0, 8]$  and  $[0.5, 1]$ , respectively.

The LRPT is computed as the desired ERPT given the long-run response of  $P_t$ . With  $\phi$  fixed, both the MRPT and the LRPT are decreasing in  $\Gamma$ , but MRPT decreases faster. It is found that the cut-off of 0.5 for the MRPT is a good rule of thumb for determining the optimality of LCP vs PCP (i.e., the comparison of value functions gives approximately the same result). See Figure 3. Keeping  $\varepsilon$  fixed, it is found that both MRPT and LRPT increase with  $\phi$ , at about the same rate. Figure 5 shows the region in  $(\Gamma, \phi)$  space in which PCP is optimal; this is the case when  $\Gamma$  is low and  $\phi$  high. It is found that the empirical MRPT micro-level regressions provide good estimates of the theoretical MRPT  $\bar{\Psi}$  (despite the fact that the ER isn't exactly a RW here).

The authors finally discuss whether there is a way to reconcile *exogenous* currency choice with the data. If firms were backward-looking, i.e., don't set prices too far from past prices, there would be differential PT conditional on price adjustment, as in the data. However, the LRPT should still be the same for similar firms. Furthermore, in the data there is distinct sorting of non-dollar pricers into certain sectors. This could be for institutional reasons, but then one would have to explain why other sectors have a mix of dollar and non-dollar pricers.

## 9.4 Gopinath and Rigobon (2008): “Sticky Borders”

**Summary** BLS data on individual imported and exported U.S. goods show that prices are sticky with an average duration of about a year. Stickiness is quite heterogeneous. 90% of U.S. imports

and 97% of exports are priced in dollars. Prices contracted in foreign currency are only slightly more sticky than dollar prices. There has been a trend decline in the probability of price adjustment for imports; this holds across different kinds of goods, so the effect is largely within-sector and not compositional. ER pass-through into U.S imports is low, 20–30%, even when one conditions on a price change. Goods with a lower frequency of price adjustment display higher pass-through. There is evidence that most prices in the data set are allocative and that quantities are more flexible in contracts: Many exporting firms specify that the country of destination is not price-determining, and when asked whether the price is specific to a quantity, most firms answer the default “no.”

**Empirics** The data set is an unpublished survey of U.S. import and export prices for individual goods that the BLS conducts to construct price indices. Price data is collected monthly, 1993–2005. It is voluntary for firms to respond. The authors only include price quotes for actual transactions and further leave out intrafirm transactions.

The frequency of price changes is calculated in various ways, depending on the adjustment for goods substitution and discontinuation. The trade-weighted median duration for imports varies between 11–14 months and 11–16 for exports. Heterogeneity in stickiness is relatively large; for imports (exports) the weighted median monthly frequency of price adjustment is 9% (7.5%), the weighted mean is 28.8% (23%) and standard deviation 27.4% (26%). The authors classify goods into three categories: organized exchange traded goods, reference priced goods (whose prices are quoted in trade magazines) and differentiated goods. The median frequency of price change is highest for exchange goods (83%), then reference (30%), then differentiated (7%). This accords with the prediction from menu cost models where the cost of not adjusting is increasing in the elasticity of demand. However, goods traded on organized exchanges are probably also subject to more speculation.

The authors note that the observed frequency of price adjustment accords well with the literature on U.S. retail prices (adjusted for sales) as well as producer prices.

The authors also focus on episodes in which the foreign currency depreciated by more than 15% in one month. The change in price durations for imports from that country is negligible, although the frequency does increase a tiny bit right after the depreciation.

The average probability of import price changes decreased by about 10 percentage points to 22% in 2004. The decline is mostly concentrated in the 1990s and is across the board for all types of goods and all countries of origin.

There are three reasons why exchange rate pass-through may be incomplete: (1) if firms have market power, they absorb part of an ER-driven cost increase by decreasing their markups; (2) if marginal costs are increasing, a price raise lowers MC, which lowers the desired price; (3) some fraction of a firm’s inputs may be imported from countries whose currencies aren’t affected by the ER movement. The authors measure pass-through as the change in price in response to the cumulative change in the ER since the last price change (i.e., conditional on a price change). Specifically, they regress the price change on the cumulative ER change, the change in MC (proxied by foreign CPI inflation) and controls (U.S. GDP and U.S. CPI inflation). They also include sector and country fixed effects. The estimation is carried out on 20 different goods categories. The average pass-through is 21%, and except for one, all sectors have pass-through below 33%. When they sort goods into ten quantiles based on price stickiness, they find that more price-sticky goods display higher pass-through. This is because differentiated goods have high stickiness and also a lower demand elasticity, making a large price adjustment more attractive for firms, conditional on

adjustment.

## 10 Open Economy Models with Nominal Rigidity

### 10.1 Chari, Kehoe and McGrattan (2002): “Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?”

**Summary** Develops two-country OE model with complete markets, LCP, monetary shocks and nominal rigidity via staggered price setting à la Taylor. The main objective is to explain the volatility and persistence of the RER. The benchmark parameterization has separability of C and leisure in the utility function and relatively high risk aversion. Through the Backus-Smith equation, the volatility and persistence (first-order autocorrelation) of the RER increases with the CRRA for a given volatility and persistence of C. It is found that the benchmark parameterization can fit the volatility of the RER but the persistence is too small, the *persistence anomaly*. Due to the B-S relation, the model’s RER is highly positively correlated with relative consumption, which goes against the data, the *consumption-RER anomaly*. A number of fixes to these anomalies are attempted. Non-separability in C and L in the utility function (a C-D-type specification) makes the persistence anomaly worse. Adding shocks to productivity and government purchases does not do much quantitatively. Adding nominal wage stickiness through monopolistic unions and Taylor-type contracting generates a bit more persistence but not enough. Making asset markets incomplete by restricting the menu of assets to an uncontingent nominal bond does not make a quantitative difference. Finally, adding habit persistence in consumption would in principle help, but simple calculations suggest that it doesn’t do nearly enough.

**Empirics** The authors use quarterly data on bilateral exchange rates between the U.S. and individual European countries as well as their aggregate. Both the nominal and real ERs are very volatile (more than 4 times more volatile than output) and they are almost perfectly correlated. The relative price ratio is quite smooth. A decomposition à la Engel (1999) indicates that the relative price of traded goods accounts for almost all the variation in the RER, whereas the relative price of nontradables to tradables contributes insignificantly. Hence, the authors abstract from nontradables in their model.

**Model** Two countries, complete markets. In each country, a nontraded final good is produced by competitive producers. Intermediate goods producers are monopolistically competitive and can price-discriminate across countries but must set prices in the *local* currency. Prices are fixed for  $N$  periods (see below).

The state of the world is  $s_t$  and the full history of shocks is denoted  $s^t = (s_0, \dots, s_t)$ . The commodities are L, a C/K good, money, a continuum of intermediate goods indexed by  $i \in [0, 1]$  in each country, and final goods. Final goods are produced using the aggregator

$$y(s^t) = \left[ a_1 \left( \int_0^1 y_H(i, s^t)^\theta di \right)^{\rho/\theta} + a_2 \left( \int_0^1 y_F(i, s^t)^\theta di \right)^{\rho/\theta} \right]^{1/\rho},$$

where  $y_H$  and  $y_F$  are intermediate goods from Home and Foreign, resp. The parameter  $\theta$  will determine the markup,  $\rho$  (and  $\theta$ ) determine the elasticity of substitution between Home and Foreign

goods, while  $a_1$  and  $a_2$  (and  $\rho, \theta$ ) will determine the import shares. Final goods producers are competitive and maximize profits

$$\max P(s^t)y(s^t) - \int_0^1 P_H(i, s^{t-1})y_H(i, s^t)di - \int_0^1 P_F(i, s^{t-1})y_F(i, s^t)di$$

subject to the production function. The prices of intermediate goods can only depend on  $s^{t-1}$  because intermediate goods producers set prices before observing the shock. The FOCs lead to demand functions for Home and Foreign intermediate goods. The zero-profit condition determines the price  $P(s^t)$  of the final good in terms of the prices of intermediate goods; since the latter only depend on  $s^{t-1}$ , so does  $P$  in equilibrium.

The technology for producing intermediate goods is

$$y_H(i, s^t) + y_H^*(i, s^t) = F(k(i, s^{t-1}), l(i, s^t)),$$

where  $F$  exhibits CRS. Investment is denoted  $x(i, s^t)$ . There are convex adjustment costs of capital. Intermediate goods producers set prices for  $N$  periods in a staggered way: a (deterministically determined) fraction  $1/N$  get to reset every period. They choose prices  $P_H(i, s^{t-1})$  and  $P_H^*(i, s^{t-1})$ ,  $L$ ,  $K$  and  $I$  to maximize profits given the common SDF  $Q$ .

Contingent, one-period nominal bonds denominated in the Home currency are traded. Residents of each country own that country's firms. Consumers choose their allocations after observing the shock  $s_t$ . They face the budget constraint

$$P(s^t)c(s^t) + M(s^t) + \sum_{s_{t+1}} Q(s^{t+1}|s^t)B(s^{t+1}) \leq P(s^t)w(s^t)l(s^t) + M(s^{t-1}) + B(s^t) + \Pi(s^t) + T(s^t),$$

where  $\Pi$  are profits and  $T$  transfers of home currency from the monetary authority. They are also subject to a lower bound on borrowing to rule out Ponzi schemes. The utility function depends on  $c$ ,  $l$  and real money balances  $M/P$ . The labor-leisure, RMB and bond FOCs are

$$\begin{aligned} -\frac{U_{l,t}}{U_{c,t}} &= w_t, \\ \frac{U_{m,t}}{P_t} - \frac{U_{c,t}}{P_t} + \beta \sum_{s_{t+1}} \pi(s^{t+1}|s^t) \frac{U_{c,t+1}}{P_{t+1}} &= 0, \\ Q(s^t|s^{t-1}) &= \beta \pi(s^t|s^{t-1}) \frac{U_{c,t}}{U_{c,t-1}} \frac{P_{t-1}}{P_t}. \end{aligned}$$

Similar expressions hold for Foreign. The bond Euler equations for Home and Foreign may be iterated back to time 0 and combined to yield

$$\frac{U_{c,t}}{U_{c,0}} \frac{P_0}{P_t} = \frac{U_{c,t}^*}{U_{c,0}^*} \frac{e_0}{e_t} \frac{P_0^*}{P_t^*},$$

which yields the B-S condition

$$q_t \equiv e_t \frac{P_t^*}{P_t} = \kappa \frac{U_{c,t}^*}{U_{c,t}}. \quad (5)$$

The money supply process for Home is  $M(s^t) = \mu(s_t)M(s^{t-1})$ , where  $\mu$  is a Markov process. New money balances are distributed lump sum to consumers,  $T(s^t) = M(s^t) - M(s^{t-1})$ . Similar for Foreign.

The resource constraint and labor market clearing condition in Home require

$$y(s^t) = c(s^t) + \int_0^1 x(i, s^t) di, \quad l(s^t) = \int_0^1 l(i, s^t) di.$$

Market clearing for bonds requires  $B(s^t) + B^*(s^t) = 0$ .

To solve the model, all nominal variables are deflated by the price level. The equations are then log-linearized.

In terms of calibration, the utility function is picked to be

$$U(c, l, M/P) = \frac{1}{1-\sigma} \left[ \omega c^{(\eta-1)/\eta} + (1-\omega)(M/P)^{(\eta-1)/\eta} \right]^{(1-\sigma)\eta/(\eta-1)} + \psi \frac{(1-l)^{1-\gamma}}{1-\gamma}.$$

It is therefore separable in C and L. The production function is  $F(k, l) = k^\alpha l^{1-\alpha}$ . They set  $\sigma = 5$  and  $\gamma = \sigma$  to get balanced growth. The FOC for RMB can be shown to imply a standard money demand function that the authors estimate from the data to calibrate parameters. The elasticity between Home and Foreign goods is  $1/(1-\rho)$ , which is calibrated to 1.5. They pick  $\theta = 0.9$ , which gives 11% markup in s.s. They set  $N = 4$  so prices are fixed for a year. The growth rate of the money stock follows an AR(1) in logs. It is estimated from U.S. data (for both countries). The standard deviation of shocks is chosen to match the output volatility in the data.

**Simulations** The benchmark model generates nominal and real ERs that are volatile, persistent and highly cross-correlated. However, the persistence and cross-correlation is still too small to match the data. RERs and relative C are highly positively correlated. The price ratio is too volatile and I isn't volatile enough.

To understand the relationships, the authors log-linearize the B-S condition (5) to obtain

$$\hat{q} = A(\hat{c} - \hat{c}^*) + B(\hat{m} - \hat{m}^*) + D(\hat{l} - \hat{l}^*), \quad (6)$$

where

$$A = -\frac{cU_{cc}}{U_c}, \quad B = -\frac{mU_{cm}}{U_c}, \quad D = -\frac{lU_{cl}}{U_c},$$

all evaluated at the s.s. Quantitatively,  $A \approx \sigma = 5$  (the only reason it's not exactly  $\sigma$  is due to nonseparability between C and RMB),  $B \approx 0$  and  $D = 0$  (separability). Thus,

$$\frac{\text{std}(\hat{q})}{\text{std}(\hat{y})} \approx \sigma \frac{\text{std}(\hat{c} - \hat{c}^*)}{\text{std}(\hat{y})}, \quad \text{corr}(\hat{q}, \hat{q}_{-1}) \approx \text{corr}(\hat{c} - \hat{c}^*, \hat{c}_{-1} - \hat{c}_{-1}^*).$$

Numerical experiments show that a  $\sigma$  of about 5 is needed to get the right relative volatility of the RER. Because C isn't very persistent in the model (in particular, less persistent than in the data), the RER in the model doesn't achieve sufficient persistence. Persistence increases as price stickiness  $N$  increases, but an unrealistic increase is necessary to match the data.

I is not volatile enough in the model, because a relatively high adjustment cost is needed to make C volatile. L is more volatile in the model than in the data because almost all the volatility in Y has to come from changes in L (since K moves sluggishly and there are no technology shocks). The cross-country correlations of Y and C are the same in the model, while in the data the Y one is highest. Finally, NX are acyclical in the model but somewhat countercyclical in the data.

**Extensions** The model is extended in various ways to try to fix the anomalies. First, another C-D-type preference specification is attempted, so that there is complete nonseparability of C, RMB and L. This generates less volatility in the RER. The reason is that under such a specification (and  $\sigma > 1$ ), we have  $U_{cl} > 0$ , so when a monetary shock raises C and L,  $U_c$  doesn't decrease as much as it would if  $U_{cl} = 0$ , so movements in L offset movements in C, and the RER (which is a ratio of  $U_c$ ) stays relatively constant. In terms of relation (6),  $D$  is now negative and quantitatively significant.

The authors instead add two additional shocks. One is a country-specific technology process whose log follows an AR(1). The other is a shock to government purchases  $g_t$  that now enter into the economy-wide resource constraint. It also follows an AR(1) in logs. The AR(1)'s are estimated from the data. The quantitative results are not very different from the benchmark case, except that relative L volatility actually increases a bit (due to technology shocks).

If the money growth rate rule is replaced with a Taylor interest rate rule, the endogenous policy reaction tends to offset exogenous shocks so much that C, and thus the RER, is too smooth.

The authors try introducing a labor market friction through sticky wages. Consumers are now assumed to belong to one of a continuum of unions that each supply a differentiated labor good. Intermediate goods firms produce using a Dixit-Stiglitz composite of all the domestic labor goods. They choose the cost-minimizing composition of different types of labor. A fraction  $1/M$  of unions set their wages each period, and the wage is set for  $M$  periods at a time. The consumers' labor-leisure condition is then replaced by an optimality condition for nominal wages. The authors calibrate  $M = 4$ . The model improves a bit on the benchmark due to the additional friction but more is needed.

Two approaches to dealing with the C-RER anomaly are attempted. First, the set of assets is restricted to only non-contingent one-period bonds. The B-S condition is then replaced with

$$E_t[\hat{q}_{t+1} - \hat{q}_t] = E_t[(U_{c,t+1}^* - U_{c,t+1}) - (U_{c,t}^* - U_{c,t})].$$

The quantitative effect of moving to incomplete markets is tiny, however.

The second approach is to add habit persistence in C by replacing  $c_t$  in the utility function with  $c_t - d\bar{c}_{t-1}$ , where  $\bar{c}_{t-1}$  is last period's aggregate consumption. Then condition (6) becomes (with  $D = 0$ )

$$\hat{q}_t = \frac{A}{1-d}[(\hat{c} - \hat{c}^*) - d(\hat{c}_{t-1} - \hat{c}_{t-1}^*)] + B(\hat{m} - \hat{m}^*).$$

The authors use data for the U.S. and Europe to compute the correlation between the RHS and the actual RER.  $d$  was varied between  $-1$  and  $1$ . The maximum attainable correlation was  $-0.19$ . Hence, it seems that habit persistence will not help in matching the data.

The authors call for more work on introducing asset market frictions to break the link between RERs and MUs.

## 10.2 Devereux and Engel (2003): “Monetary Policy in the Open Economy Revisited: Price Setting and Exchange-Rate Flexibility”

**Summary** The traditional argument for the optimality of floating exchange rates relies crucially on firms setting prices in their domestic currencies, i.e., the assumption of PCP. Under PCP (and nominal rigidities), following a bad real shock to Home, the ensuing depreciation of its currency leads to expenditure switching towards Home's goods, which mitigates the adverse effect of the original shock. However, if instead export prices are set in local currency (LCP), the NER has

no effect on relative goods prices while the stickiness lasts. Thus, expenditure switching plays no role. The authors develop a simple two-country, stochastic, complete markets model that can be solved (almost) analytically. Prices are set in each period before the realization of shocks. The governments in each country choose the money supply to maximize welfare for domestic consumers. Under PCP, governments are able to replicate the first-best flexible price solution by influencing the NER to generate sufficient relative price movements and thus expenditure switching. Under LCP, governments are unable to generate relative price movements; the optimum features perfect risk-sharing but lower expected consumption due to the inefficient price movements. The optimal policies turn out to be invariant to whether a Nash solution concept is employed or whether countries act cooperatively.

**Motivation** PCP has been the usual assumption in OE macro. There is a well-established fact that an appreciation of a country's NER tends to be associated with an improvement in the ToT. Keynesian reasoning suggests this is consistent with PCP but not LCP (Obstfeld and Rogoff, 2000): The ToT is given by

$$ToT = \frac{SP_H^*}{P_F},$$

where  $S$  is the NER in Home currency per Foreign currency,  $P_H^*$  is the Foreign currency price of Home exports and  $P_F$  is the Home currency price of Home imports. Suppose  $S$  drops, i.e., the Home currency appreciates. If there is PCP and price stickiness, then  $SP_H^*$  and  $P_F/S$  are constant in the short run, so *ToT increases* one-for-one with the drop in  $S$ , i.e., the ToT improve (and Foreign experiences a favorable expenditure switching as their export goods become relatively cheaper). However, under LCP,  $P_H^*$  and  $P_F$  are both fixed in the short run, so the ToT deteriorate.

Devereux, Engel and Tille (2003) has shown that this conclusion can be circumvented by introducing importing intermediaries who set prices in local currencies.

**Model set-up** Two countries, monopolistic firms, shocks to velocity (which multiplies utility derived from real money balances) and production technology. Producers set prices each period but before the realization of the shocks. They maximize profits using the consumers' common SDF. Consumers are infinitely lived and have utility that's separable in C, RMB and L. Markets are complete and pay off in currencies, not goods (otherwise the distinction between PCP and LCP would be inconsequential, as households could circumvent international price discrimination). The government supplies households with direct transfers of money. The production function is linear in L. Velocity and TFP follow RWs with i.i.d. independent normal innovations.

The money demand function implied by the consumers' preferences is not of a nice log-linear form, so it is log-linearized. This is the only approximation used in solving the model. Since all other relations are nicely multiplicative, logs can be taken and the log-normal distribution for velocity and productivity may be used. Due to complete markets, the B-S condition holds.

Under PCP, firms choose a single price in their domestic currency. It is not just given by a constant markup over unit costs, since there is an additional risk premium from the covariance of profits with consumption. The LOP holds.

Under LCP, the firm price discriminates between the Home and Foreign market.

**Policy assumptions and model solution** Under *flexible prices*, consumption is equalized across countries and solely determined by Home and Foreign productivity shocks.

As for government policy, it is assumed that the government controls the money supply process, i.e., picks the innovation. This is done after having seen the current-period shocks. It is assumed that the government only responds to unanticipated shocks so that shocks to the money supply are also unanticipated. Apart from a time trend that is invariant to policy, the decision rules of all agents are stationary due to the foregoing assumptions. Hence, the government can be thought of as maximizing the ex ante expected *per-period* utility. The simplifying assumption that the government doesn't care about the consumers' utility derived from RMB is imposed. Only utility from C and L is counted in the optimization. Finally, it is assumed that government can *commit* to their policy in advance.

Under *PCP*, PPP holds, so consumption is equalized across countries according to the B-S condition. The government's objective function turns out to be quadratic, while the optimal policy (i.e., choice of money innovation) can be written as a linear function of the exogenous model shocks. Because Home and Foreign consumption is identical state by state, it doesn't matter whether the policy game is construed as a Nash equilibrium or a cooperative solution. Proposition 1 says that the governments can and will implement the flexible price solution. Intuitively, there are three departures from efficiency in this model: sticky prices, monopolistic pricing and a departure from the Friedman rule of zero nominal interest rates. The second inefficiency can't be affected by policy. The government is assumed not to care about the third one. Hence, policy solely focuses on eliminating the adverse effects of nominal rigidity. This is done by steering the money supply such that the NER causes an adjustment in the ToT that generates efficient expenditure switching.

Operationally, the Home government expands the money supply after a positive domestic productivity shock to make demand for goods meet supply (money demand has to rise so C rises); this raises the Home price level, so through the LOP the NER depreciates. Optimal policy also responds to Foreign productivity shocks, because after a positive shock to Foreign, the Foreign government generates a NER appreciation (from Home's point of view), a fall in Home's price level, an increase in RMB and thus a rise in Home consumption (there's an additional effect through the nominal interest rate's effect on the demand for RMB); this must be counteracted by a reduction in Home's money supply. Finally, Home velocity shocks are perfectly accommodated to cancel their effect on the NER (since the flexible price allocation is independent of velocity).

Under *LCP*, PPP does not hold. By assumption, the price level in each country is predetermined in every period. There is no transmission of Foreign money supply or velocity innovations to Home consumption; in particular, the NER is not transmitted to the price level. However, both Home and Foreign productivity shocks influence Home consumption through the nominal interest rate (determined by the contingent assets). Since consumption is not necessarily equated across countries, the objective functions for the two governments differ. Proposition 2 shows that optimal policy in a Nash solution leaves the NER *constant*. This is because the ToT can't be affected. Policy makers can influence welfare only by affecting the aggregate consumption indexes (i.e., the *global* demand for Home and for Foreign goods, respectively). In the flexible price equilibrium, aggregate consumption is a function of only world productivity shocks, and optimal policy under LCP replicates this. In particular, consumption is equalized across countries, which is ensured by keeping the NER and thus the RER constant. However, *relative* consumption of Home and Foreign produced goods can't be kept at the first best and thus neither can employment. Due to the sub-optimal substitution between Home and Foreign goods, expected consumption is lower than under PCP.

It is also true under LCP that the cooperative solution is the same as the Nash solution. Here

the reason is that there is effectively no strategic element to government policy, as Foreign policy doesn't influence the prices that Home faces.

### 10.3 Engel (2011): “Currency Misalignments and Optimal Monetary Policy: A Reexamination”

**Summary** Previous theoretical studies of optimal monetary policy have either assumed that the LOP holds or that prices are only fixed one period, meaning that inflation can't cause price dispersion within countries. Engel seeks to evaluate the (cooperative) policy tradeoff between inflation, the output gap and currency misalignment (defined as the average deviation between prices paid by Foreign consumers compared to those paid by Home consumers for the same bundles). The latter can lead to inefficiencies even when inflation and the output gap are zero, since it implies that consumers in different countries don't face the same prices and so allocations are inefficient. A fairly general two-country model with complete markets is developed, and it's shown that a second-order approximation to welfare may be derived without assuming anything about how prices or wages are set. If the LOP does not hold, currency misalignment enters the welfare expression. The model is specialized to involve Calvo pricing with no cross-country price discrimination when prices are initially set. There are exogenous shocks to markups so the inflation-output tradeoff is nontrivial. Under PCP, the LOP holds, and optimal policy trades off domestic price inflation and the output gap, as in the closed economy. Under LCP, the LOP doesn't hold, so optimal policy targets CPI inflation rather than domestic inflation. The NER no longer serves a useful allocative role. Instead, policy influences relative final consumption in the two countries to target the currency misalignment. Unlike under PCP, when the cost-push shock is eliminated, optimal policy isn't able to replicate the first-best flexible price solution due to the relative price distortions.

**Model** Two countries of equal size, a continuum of consumers in each. Goods producers are monopolistically competitive. Each household supplies a unique type of labor to firms in its home country but are otherwise identical; hence, they set the same wage. Markets are complete.

Period utility is  $C^{1-\sigma}/(1-\sigma) - \frac{1}{1+\phi}N^{1+\phi}$ . Here  $C$  is a consumption aggregate over Home and Foreign goods bundles  $C_H$  and  $C_F$  with EoS  $\xi$  and home bias weight  $\nu/(2-\nu)$ ,  $\nu \in [0, 1]$ . The labor aggregate is  $N$  and wage  $W$ . Apart from wages, consumers each receive a share of aggregate profits from Home firms. The production function is linear in labor input with productivity  $A_t$  that is stochastic and common across firms in a country. The labor aggregate is a Dixit-Stiglitz aggregator with stochastic EoS  $\nu_t$ . A fraction  $\tau_t$  of the wage is subsidized by the government (taxes are lump sum).

The ToT are defined as  $S_t = P_{Ft}/P_{Ht}$ ,  $S_t^* = P_{Ht}^*/P_{Ft}^*$ , where for example  $P_{Ht}^*$  is the Foreign-currency price of goods exported from Home, while  $P_{Ft}^*$  is the Foreign-currency price of goods produced in Foreign for local consumption. The Backus-Smith condition  $(C_t/C_t^*)^\sigma = \mathcal{E}_t P_t^*/P_t$  holds, where  $P_t$  is the consumption price aggregate in Home.

The model is log-linearized to derive approximate welfare. For some relations second-order linearizations are necessary. Key quantities are

$$m_t = \frac{1}{2}(e_t + p_{Ht}^* - p_{Ht} + e_t + p_{Ft}^* - p_{Ft})$$

and

$$z_t = \frac{1}{2}(p_{Ft} + p_{Ht}^* - p_{Ht} - p_{Ft}^*).$$

$m_t$  is called the *currency misalignment* and measures the average deviation of the price paid by Foreign consumers relative to that paid by Home consumers (with a weight of 1/2 for each country's bundle).  $z_t$  measures the average relative price of exported goods to imported goods. Note that if the LOP holds, both  $m_t = z_t = 0$ , but otherwise neither of  $m_t = 0$  or  $z_t = 0$  implies the other.

The policy maker wishes to minimize the expected PDV of  $X_t$ , where  $X_t$  is the difference between the maximal achievable consumer utility without inefficiencies and the market-determined utility. The loss function for cooperative policy makers is derived, i.e., the one relevant for world welfare. In general, it is a linear combination of  $(\tilde{y}_t^R)^2$  (the relative output gap),  $(\tilde{y}_t^W)^2$  (the world/total output gap),  $m_t^2$ ,  $z_t^2$ ,  $\sigma_{p_{Ht}}^2$  (the cross-sectional variance of  $p_H$ ),  $\sigma_{p_{Ht}^*}^2$ ,  $\sigma_{p_{Ft}}^2$  and  $\sigma_{p_{Ft}^*}^2$ . The dispersion of prices enters because it affects the allocation of labor. Note that if the LOP holds (e.g.,  $p_{Ht}(f) = p_{Ht}^*(f) + e_t$  for every firm  $f$ ), then  $\sigma_{p_{Ht}}^2 = \sigma_{p_{Ht}^*}^2$  because  $e_t$  doesn't vary cross-sectionally, and  $m_t = z_t = 0$ , so the welfare function simplifies. For the same reason, under PCP (and LOP) price dispersion can be eliminated simply by eliminating *producer* price dispersion. Under LCP (and not LOP), all four kinds of price dispersion enter separately.

$m_t$  and  $z_t$  enter into the welfare function in the general case because they measure the extent to which consumers in the two countries face different prices for the same goods. Any such price disparity results in inefficient allocations because private tradeoffs aren't perfectly matched to relative marginal ratios of substitution (i.e., a social planner would allocate a good according to the unique shadow price of the good in the aggregate resource constraint).

The author notes that to arrive at the welfare expression it was not necessary to assume anything about how prices or wages are set.

**Optimal policy** The model is now specialized so as to analyze optimal policy in a concrete framework. Wages are set flexibly by the households who exercise their monopoly power and set real wages at a markup  $1/(\eta_t - 1)$  over the labor wedge. The government sets a constant subsidy  $\tau$  that would eliminate the distortions from monopolistic behavior in the non-stochastic steady state.

Under *PCP*, firms reset prices with Calvo probability  $1 - \theta$ . When they do, they set a single price in their own currency and the LOP holds. Engel derives the resulting NKPCs for Home and Foreign producer price inflation  $\pi_{Ht}$  and  $\pi_{Ft}^*$ .

Under *LCP*, firms still reset with Calvo probability  $1 - \theta$ , but when they do, they choose both a local price in Home currency and an export price in Foreign currency. The Home-currency value of the export price is then subject to NER fluctuations until the next time prices are reset. Engel derives four NKPCs for each of the consumption bundles. Due to the symmetry of the model (and the assumption of no trade barriers), relative prices  $p_{Ft} - p_{Ht}$  and  $p_{Ft}^* - p_{Ht}^*$  in Home and Foreign are always equal, even though the LOP doesn't hold. Hence  $z_t = 0$ , and also  $\pi_{Ft} - \pi_{Ht} = \pi_{Ft}^* - \pi_{Ht}^*$ .

The world government chooses nominal interest rates in the two countries, but we can think of it as choosing inflation and output subject to the NKPCs. Optimal policies with and without commitment are analyzed, but the main conclusions of the paper are the same regardless of the credibility of policy.

Optimal policy depends on the price-setting regime.

- *PCP*. As in Galí and Monacelli (2005), optimal policy linearly trades off domestic (producer) inflation and the domestic output gap. The relative weight is given by the EoS among goods produced in the country (the higher the EoS, the higher the weight on inflation). Home bias in preferences plays no role. The NER plays a large role as it makes up for the stickiness in producer prices. Thus, it is actually beneficial to generate movements in the non-domestic

component of CPI inflation (import prices) to influence relative consumption of different bundles.

If there are no markup (cost-push) shocks, policy can implement the first-best flexible price outcome thanks to NER flexibility.

- *LCP*. As noted above, we have  $z_t = 0$  in the model. The author makes the simplifying assumption that  $\phi = 0$  (linear disutility of labor). This makes the evolution of the ToT  $s_t$  independent of policy and each period's choice of output level does not constrain the choices for subsequent periods (MC are constant). Setting  $\phi = 0$  does not limit the comparability to the PCP case, as  $\phi$  doesn't enter into the optimal policy rule under that assumption.

When  $\phi = 0$ , policy can't influence internal relative prices as the NER doesn't affect them (if  $\phi \neq 0$ , policy could affect firms' MC and thus prices). Because the LOP doesn't hold, there is currency misalignment, which *can* be influenced somewhat by steering relative total consumption  $c_t - c_t^*$ . However, relative consumption of bundles from different countries, i.e.,  $c_H - c_F$  and  $c_H - c_H^*$  can't be influenced by policy. This remains true even if there are no markup shocks and only productivity shocks, so policy can't achieve the efficient outcome in this case; although the world output gap can be set to zero (and the RER to its efficient level), the country-specific output gaps can't.

In the LCP case, optimal policy rules involve CPI inflation rather than just domestic inflation. If the LOP held, price dispersion could be eliminated by only controlling domestic inflation, but this isn't the case under LCP. Engel's equation (42) shows that optimal policy under LCP can be thought of as trading off relative CPI inflation with inefficiencies in the change of the RER.

Finally, impulse responses are simulated for a calibrated version of the model without markup shocks. Under LCP, the ToT respond much more sluggishly to a productivity shock than under the first best. Because the right relative price movements aren't generated, output does not rise as much as potential and the output gap drops. The NER is less volatile under LCP than under PCP. Because optimal policy stabilizes CPI inflation fully under LCP (there are no cost-push shocks), the NER is perfectly correlated with the RER.

#### 10.4 Farhi, Gopinath and Itskhoki (2011): "Fiscal Devaluations"

**Summary** It's possible to replicate the real effects of a NER depreciation using revenue-neutral tax policy while holding the NER fixed, a so-called fiscal devaluation. While certain details of implementation vary depending on whether the devaluation is expected or not and on the completeness of asset markets, the policy recommendations are robust to the nature of price setting and other model details.

Two types of revenue-neutral fiscal policies can replicate the effects of a NER depreciation. Consider PCP. First, the government can impose an export subsidy and an equally sized import tariff, as this will have the desired effect on international after-tax export prices and thus the ToT. Second, the government can move the ToT by increasing the VAT (which is border-adjusted, i.e., rebated to exporters and levied on importers), while imposing a payroll subsidy so as to not distort labor demand. If markets are complete, international risk-sharing is affected by the RER. Any kind of devaluation works by making Home exports cheaper relative to Foreign exports. However, under the two above-mentioned fiscal policies, this is achieved by making Home consumption more

expensive by taxing imports, whereas under a NER depreciation, Home produced goods are made relatively cheaper. The response in the RER is therefore different (it appreciates under a fiscal devaluation), and to cancel the risk-sharing effects of this out, a consumption tax is needed.

Another effect of a NER depreciation is that the Foreign-currency value of Home-currency denominated debt is depleted. To replicate this without a NER depreciation, the Home government must partially default on its debt. All of the fiscal policies discussed require appropriate responses in the money stock. This can't be achieved in a currency union. Instead, fiscal devaluations can be achieved by having the union central bank transfer seignorage to regional governments.

The conclusions of the paper are illustrated both in a simple static model and a full-fledged NK model. The dynamic framework introduces a few more complications. If the devaluation is anticipated, interest rates will behave differently than if the devaluation is not anticipated, so required responses in the money supply differ. Under the tariff policy, it's necessary to introduce a tax on dividends because a time-varying consumption tax affects price setting through the SDF.

**Static model** The basic insights are demonstrated with a one-period model with two countries, Home and Foreign. Foreign follows a passive policy without taxation and fixed money  $M^*$ . There are two goods, one produced in each country. The utility of the Home representative consumer is

$$U = \frac{1}{1-\sigma} C^{1-\sigma} - \frac{\kappa}{1+\varphi} N^{1+\varphi}.$$

The consumption aggregator is C-D with a share  $\gamma$  for domestic goods. This allows for home bias. Due to constant consumption shares, market clearing can be written

$$Y = \gamma \frac{PC}{P_H} + (1-\gamma) \frac{P^* C^*}{P_H^*},$$

and similarly for Foreign production. Consumers are subject to the cash-in-advance constraint

$$\frac{PC}{1+\zeta_c} \leq M,$$

where  $\zeta_c$  is Home's consumption subsidy.  $P$  is the price of  $C$  exclusive of the consumption tax. Home households face the budget constraint

$$\frac{PC}{1+\zeta_c} + M + T \leq \frac{WN}{1+\tau^n} + \frac{\Pi}{1+\tau^d} + B^p,$$

where  $\tau^n$  is the labor-income tax,  $T$  is a lump sum tax and  $B^p$  are household NFA. The capital tax  $\tau^d$  plays no role in the static analysis. Note that last period's money holdings have been left out of the RHS due to the static nature of the model.

Home's government constraint is

$$M + T + TR + B^g \geq 0,$$

where  $B^g$  is the governments' NFA and  $TR$  stands for all non-lump-sum fiscal revenue (see footnote 10). Total NFA are  $B = B^p + B^g$ . International asset market clearing requires  $B + B^* = 0$ .

Consider first the model with PCP. Wages and prices are partially sticky in the beginning of the period before productivity shocks and policies are realized. Wages are set according to

$$W = \bar{W}^{\theta_w} \left[ \mu_w \frac{1+\tau^n}{1+\zeta^c} \kappa P C^\sigma \left( \frac{Y}{A} \right)^\varphi \right]^{1-\theta_w},$$

where  $\theta_w$  indexes wage stickiness,  $\bar{W}$  is the preset wage and the expression in the brackets is the flexible wage. Under PCP, the LOP holds, and producers set the domestic currency price

$$P_H = \bar{P}_H^{\theta_p} \left[ \mu_p \frac{1 - \zeta^p W}{1 - \tau^v A} \right]^{1 - \theta_p},$$

which is inclusive of the VAT. Here  $\bar{P}_H$  is the preset price and the expression in the brackets the flexible price.  $\zeta^p$  and  $\tau^v$  are the payroll subsidy and VAT, resp.

Exports are subsidized at rate  $\zeta^x$  and imports are subject to a tariff  $\tau^m$ . The VAT is reimbursed at the border for exports but levied on importers (VAT with border adjustment). Thus,

$$P_H^* = P_H \frac{1}{\mathcal{E}} \frac{1 - \tau^v}{1 + \zeta^x}, \quad P_F = P_F^* \mathcal{E} \frac{1 + \tau^m}{1 - \tau^v},$$

where  $\mathcal{E}$  is the NER (Home currency per Foreign; an increase is a depreciation).

The authors consider several degrees of openness of the capital account.

- Under *balanced trade* (financial autarky),  $P^* C^* = P_H^* C_H^*$ . Along with market clearing and the CIA constraint, this implies  $\mathcal{E} \propto M/M^*$ , where the constant of proportionality is determined by tax rates alone.
- Under *complete markets*, the Backus-Smith condition  $\lambda^{-1}(C/C^*)^\sigma = Q \equiv P^* \mathcal{E}/P(1 + \zeta^c)$  holds. Here  $Q$  is the CPI-based RER. The authors normalize  $\lambda = 1$ . Using the CIA constraint,  $\mathcal{E} = (M/M^*)Q^{(\sigma-1)/\sigma}$ .
- Under *arbitrary portfolios*, we write  $B = B^h + B^{f*} \mathcal{E}$ , where  $B^h$  are Home-currency assets and  $B^{f*}$  Foreign-currency ones. The expression for the NER is more complicated in this case, cf. their equation (14).

A *nominal devaluation* of size  $\delta$  is the outcome of an increase in the Home money supply so that  $\Delta \mathcal{E}/\mathcal{E} = \delta$  without a change in taxes. A *fiscal devaluation* of size  $\delta$  is a set of fiscal policies such that the allocation of C, L and Y is the same as under the nominal devaluation but while holding  $\mathcal{E}$  fixed.

Proposition 1 states that if trade is balanced or the country only holds foreign-currency NFA, a fiscal devaluation of size  $\delta$  can be attained in one of two ways:

1. Via an export subsidy and import tariff:  $\tau^m = \zeta^x = \delta$ ,  $\tau^c = \tau^n = \varepsilon$ ,  $\Delta M/M = (\delta - \varepsilon)/(1 + \varepsilon)$ .
2. Via a VAT increase and payroll subsidy:  $\tau^v = \zeta^p = \delta/(1 + \delta)$ ,  $\tau^c = \tau^n = \varepsilon$ ,  $\Delta M/M = (\delta - \varepsilon)/(1 + \varepsilon)$ .

In both cases,  $\varepsilon$  can be chosen arbitrarily. The proof proceeds by showing that this choice of taxes has the same effect on all equilibrium relations involving real variables as a change in  $\mathcal{E}$ . The intuition can be seen by considering the ToT

$$\mathcal{S} = \frac{P_F^*}{P_H^*} = \frac{P_F^*}{P_H} \mathcal{E} \frac{1 + \zeta^x}{1 - \tau^v},$$

using the LOP. Evidently, given relative domestic prices  $P_F^*/P_H$ , the required change in the ToT can be generated with either an export subsidy or VAT. To make sure that  $P_H$  and  $P_F$  don't

change more than required to, we must compensate with an import tariff (in the first policy) or a payroll subsidy (in the second policy). A consumption subsidy is not needed here but may be used, provided its effect on wages (through the labor wedge) is canceled out by an income tax. If it's used, a smaller expansion of the money supply is needed (cf. the CIA constraint).

If markets are complete, there is an additional effect through the RER. Fiscal and nominal devaluations make Home exports cheaper relative to Foreign exports. But nominal devaluations does this by making Home-produced goods cheaper, while fiscal devaluations makes Home consumption more expensive by taxing imports. Hence, the RER appreciates under a fiscal devaluation of the sort mentioned above, unless the consumption subsidy is used to decrease Home's CPI  $P/(1 + \zeta^c)$ . Under perfect risk-sharing, a RER movement leads to a reallocation of consumption across countries that doesn't happen under a nominal depreciation. Hence, if markets are complete, the above-mentioned two fiscal devaluation policies must be accompanied by  $\zeta^c = \tau^n = c$  (Proposition 2). The money supply must be expanded or contracted appropriately.

Finally, if the country has NFA denominated in Home currency, a NER depreciation is accompanied by a valuation effect, as the Foreign-currency value of the debt is depleted. This is a direct transfer of wealth across borders. To replicate this, a fiscal devaluation must be accompanied by a partial default on the debt, where the default fraction is given by the devaluation rate that would occur under the nominal depreciation.

Under LCP, the results are the exact same. Now firms set prices for each of the two markets. Nevertheless, the relevant wedges created by the various taxes are the same as under PCP. Proposition 4 shows that Proposition 1 and 2 apply to LCP as well. Importantly, this does *not* imply that the allocations attained by the devaluation (nominal or fiscal) are the same under PCP as LCP. For example, if prices are fully fixed in the short run, the ToT responds differently to a devaluation depending on the assumptions on currency of pricing (Obstfeld and Rogoff, 2000), leading to different behaviors of the real variables.

Proposition 5 shows that the suggested fiscal policies can be made revenue-neutral if  $\varepsilon = \delta$  and the dividend tax is set appropriately. Revenue neutrality here refers to the primary budget ( $TR$ ), as seignorage may be affected to. Also, revenue neutrality is measured relative to the fiscal position the government would be in under a nominal devaluation.

**Dynamic model** The authors lay out a full-fledged two-country NK model with Calvo pricing and Calvo monopolistic wage setting. Real money balances  $m = M(1 + \zeta^c)/P$  appear (separately) in the utility function, so money demand is interest-elastic. Markets are assumed complete within countries. Production is linear in labor and subject to country-wide and idiosyncratic technology shocks. Taxes may be time-varying. Now dividend (profit) taxes  $\tau_t^d$  matter.

A nominal devaluation is defined as a path for  $\mathcal{E}_t$  relative to the initial value  $\mathcal{E}_0$ . A fiscal devaluation is one that achieves the same dynamic allocation in terms of C, L and Y but for which  $\mathcal{E}_t \equiv \mathcal{E}_0$ . Many different nominal devaluations are possible, including anticipated and unanticipated ones.

Proposition 6 shows that under PCP and complete markets, a fiscal devaluation can be achieved with the natural time generalization of the policies from the static case. Again, the proof shows that the relevant equilibrium relations are satisfied by comparing to the situation with a changing NER. The appropriate money supply rule is simpler if the devaluation is not anticipated, because if it is anticipated, the interest rate differential will increase before the devaluation, which changes money demand.

Under the export subsidy and import tariff policy, it's necessary to use dividend taxes. The reason is that the time-varying consumption tax affects the dynamics of the SDF, which the firms use to set prices. Under the VAT and payroll subsidy policy, the effects on the SDF are exactly offset, so a dividend tax is not needed.

The authors consider a model in which international asset trade is restricted to an uncorrelated risk-free one-period bond. Now international risk-sharing only holds in expectation. Consequently, if the devaluation is unexpected, the risk-free bond does not help with sharing risk relative to financial autarky, and so the fiscal devaluation can be carried out with  $\zeta_t^c = \tau_t^n = \tau_t^d = 0$ , just as in the static model.

The valuation effect on Home-currency denominated assets is analogous to the static model, and partial default is needed to replicate the nominal devaluation.

The authors also consider trade only in Home and Foreign equities (claims to profit streams from the firms). A nominal devaluation leads to a negative valuation effect for Foreign holders of Home equity and a positive valuation effect for Home holders of Foreign equity. The VAT-plus-payroll-subsidy effectively taxes Home firms. If a consumption subsidy for Home households is added to the mix, the profit tax is undone by creating an effective subsidy on holding Foreign equity.

All results continue to hold under LCP. The authors also consider an extension with capital investment. They introduce an investment tax credit, a tax on capital income and a subsidy on the rental cost of capital. Firms rent capital from households. The production function is C-D in capital and labor. There is an additional FOC for investment. Consider the case of incomplete markets and an unexpected devaluation. Under the export-import policy, the results from before go through. However, under the VAT-payroll policy, a subsidy on the rental rate of capital is needed, for the same reason that a payroll subsidy is needed: without that adjustment, firms would have an incentive to substitute from capital to labor. In the general case, since the consumption subsidy and labor income tax must be used even when there is no capital, the investment subsidy and tax on capital income must be used so as to not distort the investment decision of households.

The authors provide a numerical illustration of an optimal devaluation when the Home country is hit by a negative productivity shock. The Foreign country is taken to be large so that all its variables remain fixed. It's assumed that prices are flexible but wages are subject to the Calvo friction. There is only trade in a risk-free bond. The optimal devaluation (whether nominal or fiscal) replicates the flexible price outcome. Home prices jump on impact, and so do all other variables since the shock is permanent. The authors then consider what happens if the NER and taxes are held fixed. Two relative prices need to adjust: the real wage and the ToT. Only prices can jump at first, so the ToT appreciates too much and the real wage doesn't adjust enough. Instead, nominal wages slowly converge to the new steady state, leading to depressed hours and a negative output gap. The exaggerated initial ToT response leads first to trade deficits (a result of lost competitiveness) and then to surpluses.

**Discussion** The authors discuss several details concerning implementation. The standard trilemma of international macroeconomics says that a country can't maintain an independent MP in an OE if the NER is fixed and capital can move freely. The results in this paper say that the loss of MP freedom does not restrict the range of real outcomes that can be achieved.

In general, the proposed fiscal devaluations require a change in Home money supply with no change in Foreign money supply. This can't be done directly in a currency union where member

states have surrendered their MP independence. Instead, the solution can be implemented by having the union CB increase the pan-union money supply  $\bar{M}_t$  as the Home government would have done, and then distribute all union seignorage  $\Delta\bar{M}_t$  to Home's national government. The authors note that if the coefficient on the RMB term in the utility function goes to zero, the requisite seignorage transfers are very small.

## 10.5 Galí and Monacelli (2005): “Monetary Policy and Exchange Rate Volatility in a Small Open Economy”

**Summary** Develop a model of an infinitesimally SOE. Prices are set in the producer's currency and subject to Calvo frictions. Markets are complete. Log-linearization of the equilibrium conditions lead to an NKPC and dynamic IS equation with the exact same reduced form as in the closed economy, while the slope parameters depend on structural parameters relating to the degree of trade openness and the elasticity of substitution between goods of different origin. Monetary policy is assumed to control the nominal interest rate. It is shown that in a special case of the model with log utility of consumption and unitary EoS between goods of different origins, optimal policy implements the flexible price equilibrium (this is so because an employment subsidy is assumed to cancel the distortions from monopolistic markups and the ToT, leaving price dispersion as the only inefficiency). This requires keeping the output gap at zero and thus full stabilization of *domestic* goods inflation (different from CPI inflation). In turn, this implies that the NER optimally is quite volatile. A second-order approximation to consumer utility is employed to gauge the welfare loss from following Taylor rules or a NER peg. The latter policy is the worst, since it implies excess smoothness of the NER and so insufficient domestic price stability. The Taylor rules perform somewhat better but still generate excess smoothness of the NER, and slightly more so if the Taylor rule is formulated in terms of CPI inflation rather than domestic inflation.

**Model** There is a continuum of world economies that are identical except for their exposure to idiosyncratic productivity shocks. Utility is derived from a C Armington/Dixit-Stiglitz aggregate of Home  $C_H$  and imported goods  $C_F$ , with prices  $P_H$  and  $P_F$ . The home bias coefficient is  $1 - \alpha$ , so  $\alpha$  indexes the degree of trade openness. The foreign goods bundle is a D-S aggregate of bundles from each foreign country.

Markets are complete so the SDF  $Q$  is common for all consumers. Due to the D-S aggregators, all demand functions are isoelastic, while the demand for Home and imported goods are also multiplied by  $(1 - \alpha)$  and  $\alpha$ , resp. Period utility takes the form

$$U(C, L) = \frac{C^{1-\sigma}}{1-\sigma} - \frac{N^{1+\varphi}}{1+\varphi}.$$

The effective terms of trade are  $S_t = P_{F,t}/P_{H,t}$ . Log-linearization of the CPI formula around the PPP condition  $P_H = P_F$  gives

$$p_t = (1 - \alpha)p_{H,t} + \alpha p_{F,t} = p_{H,t} + \alpha s_t.$$

This implies a relationship

$$\pi_t = \pi_{H,t} + \alpha \Delta s_t$$

between CPI and domestic inflation.

The LOP holds for all goods at all times:  $P_{i,t}(j) = \mathcal{E}_{i,t}P_{i,t}^i(j)$  (the LHS refers to the price of country  $i$ 's good  $j$  in the domestic currency). This implies that

$$p_{F,t} = e_t + p_t^*,$$

where  $e_t$  is the log nominal effective ER (summing over all countries) and  $p_t^*$  is the log world price index. Then,

$$s_t = p_{F,t} - p_{H,t} = e_t + p_t^* - p_{H,t}.$$

The bilateral RER is  $Q_{i,t} = \mathcal{E}_{i,t}P_t^i/P_t$ . The log effective RER is the sum over the logs of the bilateral ones. It is

$$q_t = e_t + p_t^* - p_t = s_t + p_{H,t} - p_t = (1 - \alpha)s_t.$$

The Backus-Smith condition holds (the constant is assumed to be 1), so

$$c_t = c_t^* + \frac{1}{\sigma}q_t.$$

The home currency price of a riskless bond denominated in foreign currency is  $\mathcal{E}_{i,t}(R_t^i)^{-1} = E_t[Q_{t,t+1}\mathcal{E}_{i,t+1}]$ , while the price of a domestic bond is  $(R_t)^{-1} = E_t[Q_{t,t+1}]$ . We then get the uncovered interest parity condition

$$E_t[Q_{t,t+1}(R_t - R_t^i(\mathcal{E}_{i,t+1}/\mathcal{E}_{i,t}))] = 0,$$

or

$$r_t - r_t^* = E_t[\Delta e_{t+1}]$$

in log-linear form. Notice that this is not an additional equilibrium condition; it follows from the C Euler equation and risk sharing conditions.

Technology is linear in labor with productivity  $A_t$  that is common to all producers in a given country and evolves as an AR(1) in logs. Employment is subsidized at rate  $\tau$ , so real MC (in terms of domestic prices) are

$$mc_t = -\nu + w_t - p_{H,t} - a_t, \quad \nu = -\log(1 - \tau).$$

Each period firms reset prices *in their domestic currency* with Calvo probability  $1 - \theta$ . As in the closed economy, the log-linear reset pricing rule is

$$\bar{p}_{H,t} = \mu + (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k E_t[mc_{t+k} + p_{H,t}],$$

since the expression inside the expectation is nominal MC.  $\mu$  is the log gross markup. This leads to exactly the same NKPC as in the closed economy, when expressed in terms of MC and domestic goods inflation,  $\pi_{H,t} = \beta E_t \pi_{H,t+1} + \lambda \widehat{mc}_t$ .

Market clearing and the Backus-Smith condition imply various relationships between production and consumption. Real MC is affected positively by world output and the ToT, since these impact domestic consumption (through risk-sharing) and thus labor supply. Changes in the ToT also have a direct effect on the product wage  $W/P_H$  for any given real wage  $W/P$ . The natural (flexible

price) level of output also depends on world output. It can be derived that domestic real MC is proportional to the output gap,

$$\widehat{mc}_t = (\sigma_\alpha + \varphi)x_t, \quad \sigma_\alpha \propto \sigma.$$

This may be substituted into the MC NKPC to get an output gap specification whose slope depends on the OE parameters. Intuitively, if elasticities are relatively high, an increase in trade openness  $\alpha$  lowers the adjustment in the ToT that is necessary to absorb a given change in relative domestic output (i.e., to generate enough expenditure switching), which dampens the impact of the ToT adjustment on MC and inflation.

One can also substitute into the C Euler equation to obtain a dynamic IS equation. The natural rate of interest depends on expected foreign output. Again, the OE parameters influence only the natural rate of interest and the slope (the interest rate sensitivity). Intuitively, if the various EoSs are high, an increase in trade openness raises the sensitivity of demand to the ToT, and thus also to the nominal interest rate.

**Optimal policy** There are three distortions relevant for policy here: (1) price dispersion, (2) monopolistic pricing and (3) an incentive to manipulate the ToT to benefit the home country (due to sticky prices and imperfect substitutability of Home and Foreign goods). The employment subsidy can be chosen so as to eliminate the latter two.

The authors consider the special case in which  $\sigma = 1$  and there is unitary EoS between both Home and Foreign goods and between different bundles of foreign country bundles. This allows the equilibrium relations to be expressed exactly rather than as log-linearizations. The optimal allocation for a SOE, who takes world Y and C as given, must satisfy

$$-\frac{U_{N,t}}{U_{C,t}} = (1 - \alpha) \frac{C_t}{N_t}.$$

The LHS is the MRS between leisure and consumption, and the RHS is the MPL for production of the composite consumption good (uses market clearing and risk-sharing). Given the choice of utility specification, C drops out and one can solve for an optimal constant level of labor. It is also possible to substitute into the formula for flexible price MC,  $\overline{MC}_t = e^\mu$ , to show that for the right choice of  $\tau$ , the above-mentioned optimal constant level of labor  $N$  does indeed generate the flexible price MC. This implies that in the optimum, the output gap is zero and so, by the NKPC, domestic inflation  $\pi_{H,t} = 0$  for all  $t$ . Thus, strict domestic inflation targeting (DIT) is optimal. This may be implemented by ensuring that the nominal interest rate equals the natural rate at all times. This, in turn, is ensured if, say, policy follows a Taylor rule that leads to determinacy.

Under DIT, real variables behave as in the flexible price equilibrium. In this case, an increase in world output generates a direct positive demand effect on Home goods as well as a windfall from financial risk-sharing; however, the ToT appreciate as a result, generating expenditure switching and risk-sharing effects. The natural exchange rate is  $\bar{e}_t = \bar{s}_t - p_t^*$ , while the level of the CPI, given the constancy of domestic prices, is proportional to the ToT, with the proportionality being the degree of openness,  $\bar{p}_t = \alpha \bar{s}_t$ . Hence, if the economy is very open and domestic productivity is not very synchronized with world output, optimal policy is consistent with a highly volatile CPI index.

The authors derive a second-order approximation to welfare. Its expectation depends (negatively) on the variance of domestic inflation and the output gap. This welfare criterion is then used to numerically evaluate three alternative monetary policy rules. The *domestic inflation-based*

*Taylor rule* (DITR) sets  $r_t = \rho + \varphi_\pi \pi_{H,t}$ , the *CPI inflation-based Taylor rule* (CITR) is similar but with CPI-inflation  $\pi_t$ , and the *exchange rate peg* (PEG) sets  $e_t = 0$ . The calibration is standard and sets  $\varphi_\pi = 1.5$ .

Consider a shock to domestic productivity. Under the optimal policy (the flexible price allocation), this shock should lead to a transitory expansion in C and Y. To generate this response, the domestic interest rate must be lowered. Due to the constant world interest rate, uncovered interest rate parity implies that the NER depreciates at first and is subsequently expected to appreciate. The ToT respond strongly initially since  $p_{H,t}$  is constant.

The DITR generates a response that is qualitatively similar to the optimal policy, although the movements in the ToT are slightly muted. This is even more so the case for CITR, since CPI stability requires a smaller movement in the ToT and in fact a fall in domestic prices. Finally, under the PEG, the currency can't depreciate to support the expansion in C and Y. The response in the ToT is very muted while inflation and the output gap are volatile. Hence, the NER and ToT exhibit *excess smoothness* (consistent with Mussa, 1986), given that prices are sticky and so can't provide the adjustment themselves.

The authors finally note that due to PCP, they have implicitly assumed full NERPT.

## 11 Currency Unions

### 11.1 Chari and Kehoe (2007): “On the need for fiscal constraints in a monetary union”

**Summary** In a monetary union with benevolent national governments and a union CB that can't commit to its future policies, a time inconsistency problem arises for MP that leads to a free-rider problem for fiscal policy. If debt levels are high, the union CB has an incentive to inflate, although this is balanced against the resulting output losses. If national governments take on more debt, they can thus expect higher inflation going forward. However, each national government does not internalize the cost of inflation imparted on the other union members. This free-rider problem leads to excessive debt and inflation relative to the coordinated optimum. Hence, debt limits can raise welfare in each country. However, if the union CB is able to commit to an inflation rate before national governments choose their debt levels, the problem goes away and debt limits can only weakly decrease welfare.

**Model** Two periods, 0 and 1. There are  $I$  identical, small countries that form a monetary union with a central MP authority. In period 0, countries start with price level  $p_0 = 1$ . MP chooses the common time-1 price level  $p_1$ , leading to gross inflation  $\pi = p_1/p_0 = p_1$ . In period 0, country output is given by  $\omega$ , while in period 1 it's  $y(\pi)$ , where  $y(\cdot)$  is decreasing and convex. It's assumed that  $\pi \geq 1$  and  $\omega$  is sufficiently small that the countries wish to borrow in period 0.<sup>6</sup> Each country  $i$  faces budget constraints

$$c_{i0} = \omega + b_i, \quad p_1 c_{i1} = p_1 y(\pi) - x_i,$$

where  $x_i$  is the nominal repayment on debt. A debt contract is denoted  $(b_i, x_i)$  with implicit gross return  $R_i = x_i/b_i$ . Note that we can write

$$c_{i1} = y(\pi) - x_i/\pi.$$

---

<sup>6</sup>If they were net creditors, the optimal gross inflation rate would be 0, as deflation would be beneficial both in terms of output and in terms of the real value of debt.

The government of country  $i$  maximizes

$$U(c_{i0}) + \beta U(c_{i1}).$$

Governments borrow from risk neutral, competitive international lenders, who will agree to a contract  $(b_i, x_i)$  iff.  $-b_i + \beta x_i/\pi \geq 0$ . The monetary authority maximizes an equally weighted average of the objective functions of all governments in the union.

Consider the case without commitment. The monetary authority moves last. Taking repayments  $\bar{x} = (x_1, \dots, x_I)$  as given, it chooses  $\pi$  to solve

$$\max_{\pi} \frac{1}{I} \sum_{i=1}^I U(y(\pi) - x_i/\pi).$$

The resulting MP rule is denoted  $\pi(\bar{x})$ . In a *non-cooperative* equilibrium, national governments choose their debt simultaneously in order to maximize their individual objective functions,

$$\max_{b_i, x_i} U(\omega + b_i) + \beta U(y(\pi(\bar{x})) - x_i/\pi(\bar{x})),$$

subject to the lending constraint, and recognizing that their choice of  $x_i$  influences  $\pi(\bar{x})$ . In a *cooperative* equilibrium, all countries go together to solve

$$\max_{b, \bar{x}} \frac{1}{I} \sum_{i=1}^I [U(\omega + b_i) + \beta U(y(\pi(\bar{x})) - x_i/\pi(\bar{x}))].$$

Under both types of equilibria, the lending constraint will bind, so  $b_i = \beta x_i/\pi(\bar{x})$ .

The authors restrict attention to symmetric equilibria. Proposition 1 shows that welfare, as measured by average total utility, is larger in the cooperative equilibrium than in the non-cooperative one. Because the non-cooperative solution is feasible in the cooperative optimization problem, all that needs to be shown is that the FOCs for the two problems don't coincide. Indeed, they don't, as they differ by a term (once the monetary authority's FOC has been substituted in)

$$(I - 1)y'(\pi) \frac{\partial \pi}{\partial x_i},$$

which is the inflation cost that a marginal unit of government  $i$  debt induces on the remaining countries in the union.

Proposition 2 shows that at the cooperative outcome, each national government would individually prefer to raise its debt level. Hence, if a debt limit  $b \leq b_C$  were imposed, where  $b_C$  is the optimal cooperative debt, then the cooperative outcome would be a non-cooperative equilibrium (at least each union member would be optimizing locally). The authors note that as the number of countries  $I$  grows, the induced inflation cost rises, so the free-rider problem should be more of an issue in a large union.

The model with commitment instead lets the monetary authority choose  $\pi$  first, and then national governments choose debt levels. Working backwards, given an inflation rate  $\pi$ , non-cooperative governments will each maximize

$$\max_{b_i, x_i} U(\omega + b_i) + \beta U(y(\pi) - x_i/\pi).$$

A cooperative union would maximize

$$\max_{\bar{b}, \bar{x}} \frac{1}{I} \sum_{i=1}^I [U(\omega + b_i) + \beta U(y(\pi) - x_i/\pi)].$$

But these two problems have the exact same solutions, since  $\pi$  is given and the union objective function is the average of the individual objective functions with no strategic interactions between terms. Hence, the cooperative and non-cooperative equilibria coincide. It also then follows that imposing debt limits will weakly reduce welfare.

## 11.2 Galí and Monacelli (2008): “Optimal monetary and fiscal policy in a currency union”

**Summary** Lays out a NK model of a currency union with a continuum of small member countries subject to imperfectly correlated shocks. There is no trade with outside countries. Prices are Calvo-staggered. There is room for fiscal policy decided by national governments as it yields utility to private households. There is home bias in consumption and markets are complete. The efficient allocation under flexible prices is compared with that under nominal rigidity, using a second-order approximation to union-wide household utility. Under flexible prices, the optimal allocation implies that the bilateral ToT equal the ratio of TFP, which implies that the inflation differential (i.e., relative to unionwide inflation) is inversely proportional to the productivity growth differential (again relative to aggregate union productivity). An NKPC, featuring both the output gap and fiscal gap (i.e., relative to the flexible price allocation), is derived along with an expression for the output gap differential. The utility approximation involves both inflation, the output gap and the fiscal gap. Optimal coordinated policy under commitment involves fully stabilizing union-wide inflation and setting the *aggregate* output and fiscal gaps to zero. This in turn implies that optimal fiscal policy doesn’t set both the *individual* output and fiscal gaps to zero. Instead the authors show, partly through simulations, that the optimal policy mix requires using fiscal policy to smooth the adjustment of prices over time, leading to a positive fiscal gap and negative output gap following a positive productivity shock.

**Model** The currency union is modeled as a closed system with a continuum of SOEs  $i \in [0, 1]$ , each infinitesimally small so that they take foreign aggregates as given. Each country has a continuum of monopolistically competitive producers  $j \in [0, 1]$ . The LOP holds across all individual goods.

Households derive utility  $U(C, N, G) = (1-\chi) \log C + \chi \log G - N^{1+\varphi}/(1+\varphi)$  from  $C$ , government purchases and  $L$ .  $C_t^i$  is C-D consumption index with relative weight  $\alpha$  on Foreign-produced goods. The bundle of Home-produced goods is a CES aggregate with EoS  $\varepsilon$ . The bundle of Foreign-produced goods is given by

$$C_{F,t}^i = \exp \int_0^1 c_{f,t}^i df,$$

where  $c_{f,t}^i$  is the log of a CES aggregate of goods produced in country  $f$ , also with EoS  $\varepsilon$ . Because the economy is infinitesimally small, an  $\alpha < 1$  implies home bias in consumption. Hence, countries have different consumption baskets so CPI inflation differentials can emerge even when the LOP holds for each good.

Taxes are lump-sum and because households are infinitely lived, Ricardian equivalence holds. Households can trade a complete set of contingent claims. The common SDF is denoted  $Q_{t,t+1}$ . The

gross nominal (unionwide) risk-free return is  $R_t^* = 1/E_t[Q_{t,t+1}]$ , which the union CB is assumed to set. Optimal allocations of expenditures are given by the usual isoelastic and C-D formulas. The optimal allocation of expenditures on imported goods by country of origin, implied by the form of  $c_{f,t}^i$ , is

$$P_t^f C_{f,t}^i = P_t^* C_{F,t}^i,$$

where  $P_t^f$  is the price index for goods produced in country  $f$  and  $P_t^*$  is the union-wide price index. The CPI for country  $i$  is  $P_{c,t}^i = (P_t^i)^{1-\alpha}(P_t^*)^\alpha$ . The bilateral ToT are  $S_{f,t}^i = P_t^f/P_t^i$ , while the effective ToT are  $S_t^i = P_t^*/P_t^i = \exp \int_0^1 s_{f,t}^i df$ . We have  $P_{c,t}^i = P_t^i (S_t^i)^\alpha$ , so

$$\pi_{c,t}^i = \pi_t^i + \alpha \Delta S_t^i.$$

International risk-sharing (and symmetry) implies the Backus-Smith condition

$$C_t^i = C_t^f (S_{f,t}^i)^{1-\alpha}.$$

Public consumption  $G_t^i$  is a CES aggregate over *domestic* goods with EoS  $\varepsilon$ . The government is assumed to allocate expenditures across goods so as to minimize total cost, which gives an isoelastic government demand function for individual goods.

The technology is linear in labor,  $Y_t^i(j) = A_t^i N_t^i(j)$ . Country-wide TFP follows an AR(1) in logs with coefficient  $\rho_a$ . There is a constant employment subsidy  $\tau^i$  such that real MC (in terms of domestic goods) is

$$mc_t^i = \log(1 - \tau^i) + w_t^i - p_t^i - a_t^i,$$

and this must be common across firms due to the linear technology. The output aggregate  $Y_t^i$  is a CES aggregate of firm outputs with EoS  $\varepsilon$ .

Firms reset prices with Calvo probability  $1-\theta$ . As usual, the optimal reset price is approximately

$$\bar{p}_t^i = \mu + (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k E_t[mc_{t+k}^i + p_{t+k}^i],$$

where  $\mu = \log[(\varepsilon - 1)/\varepsilon]$  is the steady-state and flexible price log markup.

**Results** A log-linear approximation (around the symmetric steady state) to the market-clearing condition for goods gives

$$\hat{y}_t^i = \gamma \hat{g}_t^i + (1 - \gamma) \hat{c}_t^* - (1 - \gamma)(p_t^i - p_t^*),$$

where  $\hat{c}_t^*$  is aggregate consumption (a proxy for the strength of union demand) and  $\gamma = G/Y$  is the steady-state share of government purchases in GDP. By integrating the above over  $i$  and using the consumption Euler equation, it's possible to derive the union-wide DIS equation:

$$\hat{y}_t^* = E_t[\hat{y}_{t+1}^*] - (1 - \gamma)(r_t^* - E_t[\pi_{t+1}^*] - \rho) - \gamma E_t[\Delta \hat{g}_{t+1}^*],$$

where  $\pi_t^* = \int_0^1 \pi_t^i di$ . Hence, fluctuations in union-wide output are determined by variations in current government purchases and expected real long-term interest rates.

The PPI-based NKPC written in terms of real MC is exactly the same as in the closed economy. Real MC can be written

$$\widehat{mc}_t^i = \left( \frac{1}{1 - \gamma} + \varphi \right) \hat{y}_t^i - \frac{1}{1 - \gamma} \hat{g}_t^i - (1 + \varphi) a_t^i.$$

Given output, an increase in  $g_t^i$  either crowds out consumption or generates a real appreciation, both of which reduce the product wage (real wage, deflating by the price of the good) and thus real MC. The elasticity of MC wrt. output is increasing in  $\gamma$ : For any given percent increase in output, and given government purchases, a larger  $\gamma$  is associated with a larger percentage increase in either consumption or the ToT. Both of these reduce the product wage and thus real MC. Using the above expression for real MC, it's possible to derive a union-wide NKPC.

Suppose prices are *flexible* and consider the social planner's problem. He sets the marginal utility loss of producing an extra unit of the composite good equal to the utility gain from each of its three uses: (1) consumption by domestic households, (2) consumption by all households in the union and (3) domestic government spending. The optimal allocation (denoted by bars) turns out to have  $\bar{N}_t^i = 1$  and

$$\bar{C}_t^i = (1 - \chi)(A_t^i)^{1-\alpha}(A_t^*)^\alpha,$$

where  $A_t^* = \exp \int_0^1 a_t^f df$  is aggregate productivity. This allocation can be decentralized as in Galí and Monacelli (2005) by setting a constant employment subsidy to offset the markup distortion (note that the incentive to distort the ToT isn't an issue here as only cooperative optima are considered). From the Backus-Smith condition, in the flexible price optimum, the ToT are given by

$$\bar{S}_t^i = (\bar{C}_t^i / \bar{C}_t^*)^{1/(1-\alpha)} = A_t^i / A_t^*,$$

so the inflation differential is inversely related to the productivity growth differential,

$$\pi_t^i - \pi_t^* = -(\Delta a_t^i - \Delta a_t^*).$$

Now consider the case with *nominal rigidities*. Let tildes denote differences from the flexible price allocation. Hence,  $\tilde{y}_t^i$  is the output gap and  $\tilde{f}_t^i$  the *fiscal gap*. It's assumed that relative government consumption in steady state is efficient, which means  $\gamma = \chi$ . One can derive an NKPC

$$\pi_t^i = \beta E_t[\pi_{t+1}^i] + \lambda(1 + \varphi)\tilde{y}_t^i - \frac{\lambda\chi}{1 - \chi}\tilde{f}_t^i, \quad (7)$$

and an expression for the output gap differential,

$$\Delta\tilde{y}_t^i - \Delta\tilde{y}_t^* = \frac{\chi}{1 - \chi}(\Delta\tilde{f}_t^i - \Delta\tilde{f}_t^*) - [(\pi_t^i - \pi_t^*) - (\Delta a_t^i - \Delta a_t^*)]. \quad (8)$$

Suppose the unionwide gaps are zero. We see that closing the individual output and fiscal gaps is consistent with zero domestic inflation in the NKPC. But if domestic inflation is zero and productivity shocks are asymmetric, the second equation implies that closing the output gap while keeping the fiscal gap closed requires a ToT adjustment and thus a movement in prices (in Galí and Monacelli, 2005, the movement could come from the NER). There is then a tradeoff between stabilizing the gaps and domestic inflation.

If we integrate the above conditions across  $i$ , the ToT term disappears. Hence, zero unionwide inflation is consistent with zero unionwide output and fiscal gaps. The authors derive a second-order approximation to aggregate welfare; it features quadratic terms in domestic inflation and the two gaps. They then seek to characterize optimal *coordinated* policy. They set up the Lagrangian and optimize with respect to inflation and the two gaps, both individual and unionwide. The constraints are (7)–(8) and the expressions for the aggregate in terms of the individual decision variables. The FOCs imply that the optimum has  $\pi_t^* = \tilde{y}_t^* = \tilde{f}_t^* = 0$ , i.e., at the union level inflation and the two

gaps should be zero, which is implemented by setting the interest rate to the unionwide natural interest rate at all times (and using some Taylor rule off the equilibrium path). Under this optimal aggregate policy, equation (8) (in non-differenced form) simplifies to

$$\tilde{y}_t^i - \frac{\chi}{1-\chi} \tilde{f}_t^i + p_t^i = -a_t^i,$$

under the normalization  $p_t^* = 0$ . As the price level adjusts only gradually, positive productivity shocks are thus smoothed by a combination of positive fiscal gaps and negative output gaps. Intuitively, the sluggish adjustment in prices and the ToT leads to too little demand switching over to Home-produced goods, leading to a negative output gap, which must be alleviated using fiscal policy. This is illustrated in a simulation with three different values of  $\theta$ : 0, 0.4 and 0.75. The larger  $\theta$  (the more price stickiness), the longer does the adjustment to a positive productivity shock take. For  $\theta = 0$ , the price level drops instantaneously and the output and fiscal gaps remain closed as the ToT depreciate efficiently. For  $\theta > 0$ , the above-mentioned mix of fiscal policy and output below potential are used to smooth the slow transition to the new efficient allocation.

Galí and Monacelli mention that their model leaves out wage stickiness, distortionary taxation, non-Ricardian behavior of households and incompleteness of markets. If markets weren't complete, optimal policy would presumably involve a complementary role for fiscal policy as a tool for cross-country insurance.

### 11.3 Mundell (1961): “A Theory of Optimum Currency Areas”

**Summary** Proponents of flexible exchange rates say that its equilibrating function on the terms of trade can alleviate unemployment problems in deficit countries and inflation in surplus countries. However, flexibility of price adjustment is needed primarily across areas where factor inputs are immobile, not between those where capital and labor can move freely. Mundell defines an area within which factor inputs are mobile, but whose borders limit factor mobility, as a *region* and argues that, in a world with flexible exchange rates, the optimal currency area is a region, not a nation. In practice, an upper limit on the optimal number of currency areas is placed by requirements that currency should serve a useful function as a unit of account and medium of exchange, not be too prone to speculation, and imports must be a sufficiently small part of the region's expenditures to reasonably allow the kind of money illusion in wage bargaining necessary for the flexible exchange rate argument to work.

**Theory** The discussion on flexible ERs can be split into two questions. First, one must ask whether a system of flexible ERs can work efficiently. This requires (1) robustness to speculation, (2) that required ER changes are not too violent, (3) risks can be covered in forward markets, (4) CBs will refrain from monopolistic speculation, (5) monetary discipline will be maintained by the threat of adverse effects from depreciation, (6) protection of debtors and creditors, and (7) that wages and profits aren't indexed to the price of imports. The main focus of this paper is the second question, which is: What determines the size of an optimal currency area?

First Mundell lays out a distinction between (1) a currency area containing many nations with their own currencies but fixed ERs and (2) a currency area with a common currency and thus a single central bank. In case (1), if there's a shift of demand from country B to country A, and the surplus country (A) isn't willing to allow its price level to rise and so tightens credit conditions, the necessary adjustment is a decrease in B's real income, which implies a rise in unemployment.

In case (2), the union CB must increase the money supply to fight the recessionary tendency in B, which exacerbates the inflationary tendency in A. In either case, it's impossible to prevent both unemployment and inflation among all members of the currency area.

In fact, adjustment by means of a flexible ER is only necessary between *regions*, as defined above: If factor inputs are freely mobile between localities, exchange rate flexibility is not necessary to bring about the adjustment. Hence, the optimum currency area is the region.

This does not imply that the optimal number of currency areas is equal to every single minor pocket of unemployment arising from labor immobility. The above arguments deal only with stabilization concerns. However, the fewer currencies there are, the better do each of them serve their functions as a unit of account and a medium of exchange. Currencies of small areas will tend to be more subject to speculation. Lastly, one of the mechanisms by which the NER brings about adjustment works through money illusion in wage bargaining; unions are supposed to care more about nominal wages and the domestic price level than variations in the NER. But this is not realistic if the currency area is so small that imports constitute a large share of total expenditures.

#### 11.4 Santos Silva and Tenreyro (2010): “Currency Unions in Prospect and Retrospect”

**Summary** Summarizes the empirical evidence on the benefits and costs of joining a currency union. Particular emphasis is placed on the Eurozone. The literature has mostly found that currency unions benefit the volume of trade within the union. The authors provide a new set of diff-in-diff result, comparing the trade statistics of EZ countries to similar countries outside of the zone. Graphically, while trade among EZ countries has increased since the creation of the euro, this is also the case for these countries' trade with non-EZ countries. Regression results confirm that the partial effect of the EZ dummy is insignificant and that it's important to control for the fact that EZ countries traded a lot with each other before the creation of the euro. The literature has found positive effects of the euro on capital integration. While it's hard to test, there's no evidence of improvements in diversification of risks. Inflation rate differentials have actually increased with the adoption of the euro. There is some evidence that the euro has led to structural changes in labor and product markets.

**Benefits of currency unions** Most studies that estimate the effect of currency unions on trade use a gravity model where the LHS variable is some measure of trade flows (uni- or bidirectional) and the explanatory variables are distance (both physical and other trade barriers), size of the countries and a CU dummy. Typically, a multiplicative specification is used. Rose (2000) kicked off the literature by finding a huge effect (200%) on bilateral trade. Many studies pointed out the potential for endogeneity as CUs are not randomly formed due to cultural and historical links. Various matching, Heckman and IV approaches have been employed. Studies typically find positive effects, although the range is sizeable. For the EZ, estimates range from 2% to 70%.

The authors use trade data on the Euro-12 countries to provide another set of results. A diff-in-diff approach is used, using data on three groups of similar countries outside the EU-12. The sample is 1993–2007. Figure 1 shows that trade among EU-12 countries has increased over time, but so has trade between EU-12 countries and the other countries in the sample. The authors run a gravity regression with an added dummy indicating whether a country pair is in the EU-12. The EZ dummy then measures the partial effect of those countries joining the euro. The EU-12 dummy

is strongly significant, indicating that past trade ties are important. However, the EZ dummy is not significant, so there's no added effect of adopting the euro.

Another supposed benefit a CU is the scope for increased capital integration. Indeed, with the adoption of the euro, yield differentials dropped and cross-border bond and equity holdings as well as FDI increased. As measured by consumption volatility and output growth, risk sharing has not improved, however. This may be due to (1) there being little scope for diversification as the countries face symmetric shocks, or (2) the elimination of the NER margin has removed a useful insurance instrument.

Alesina and Barro (2002) argued that joining a common currency may eliminate the inflationary bias resulting from time inconsistency in MP. Fleming (1971) argued that low and stable inflation rates among member countries were a precondition for a successful CU, since this reduces volatility in the ToT and thus external imbalances. However, inflation differentials among EZ countries have actually increased since its creation, with Germany on the low end and Ireland on the high. This is due to differences in productivity growth, changing competition and different exposure to shocks outside the union, all affecting prices through the Balassa-Samuelson effect.

**Costs of and criteria for currency unions** The main cost of joining a CU is the loss of MP independence. Mundell (1961) argued that OCAs should feature synchronized shocks so that the direction of optimal stabilization policy is unambiguous. European countries have historically been subject to less correlated shocks than U.S. states; however, in principle it is *ex post* shocks that matter. The literature has found that adoption of a CU increases the comovement of relative price shocks but decreases that of output shocks. This may be due to increased specialization, although the evidence is ambiguous. Mundell (1973), in somewhat of an about-face, instead argued that due to deeper capital markets integration, countries with asynchronous shocks make ideal CU partners. Theoretical work (Ching and Devereux, 2003) has found that the diversification and common-stabilization-policy motives may cancel each other.

Mundell (1961) emphasized labor mobility as an important adjustment mechanism in a CU. Mobility among EZ countries has historically been lower than among U.S. states, and the main equilibrating mechanism in Europe seems to have been labor-force participation. Alesina, Ardagna and Galasso (2010) found that the euro's creation has brought about subtle reforms, as a secondary flexible labor market has emerged in many EZ countries and there seems to be less indexation of wages to past inflation.

Kenen (1969) argued that industrially diversified countries would be better candidates for a CU, as this would lead to asymmetric effects of shocks. There does not seem to be significant differences between industrial diversification in the U.S. and the EZ.

Friedman (1953) argued in favor of a flexible NER due to the existence of nominal rigidities. However, this rested on an assumption of PCP. When there is LCP and pricing-to-market, ERPT may be low, so that NER flexibility is inconsequential for stabilization. Empirical studies by Engel and coauthors find low ERPT.

Finally, Kenen (1969) argued that fiscal transfers are essential to maintaining a CU. The authors do not summarize the empirical literature on this topic.

## 12 Exchange Rate Disconnect

### 12.1 Bachetta and van Wincoop (2006): “Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?”

**Summary** Develop a noisy rational expectations model of exchange rate determination in which information about the fundamental (relative money supplies) is symmetrically dispersed among traders. Traders face different idiosyncratic ER exposures on their non-asset income. These exposures are private information, so the ER has an aggregate hedging component that is unobservable. When the ER changes, private investors can’t tell whether this is driven by hedging or new information about the fundamental, so they rationally ascribe part of the change to the latter. This means that effects of hedging trades on the ER are magnified (relative to the benchmark with identically informed traders), leading to a disconnect from fundamentals. Because the ER will in general depend on aggregate higher-order expectations, shocks to the unobservable hedging trades lead to persistent effects on the ER. Eventually, information about the fundamentals will disseminate to all investors through price discovery, so in the long run ERs and fundamentals are closely linked. Finally, the authors are explicit about the mechanism by which prices clear markets: Limit orders are passively placed based on public information and are conditional on the ER; they are then matched by market orders, which are the pure private component of asset demand and are not conditional on the ER. These latter orders are seen as the active ones, so the order flow is defined to be equal to market orders. The authors then show that the ER can be written as a linear combination of public information and order flow. Since the ER is very sensitive to private information in the model, it is closely linked to order flow in the short run, as in the data.

**Model** Two countries, one good, PPP holds. There is a continuum of investors in each country. There are overlapping generations of investors that each live for two periods and make only one investment decision; when they die, they bequeath their private information to their heir. Investors can invest in four assets: money in their home country, nominal bonds of either country and a technology with fixed real return (the latter just pins down the Home interest rate, for simplicity, as Home is assumed to be large and Foreign infinitesimally small). Each investor’s non-asset income has a stochastic exposure to the ER  $s_t$  (Home per Foreign currency) so that a hedging demand arises. The individual exposures are symmetrically dispersed around the aggregate mean, and it’s assumed that the variance of the dispersion term is so large that individual exposure provides no signal about average exposure.

Under the assumptions of the model, money market equilibrium is given by

$$m_t - p_t = -\alpha i_t, \quad m_t^* - p_t^* = -\alpha i_t^*,$$

while bond market equilibrium yields

$$\bar{E}_t[s_{t+1}] - s_t = i_t - i_t^* + \gamma \sigma_t^2 b_t.$$

Here  $\bar{E}_t$  denotes the cross-sectional average expectation, and the last term above (the deviation from UIP) is the aggregate hedging demand, which is unobservable. Define the observable fundamental  $f_t = m_t - m_t^*$ . Then one can shift the above equation forward and use money market equilibrium to get

$$s_t = \frac{1}{1 + \alpha} \sum_{k=0}^{\infty} \left( \frac{\alpha}{1 + \alpha} \right)^k \bar{E}_t^k [f_{t+k} - \alpha \gamma \sigma_{t+k}^2 b_{t+k}],$$

where higher-order expectations are given by  $\bar{E}_t^k[x_{t+k}] = \bar{E}_t \bar{E}_{t+1} \dots \bar{E}_{t+k-1}[x_{t+k}]$ .

The information structure is as follows. At time  $t$ , each investor  $i$  receives a signal

$$v_t^i = f_{t+T} + \varepsilon_t^{vi}, \quad \varepsilon_t^{vi} \sim \mathcal{N}(0, \sigma_v^2),$$

about the fundamental  $T$  periods ahead. Hence, the investor has private signals about future fundamentals  $f_{t+1}, \dots, f_{t+T}$ , while the current and past fundamentals  $f_t, f_{t-1}, f_{t-2}, \dots$  are common knowledge. The MA( $\infty$ ) process (not its realizations) for  $f_t$  is common knowledge; let the innovations be denoted  $\varepsilon_t^f$ . The ER at time  $t$  is conjectured to only depend on  $\varepsilon_{t+T}^f, \varepsilon_{t+T-1}^f, \dots$  as well as innovations to aggregate hedging demand:

$$s_t = A(L)\varepsilon_{t+T}^f + B(L)\varepsilon_t^b.$$

Because only the innovations  $\varepsilon_{t+1}^f, \dots, \varepsilon_{t+T}^f$  to  $f_t$  are unknown, the investor solves a signal extraction problem for a finite number of innovations. Having solved this signal extraction problem, the average expectation of  $s_{t+1}$  can be computed and substituted back into the above equations. Then  $A(L)$  and  $B(L)$  can be determined by solving a fixed point problem.

In recent years, ER trades have been carried out on organized exchanges in the form of auctions. Some investors place *limit orders*, which passively supply liquidity and are displayed to other investors. These other investors may then submit *market orders* to execute a deal at the posted terms (and thus drain liquidity). This kind of market is called *order driven* and stands in contrast to a *quote driven* market in which market makers are in charge of posting single bid and ask prices. Bachetta and van Wincoop's model is tied to the order driven market structure in the following way. Limit orders are assumed to only depend on public information, but they can be conditional on the ER at the time of deal execution. Market orders are defined as the pure private component of asset demand, and they can't be conditioned on the realized ER. The demand for foreign bonds

$$b_{Ft}^i = \alpha_1 I_t^p + \alpha_2 s_t + \alpha_3 I_t^i,$$

where  $I_t^p$  represents public information and  $I_t^i$  private, can therefore be split into a market order component  $\alpha_3 I_t^i - E(\alpha_3 I_t^i | I_t^p)$  and a limit order component, which is the residual. Because a market order is the active side of the trade in this interpretation, the *order flow*  $\Delta x_t$  is taken to be the aggregate of individual market orders at time  $t$ . Note that if a shift in demand is due only to new public information, the order flow is zero but the ER will still change.

Because  $s_{t-1}$  is part of  $I_t^p$ , we can write

$$\Delta s_t = \eta_1 I_t^p + \eta_2 \Delta x_t$$

for appropriate constants. The two terms on the RHS are orthogonal by construction, so a regression of changes in the ER on order flow leads to an unbiased estimate of  $\eta_2$ . Causality runs from quantity to prices in this model, since order flow decisions are determined before the ER is known.

**Results** The authors first consider the simplest case  $T = 1$  (investors receive one signal each period about next period's fundamental). The model may then be solved analytically. Only average first-order expectations matter in this situation since  $f_{t+1}$  is common knowledge at time  $t+1$ , so it doesn't matter whether an investor thinks that he is differentially informed at time  $t$ . The investor's signal extraction problem can be solved analytically by first conjecturing a solution for

$s_t$ . Relative to a benchmark model in which all investors receive the same signal, in the expression for  $s_t$  in the heterogeneous information case, the coefficient on aggregate hedging demand is larger (more negative) by a factor of  $z$ . This is because investors are rationally confused, as explained under “Summary.” Thus, when aggregate non-asset ER exposure  $b_t$  changes, it affects  $s_t$  directly through the risk-premium channel but also through the indirect effect on  $E_t f_{t+1}$ . The magnification  $z$  depends non-monotonically on the private signal variance. It rises at first, since the precision of the private signal is reduced and so more weight is given to the ER signal. Eventually it declines, because larger variance means that there is less total information about next period’s fundamental, so the ER will depend less on  $E_t f_{t+1}$ .

In the benchmark model where all investors receive the same signal, investors can perfectly deduce  $b_t$  from  $s_t$  since they know everyone else’s information set, i.e., the ER is fully revealing. The  $R^2$  from a regression of the ER on observed fundamentals is therefore higher in the common-information case than in the heterogeneous-information case (in which  $b_t$  is unknown).

The authors solve for the order flow by identifying the private information component of the equilibrium demand for foreign bonds. The ER can then be written as a linear combination of  $f_t$  and  $\Delta x_t$ . The more magnification  $z$ , the higher the weight on order flow. The order flow captures the extent to which the ER changes due to the aggregation of private information. Thus, the ER is closely linked to order flow in the heterogeneous information model, while it is more closely linked to the observed fundamental in the common information model.

The authors go on to numerically solve the model for  $T > 1$ . The higher  $T$ , the more persistent are the effects of even transitory shocks to  $b_t$ , as investors only gradually realize that the change in the ER is not caused by an innovation in the fundamental. As learning proceeds over time, the impact of the initial shock  $b_t$  dissipates. The model is calibrated numerically. The model is found to be consistent with a large magnification  $z$ . In the short run, most of the variation in ER changes is due to hedging demand, whereas long-run regressions of  $s_{t+k} - s_t$  on lags of observed fundamentals approaches 100%. The  $R^2$  of a regression of the ER on order flow is high at short horizons and grows even higher with time. This is not inconsistent with the observation that the ER is closely linked with fundamentals in the long run, as order flow is the mechanism by which information about said fundamentals is expressed over time. Evans and Lyons (2002) also found such patterns in the data.

## 12.2 Engel, Mark and West (2008): “Exchange Rate Models Are Not as Bad as You Think”

**Summary** Much has been made in the literature about the lack of forecasting power of ER models based on fundamentals; in particular, Meese and Rogoff (1983) found that no model was able to beat a simple RW model in terms of out-of-sample performance, even though *realized* fundamentals were generously used to generate the macro model forecasts. Engel et al. review the theoretical and empirical results from Engel and West (2005), which explain the poor short-term performance of the macro models and provide evidence that they are not all bad. The authors emphasize that predictions from ER models depend crucially on whether MP is modeled as endogenous: If the CB follows a Taylor rule, higher inflation leads to an appreciation due to expected policy tightening; this is not the case with passive money supply. The authors construct proxies for the PDVs using survey expectations and find that these proxies can account for a non-negligible amount of the high ER volatility. Finally, it’s shown that if ER models are estimated using panel data (as opposed to country-by-country), the long-run out-of-sample forecasting ability beats the RW model.

**Theory** Engel et al. provide an additional example of the theoretical result in Engel and West (2005). The model is a version of the Dornbusch (1976) overshooting model. Suppose that (10) holds with  $\rho_t = 0$  (UIP), and that the fundamentals follow a random walk, i.e.,

$$\Delta[m_t - m_t^* - \gamma(y_t - y_t^*) - (v_{mt} - v_{mt}^*)] = u_t,$$

where  $u_t$  is i.i.d. As in Dornbusch, a mechanism for RER adjustment is derived using price stickiness. Assume that PPP holds in the long run, so  $p_t - p_{t-1}$  eliminates part of the time- $(t-1)$  PPP deviation  $s_{t-1} + p_{t-1}^* - p_{t-1}$ . There is also a forward-looking trend term that is equal to the expected change in the market-clearing ER. Hence,

$$p_t - p_{t-1} = \theta(s_{t-1} + p_{t-1}^* - p_{t-1}) + E_{t-1}s_t + p_t^* - (s_{t-1} + p_{t-1}^*).$$

The analogous relation holds for the other country. The above equation implies that  $E_{t-1}q_t = (1 - \theta)q_{t-1}$ . Substituting back into the PDV formula for  $s_t$ , one can get

$$s_t - s_{t-1} = -\frac{\theta}{1 + \alpha\theta}q_{t-1} + \frac{1 + \alpha\theta}{\alpha\theta}u_t.$$

In response to a shock  $u_t$ , the expected ER jumps more than one for one; this is the overshooting. Intuitively, the lack of adjustment in prices causes overadjustment in financial variables. The volatility is greater, the stickier prices are (the smaller  $\theta$  is). It is also seen that the Engel-West RW theorem holds, because  $s_t - s_{t-1} \approx u_t$  as  $b = \alpha/(1 + \alpha) \rightarrow 1$ . The relation above can be rewritten to isolate the RER and the  $R^2$  of a regression of  $q_t$  on  $q_{t-1}$  can be calculated. For a reasonable calibration, the  $R^2$  is tiny.

The authors also derive a formula for the ER where the interest rate is determined by a Taylor rule. Such models can have quite distinct implications. For example, if the coefficient on inflation in the Taylor rule is greater than 1, an increase in inflation leads to an *appreciation* of the ER because the CB is expected to sufficiently tighten MP by raising  $i_t$ . In traditional money-demand models, an increase in expected inflation lowers the demand for money, which lowers the market-clearing  $i_t$ , thus causing a depreciation of the currency.

ER models have been criticized because they imply too little persistence in the RER for reasonable amounts of price stickiness. E.g., for the Dornbusch model above, the persistence of the RER equals the persistence of prices. However, Engel et al. note that other models are able to delink price persistence and RER persistence. This was pointed out by Benigno (2004). The authors give another example, using the framework in Galí and Monacelli (2005). In this model, the RER is proportional to the MP shock and the autocorrelation does not depend on price stickiness (except that the afore-mentioned constant of proportionality is zero if prices are flexible). Intuitively, after a positive MP shock, the firms that first get to adjust their prices only adjust upward moderately due to other firms being stuck at a lower price. This generates persistence.

The authors give an example (as in Campbell, 2001) that explains why forecasting power starts low but increases with the forecasting horizon. Suppose that the ER is given by

$$s_t = (1 - b)x_t + b\rho_t + bE_t s_{t+1},$$

and observed fundamentals follow a RW  $x_t - x_{t-1} = \varepsilon_t$ , while unobserved fundamentals follow an AR(1),  $\rho_t = \alpha\rho_{t-1} + u_t$ . The forward solution is

$$s_t = x_t + \frac{b}{1 - b\alpha}\rho_t,$$

which implies the  $k$ -period-ahead equilibrium correction model

$$s_{t+k} - s_t = \sum_{j=1}^k \varepsilon_{t+j} + (1 - \alpha^k)z_t + \frac{b}{1 - b\alpha} \sum_{j=1}^k \alpha^{k-j} u_{t+j}, \quad z_t = x_t - s_t \propto \rho_t.$$

Note that if the model is correct, the econometrician observes  $z_t$ . One can now calculate the  $R^2$  of the above population regression of  $s_{t+k} - s_t$  on  $z_t$ . It may take on a hump-shaped pattern in  $k$ : As  $k$  grows, the mean reversion in  $\rho_t$  makes it easier to forecast, but the variance of  $x_{t+k} - x_t$  grows without bound and eventually dominates. This intuition suggests that long-term forecasts may do better than short-term ones.

**Empirics** Engel et al. replicate the Granger causality results found by Engel and West (2005). They also use VARs to generate measures of the PDV in various models of the ER and compute how much of the actual volatility in ER changes can be explained by changes in the PDV. Monetary models of the ER can account for a fairly large fraction of volatility, about 50%.

The authors review the literature on high-frequency event studies of ER movements. These studies have generally found that when news about U.S. activity variables are released by statistical bureaus, the ER moves in the direction implied by Taylor rule models (i.e., news that suggest an expanding economy cause an appreciation).

The authors also construct the relevant PDVs using survey forecasts of inflation and output from Consensus Forecasts. Surveys are conducted every six months and contain predictions for up to 10 years out. The change in the RER is regressed on changes in the expected PDV of inflation and of output. For most countries in the sample, the coefficients get the right sign.

Finally, the authors estimate short-horizon predictive regressions using panel methods. The literature has mostly been using country-by-country regressions. The panel model is

$$s_{i,t+k} - s_{it} = \beta_k(x_{it} - s_{it}) + \varepsilon_{i,t+k}, \quad \varepsilon_{i,t+k} = \zeta_i + \theta_t + u_{it}.$$

As for the fundamentals  $x_t$ , the authors use measures based on a monetary model  $m_t^* - m_{it} - \gamma(y_t^* - y_{it}^*)$ , a PPP model  $p_t^* - p_{it}$  and a Taylor rule model (whose coefficients are simply postulated rather than estimated). 1- and 16-quarter-ahead out-of-sample forecasts are generated and compared to those from a RW model. Relative forecast accuracy is measured by the ratio of root mean square prediction error (RMSPE) relative to the RW model (called Theil's U statistic). They calculate t-statistics from Clark and West (2006, 2007). None of the models do good at the 1-quarter horizon. At the 16-quarter horizon, the monetary and PPP models outperform the RW (with and without drift) for most currencies in the sample. The Taylor rule model's relative performance is more ambiguous. The authors conclude that the macro models perform well at the longer horizon, at least if estimated using panel methods, and that the ER evidence is consistent with interest rate movements being mainly driven by monetary factors.

### 12.3 Engel and West (2005): "Exchange Rates and Fundamentals"

**Summary** Meese and Rogoff (1983) showed that most fundamentals-based macro models of the NER don't outperform a simple RW model out of sample. In most macro models, the ER is given by a PDV of expected future fundamentals. Engel and West prove that, if the fundamentals are I(1), as the discount factor goes to 1, the NER will get close to behaving like a RW. This explains

the poor short-run performance of fundamentals-based models. However, PDV models imply that the reverse causation may still be at work: the ER should Granger cause future fundamentals. This does in fact hold up moderately well in quarterly data for bilateral G-7 exchange rates, although statistical significance is achieved inconsistently across countries and measures of fundamentals. Nominal variables seem to be better predicted by the ER. Finally, the authors construct empirical measures of the PDV implied by various measures of fundamentals, using VAR estimates of the expectations. If only the particular fundamental variable's own lags is used to generate expectations, the correlation between the generated PDV and the actual NER is negative. If the ER is also used in generating the VAR expectations, the correlation turns moderately positive. Unobserved fundamentals still seem to account for a large fraction of ER movements.

**Theory** Most macro models of the NER  $s_t$  (the log of the domestic currency price of foreign currency) can be written

$$s_t = (1 - b)a'_1x_t + ba'_2x_t + bE_t s_{t+1},$$

where  $x_t$  are fundamentals and  $b \in (0, 1)$ . The no-bubbles forward solution is

$$s_t = (1 - b)E_t \sum_{j=0}^{\infty} b^j a'_1 x_{t+j} + bE_t \sum_{j=0}^{\infty} b^j a'_2 x_{t+j}. \quad (9)$$

Suppose for simplicity that  $E\Delta x_t = 0$  and let  $\Delta x_t = \theta(L)\varepsilon_t$  be the Wold decomposition of the first differenced fundamentals. The authors' theorem shows the following.

- If  $a_2 = 0$ , then  $\text{plim}_{b \rightarrow 1}[\Delta s_t - a'_1\theta(1)\varepsilon_t] = 0$ . Thus,  $\Delta s_t$  will behave approximately like the unpredictable sequence  $a'_1\theta(1)\varepsilon_t$  as the discount factor approaches 1.
- If  $a_1 = 0$ , then  $\text{plim}_{b \rightarrow 1}[(1 - b)\Delta s_t - ba'_2\theta(1)\varepsilon_t] = 0$ . Hence, the correlation of  $(1 - b)s_t$  (and thus  $s_t$ ) with time- $(t - 1)$  information will be close to zero.
- More generally, we have  $\text{plim}_{b \rightarrow 1}[(1 - b)\Delta s_t - ((1 - b)a_1 + ba_2)'\theta(1)\varepsilon_t] = 0$ .

Hence, the ER will behave approximately like a RW for large  $b$  if one of these two conditions hold: (a)  $a'_1x_t$  is I(1) and  $a_2 = 0$ , or (b)  $a'_2x_t$  is I(1) ( $a'_1x_t$  can be either stationary, I(1) or identically zero).

Intuitively, as long as one of the fundamentals is I(1), we can write  $s_t$  as a PDV of an I(1) component plus a PDV of stationary components. As  $b \rightarrow 1$ , changes in the former will attain infinite variance and thus dominate. Here's a somewhat more precise explanation. Suppose first  $a_2 = 0$  and  $a'_1x_t$  is I(1). We can Beveridge-Nelson decompose  $a'_1x_t = \tau_t + \mu_t$ , where  $\tau_t$  is a RW and  $\mu_t$  a stationary RV. As  $b \rightarrow 1$ , more weight is placed on expectations far into the future, so  $\tau_t$  dominates variance-wise relative to  $\mu_t$ . We then get

$$s_t \approx (1 - b) \sum_{j=0}^{\infty} b^j E_t[\tau_{t+j}] = (1 - b) \sum_{j=0}^{\infty} b^j \tau_t = \tau_t.$$

If instead  $a'_2x_t$  is I(1), the first PDV in expression (9) will dominate as  $b \rightarrow 1$ , because the first one is multiplied by  $(1 - b)$ . We can then use logic that is similar to the above to establish that  $s_t$  behaves like a RW.

A model that fits into the above framework is the following. Suppose we have the money demand relation

$$m_t = p_t + \gamma y_t - \alpha i_t + v_{mt},$$

(which holds similarly for the foreign country), the RER formula

$$s_t = p_t - p_t^* + q_t$$

and the interest parity relationship

$$E_t s_{t+1} - s_t = i_t - i_t^* + \rho_t,$$

where  $\rho_t$  measures deviations from UIP (e.g., a risk premium or expectational error). Then

$$s_t = \frac{1}{1+\alpha} [m_t - m_t^* - \gamma(y_t - y_t^*) + q_t - (v_{mt} - v_{mt}^*) - \alpha \rho_t] + \frac{\alpha}{1+\alpha} E_t s_{t+1}. \quad (10)$$

Here the discount factor is  $b = \alpha/(1+\alpha)$ . The literature has typically estimated  $\alpha$  between 15 and 60, so  $b$  is very close to 1.

Instead of the money demand equation, we can assume that Home and Foreign CBs follow Taylor rules that put some weight on the bilateral NER's deviation from some target. Then a similar expression for the ER can be derived, with  $\alpha$  replaced by the coefficient on the ER gap in the Taylor rule.

The authors note that the efficient markets model of Samuelson (1965) does not predict a random walk in ERs. Indeed, this model predicts that UIP holds with  $\rho_t = 0$ , so a population regression of  $\Delta s_{t+1}$  on  $i_t - i_t^*$  should give a coefficient on 1. Instead, Engel and West's result states that the coefficient will get arbitrarily close to zero as  $b \rightarrow 1$ .

**Empirics** The authors have thus argued that a failure to beat a RW in short-term forecasting exercises is not damning for macro fundamental models. How then to test the theory? Because  $s_t$  is a PDV of expected future fundamentals, it should provide extra forecasting power for predicting these fundamentals, as long as it contains *some* extra information above what is contained in lags of the fundamental in question.

The authors use quarterly data on bilateral U.S. exchange rates with the other six G-7 countries. Five measures of fundamentals  $f_t$  are used:  $m_t, p_t, i_t, y_t, m_t - y_t$ . They are each examined separately. First, it's found that a Dickey-Fuller test can't reject a unit root in any of the five measures. Consistent with their theory, changes in the NER are essentially uncorrelated and very volatile. There does not appear to be cointegrating relations between the ER and the fundamentals, indicating that the unobserved fundamentals are I(1). The null of  $\Delta s_t$  not causing  $\Delta(f_t - f_t^*)$  is rejected at the 10% level in 12 out of the 30 cases (six countries times five measures of  $f_t$ ). The strongest rejections pertain to prices. On the other hand, the null that the fundamentals don't cause the ER can't be rejected in 28 of the cases. There is therefore moderate evidence that ERs help predict fundamentals, at least compared to the evidence of reverse causation.

Finally, the authors write

$$s_t = \sum_{j=0}^{\infty} b^j E_t f_{t+j} + \sum_{j=0}^{\infty} b^j E_t z_{t+j} = F_t + U_t,$$

where  $F_t$  is the observed component and  $U_t$  the unobserved one. Two empirical proxies for  $F_t$  are generated:

$$F_{1t} = E \left( \sum_{j=0}^{\infty} b^j E_t f_{t+j} \middle| f_t, f_{t-1}, \dots \right), \quad F_{2t} = E \left( \sum_{j=0}^{\infty} b^j E_t f_{t+j} \middle| f_t, s_t, f_{t-1}, s_{t-1}, \dots \right).$$

The former is computed using a univariate AR, while the second uses a bivariate VAR. They try different discount factors and different choices for the observed fundamental  $f_t$ . Regarding univariate results,  $\Delta F_{1t}$  is actually negatively correlated with  $\Delta s_t$ . Regarding the bivariate results, the correlation between  $\Delta F_{2t}$  and  $\Delta s_t$  is positive and around 0.1–0.25. Since the fundamentals are used one by one, it’s not surprising that the correlation isn’t close to 1, but there’s likely still a large unexplained component to ERs.

## 12.4 Rogoff (2008): “Comment on ‘Exchange Rate Models Are Not as Bad as You Think’”

**Summary** Rogoff doesn’t see it as controversial that ERs are forecastable at very short and very long horizons. However, the big challenge is the 1 month to 1 year horizon, which is the most relevant for policy. Chen and Rogoff (2003) showed that so-called “commodity currencies” (e.g., Australia, New Zealand and Canada) are somewhat forecastable even at this horizon. Rogoff doesn’t think that the RW theorem of Engel and West (2005) fully explains why macro models fare poorly out of sample. For example, if

$$s_t = (1 - b)a'_1 x_t + ba'_2 x_t + bE_t s_{t+1},$$

and  $i_t - i_t^* = E_t \Delta s_{t+1} + \rho_t$ , then

$$(1 - b)(s_t - a'_1 x_t) = b\rho_t + b(i_t - i_t^*),$$

where  $a'_2 x_t$  has been associated with the risk premium  $\rho_t$ . This type of equation for the ER was considered by Meese and Rogoff without much success. But because expectations of future ERs are embodied in the observed interest rate spread, there is no concern about RW behavior of the ER in the above model.

## 13 Forward Premium, Time-Varying Risk

### 13.1 Alvarez, Atkeson and Kehoe (2009): “Time-Varying Risk, Interest Rates, and Exchange Rates in General Equilibrium

**Summary** To explain the forward premium puzzle, it’s necessary to have the risk premium be even more volatile than interest rate differentials and the NER must be expected to appreciate when the risk premium falls. Because aggregate consumption tends to be smooth in most models, representative agent models can’t generate enough variability in the risk premium. The authors instead develop a two-country, cash-in-advance model with flexible prices and endogenously limited participation in asset markets. The model has a variable SDF despite aggregate C being almost constant, since the SDF is determined by the marginal investor. MP changes the inflation rate, which increases the benefit of participation. As more households participate, the effect of a given money injection on the active households’ MU is reduced, which causes a fall in the risk premium.

**Intuition** The only source of uncertainty is the money growth rate. Households face an idiosyncratic fixed cost of transferring resources between the asset and goods markets. *Inactive* HHs (who choose not to pay the fixed cost) must consume their RMB, and so their consumption is decreasing in the money growth rate. As the money growth rate increases, more HHs pay their fixed cost and become active. *Active* HHs trade in a complete set of contingent contracts and therefore all have the same level of consumption. As their number increases, the social planner allocates more resources to them, although the average (deadweight) fixed cost paid to participate also increases. Hence, the consumption of active HHs is increasing and concave in the money growth rate. This implies that the MU of active HHs is more variable at lower money growth rates than higher ones. Because the SDF is determined by the MU of active HHs, this means that the ER risk premium is variable and decreasing with the money growth rate.

To explain the FPP, it remains to be shown that an increase in the money growth rate causes an expected ER appreciation that is smaller than the fall in the risk premium. An increase in the money growth rate increases the expected inflation rate, which c.p. would cause the NER to be expected to depreciate, but the authors show that for not-too-large increases in the money growth rate, the RER appreciates enough to offset the expected inflation effect (in a model without limited participation and flexible prices, the RER wouldn't change).

**Theory: FPP** The risk premium  $p_t$  is given by the difference between the Home currency return on holding a Foreign bond minus the return of holding a Home bond:

$$p_t = i_t^* + E_t \Delta e_{t+1} - i_t,$$

where the ER  $\exp(e_t)$  is in Home per Foreign units. If the ER follows a RW,  $E_t \Delta e_{t+1}$  is constant, and so all movements in the RP are movements in interest rate differentials,  $\text{Var}(p_t) = \text{Var}(i_t - i_t^*)$ . In the data, when one regresses  $\Delta e_{t+1}$  on  $(i_t - i_t^*)$  on U.S. data vis-à-vis other countries, the estimated regression coefficient is typically negative, so that the NER is not in fact quite a RW. However, this leads to the *forward premium puzzle*: Suppose

$$\text{Cov}(i_t - i_t^*, \Delta e_{t+1}) = \text{Cov}(i_t - i_t^*, E_t \Delta e_{t+1}) \leq 0. \quad (11)$$

By the expression for the RP, this implies

$$0 \geq \text{Cov}(i_t - i_t^*, p_t + i_t - i_t^*) = \text{Var}(i_t - i_t^*) + \text{Cov}(i_t - i_t^*, p_t),$$

and so by C-S,

$$\text{Var}(i_t - i_t^*) \leq |\text{Cov}(i_t - i_t^*, p_t)| \leq \sqrt{\text{Var}(i_t - i_t^*) \text{Var}(p_t)} \Rightarrow \text{Var}(p_t) \geq \text{Var}(i_t - i_t^*).$$

So the RP has to be even more variable than it would have to be if the ER were a RW. Similarly, (11) implies

$$\text{Cov}(E_t \Delta e_{t+1}, p_t) \geq \text{Var}(\Delta e_{t+1}) > 0$$

and by C-S,

$$\text{Var}(p_t) \geq \text{Var}(E_t \Delta e_{t+1}),$$

so the ER is expected to appreciate when the RP falls, but the movement in the RP tends to be larger than the expected appreciation.

### 13.2 Bachetta and van Wincoop (2010): “Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle”

**Summary** In the data, high-interest currencies tend to predictably appreciate over the next 5–10 quarters before slowly depreciating (“delayed overshooting”), more or less reverting to UIP over the long run. The authors develop a two-country GE model in which overlapping generations of investors make one portfolio decision for the next  $T$  periods. In between decisions, the portfolio shares are passively rebalanced.

Consider a persistent, mean-reverting rise in the Foreign interest rate. Foreign bonds are more attractive, so new investors have a higher demand for those. Over the next few periods, in equilibrium, new generations of investors keep buying Foreign bonds, while older investors (who earn higher returns on their holdings of Foreign bonds) passively sell them to keep portfolio shares constant. However, investors correctly foresee that in the future, when the interest rate differential is lower, the aggregate demand for Foreign bonds will shift inward, causing an expected future depreciation. This expected future depreciation will eventually outweigh the interest rate differential. Thus, the delayed overshoot pattern emerges. The model has noise traders that create realistic volatility and limited predictability. Due to the high risk, the cost of active management (calibrated to real-life fees) is sufficiently high that it is in fact optimal for investors to re-optimize their portfolio shares infrequently.

The authors consider an extension in which infrequently trading investors only base their decisions on the current interest rate differential (as in a carry trade). This generates even more overshooting (closer to the data), as investors keep demanding Foreign bonds for longer without heeding the expected future depreciation.

**Motivation** Banks trade extensively in the Forex market but hold virtually no overnight positions. Mutual funds’ currency positions are constrained since they are only allowed to trade within certain asset classes.

### 13.3 Brunnermeier, Nagel and Pedersen (2009): “Carry Trades and Currency Crashes”

**Summary** The carry trade entails going long in a high-interest “investment” currency financed by a short position in a low-interest “funding” currency. The authors present empirical evidence on the carry trade that is consistent with a theory of currency traders as being subject to funding constraints that limit their ability to erase UIP deviations and occasionally force them to prematurely unwind their positions, leading to asymmetric crash risk. On average, the carry trade is profitable in expectation and its Sharpe ratio dominates that for equities. However, positive interest rate differentials predict future negative skewness of ER movements, i.e., disproportionate downside risk. Speculators’ (i.e., investors that engage in futures markets for reasons other than commercial hedging) positions increase with the interest rate differential, and net speculative positions predict increased negative skewness. Times of higher global risk, as measured by the VIX or TED spread, coincide with carry trade losses and reduced speculative carry positions, while the expected future excess carry return increases. Using options data, the authors find that the price of insuring against crash risk rises after carry trade losses (despite the conditional skewness moving closer to zero), perhaps because investors who are forced to unwind their positions become more cautious. Controlling for the VIX or TED spread, the predictive power of interest rates for excess

returns decreases, possibly resolving part of the FPP. Finally, the authors show that currencies with similar interest rates co-move closely, indicating that funding constraints have a large global component and thus can't be diversified away.

**Options data** The authors use so-called *risk reversals*, which is the implied volatility of an out-of-the-money call minus the implied volatility of an equally out-of-the-money put, both written on the Foreign currency. Such an options position pays off when the Foreign currency appreciates a lot, while they lead to negative payoff when it depreciates a lot. If the risk-neutral conditional distribution of the ER is symmetric, the risk reversal is zero. Otherwise, its sign equals that of the risk-neutral conditional skew of the ER. One can think of the risk reversal contract as providing insurance against a large Foreign appreciation, i.e., it's the price of insuring against the crash risk associated with a carry trade that uses the Home currency as investment currency and the Foreign currency as Funding currency.

### 13.4 Burnside (2009): “Comment on ‘Carry Trades and Currency Crashes’”

**Summary** The author holds a favorable view toward the Brunnermeier et al. (2009) paper, although he has reservations about the quantitative ability of the underlying theory that the empirics are supposed to support. First, Burnside criticizes other explanations of the FPP. The Lustig and Verdelhan (2007) thesis is dismissed, because their estimated SDF is apparently uncorrelated with the purported consumption risk factors. “Peso problem” (rare disaster) explanations are also dismissed, since hedged portfolio carry trades tend to generate about the same average return as unhedged ones (Burnside et al., 2008). Turning to the Brunnermeier et al. (2009) explanation, Burnside notes that the relationship between the forward premium and skewness is stronger using realized skewness than using options-implied skewness. This calls into question whether speculators really anticipate changes in crash risk. Because the story depends so much on time-varying conditional ER skewness, and this variation seems modest in the data, it's important to evaluate the funding-constraint model quantitatively to see whether it can generate enough amplification.

**Theory** Burnside gives a short exposition of the FPP. The return of an investment strategy that converts Home to Foreign currency, invests in Foreign bonds and then converts back to Home currency is

$$R_{t+1}^c = (1 + i_t^*) \frac{S_{t+1}}{S_t}.$$

$S_t$  is measured in Home per Foreign units. *Covered* interest parity requires

$$\frac{1 + i_t}{1 + i_t^*} = \frac{F_t}{S_t}.$$

Absence of arbitrage implies the existence of an SDF  $M_{t+1}$  such that  $1 = E_t[R_{t+1}M_{t+1}]$  for any gross return  $R_{t+1}$ . This yields  $1 + i_t = 1/E_t M_{t+1}$ . Inserting  $R_{t+1} = R_{t+1}^c$  and rearranging, one gets

$$\frac{F_t}{S_t} E_t M_{t+1} = E_t \left[ M_{t+1} \frac{S_{t+1}}{S_t} \right] \Rightarrow E_t \delta_{t+1} = f_t - \frac{\text{Cov}_t(M_{t+1}, \delta_{t+1})}{E_t M_{t+1}},$$

where  $\delta_{t+1} = (S_{t+1} - S_t)/S_t$  and  $f_t = (F_t - S_t)/S_t$ . The last term on the RHS above is the risk premium  $p_t$ . The *unbiasedness hypothesis* says that the forward rate  $F_t$  is an unbiased predictor of

the future spot rate  $S_{t+1}$ , which we see is equivalent with  $p_t = 0$ . From covered interest rate parity, we also have

$$f_t = \frac{1 + i_t}{1 + i_t^*} - 1 \approx i_t - i_t^*.$$

So we can write  $E_t \delta_{t+1} \approx i_t - i_t^* - p_t$ , and UIP is also approximately equivalent with  $p_t = 0$ .

### 13.5 Burnside (2011): “Carry Trades and Risk”

**Summary** The author considers various explanations for the FPP. He constructs various carry trade portfolios and estimates linear factor models by GMM. First, it’s found that common factor models that price U.S. equities—such as the CAPM (single factor is excess market return), CCAPM (single factor is consumption growth), Fama-French three-factor model (factors are market return, size premium SML and value premium HML) and Yogo (2006) extended EZ-CCAPM—do not price carry returns, as the betas are mostly insignificant and pricing errors too large (i.e., J-test rejects). Second, it’s found that some proposed models with factors constructed from currency returns do seem to perform well. As factors, the literature has used the average carry return, the HML carry return, global currency volatility and global currency skewness for high-interest currencies. While such factor models price carry returns, they perform badly on stock returns. This could just be due to market segmentation, but the currency-factor models also fail to price other currency-based portfolios such as momentum strategies. Hence, while these findings are suggestive of a connection between carry returns and factors such as volatility and skewness, Burnside argues that it’s unsatisfactory that no unifying SDF specification for both stock and currency returns has been found. Previous papers have suggested that conditional betas grow large during times of overall crisis, but Burnside uses rolling regressions to show that the degree of magnification is insufficient, as many stock market declines did not feature negative carry returns (although the 2008 crisis did). Finally, the author views “peso problem” models (estimated using options data) favorably, although he cautions that these models will need to explain the correlations between ERs, volatility and skewness, and should also work for the stock market. This is difficult, because rare consumption disasters ought to *reduce* the cross-sectional spread in betas (Julliard and Ghosh, 2010).

### 13.6 Engel (2011): “The Real Exchange Rate, Real Interest Rates, and the Risk Premium”

**Summary** The FPP literature finds that high-interest currencies tend to appreciate in the short run, thus offering high expected carry returns. This suggests that those countries’ bonds are risky. On the other hand, a separate part of the international macro literature has documented that when a country’s relative real interest rate rises, its currency tends to be stronger than its long-term average, and indeed by more than UIP would imply. Because the RER will have to revert to its long-term average (under stationarity), this suggests that such a country’s bonds are *less* risky than others. Engel gives a decomposition of the RER into a component due to expected future short-term real interest rate differentials and a component due to expected future excess returns (the latter component vanishes if UIP holds). Both these components, which are estimated using VARs, are negatively correlated with the RER in the data. Existing models designed to explain the FPP imply that after an interest rate hike, the NER appreciates gradually over time. However, the above fact about the RER implies that the NER should in fact jump by *more* than implied by UIP initially and then ultimately revert back to a long-run level, so any risk-based explanation must

include a risk reversal, by which the conditional correlation of excess returns and the real interest rate differential shifts sign at some point. Engel argues that this pattern can't be generated by a one-factor model of the expected depreciation and the interest rate differential, since such models automatically imply proportionality between excess returns and the real interest rate differential. However, Engel also shows that some more sophisticated models have a hard time generating the risk reversal, as these models are typically engineered to generate a risk premium that is very sensitive to interest rates in the short run, without worrying about the long-run implications.

**Theory and empirics** Let  $s_t$  be the log NER in units Home per Foreign. Let  $\lambda_t = i_t^* + E_t \Delta s_{t+1} - i_t$  be the excess return on Foreign bonds. The FPP literature has documented that typically  $\text{Cov}(\Delta s_{t+1}, i_t - i_t^*) < 0$ , which by definition implies  $\text{Cov}(\lambda_t, i_t - i_t^*) < 0$ . Let the RER be  $q_t = s_t + p_t^* - p_t$  and the ex ante real interest rate be  $r_t = i_t - E_t \pi_{t+1}$ . Then

$$\lambda_t = r_t^* - r_t + E_t q_{t+1} - q_t. \quad (12)$$

Suppose  $r_t - r_t^*$  and  $\lambda_t$  are stationary with means  $\bar{r}, \bar{\lambda}$ . Suppose also  $E[\Delta q_{t+1}] = 0$ . Then

$$q_t - E_t q_{t+1} = -(r_t - r_t^* - \bar{r}) - (\lambda_t - \bar{\lambda}).$$

Iterating forward, we get

$$q_t - \lim_{j \rightarrow \infty} E_t q_{t+j} = -R_t - \Lambda_t, \quad R_t = \sum_{j=0}^{\infty} E_t [r_{t+j} - r_{t+j}^* - \bar{r}], \quad \Lambda_t = \sum_{j=0}^{\infty} E_t [\lambda_{t+j} - \bar{\lambda}].$$

If the RER is stationary (which is debatable in the data; certainly the half-life is long), we have  $\lim_{j \rightarrow \infty} E_t q_{t+j} = \bar{q}$ , the long-run mean of the RER.  $R_t$  is the prospective real interest rate differential, while  $\Lambda_t$  is the *level risk premium*. Note that  $\Lambda_t \equiv 0$  if UIP holds, as in the classic Dornbusch model. In this model, a high real interest rate is associated with a comparatively strong currency in real terms due to the favorable investment opportunities in the Home country.

Engel estimates VARs to obtain empirical counterparts to  $r_t$ ,  $R_t$  and  $\Lambda_t$ . Unsurprisingly,  $\text{Cov}(R_t, r_t - r_t^*) > 0$  for most currency pairings with the U.S. dollar. It's more surprising that also  $\text{Cov}(\Lambda_t, r_t - r_t^*) > 0$ . Since the FPP result  $\text{Cov}(\lambda_t, r_t - r_t^*) < 0$  holds robustly (both in nominal and real terms), we must see a *risk reversal*  $\text{Cov}(\lambda_{t+j}, r_t - r_t^*) > 0$  for some  $j$ , and this is indeed the case in the data for sufficiently long horizons. Since  $\text{Cov}(\Lambda_t, r_t - r_t^*) > 0$ , the short-run reaction of the RER to the interest rate is *more* sensitive to the real interest rate than if UIP held.

The above evidence is inconsistent with most existing models. For example, suppose  $E_t \Delta q_{t+1} = a_1 \phi_t$ ,  $r_t - r_t^* - \bar{r} = c_1 \phi_t$ , so the expected real depreciation and the real interest rate differential are driven by the same factor  $\phi_t$ . Then the excess return equation (12) implies that also  $\lambda_t = g_1 \phi_t$  for some constant  $g_1$ . But then  $\text{Cov}(\lambda_t, r_t - r_t^*) = c_1 g_1 \text{Var}(\phi_t)$  and no risk reversal can occur.

Even most models with more than two factors will have trouble squaring with the empirical evidence. The reason is that these models have been constructed to generate a risk premium that is very sensitive to the (factors driving the) interest rate differential, such that  $\text{Var}(\lambda_t) > \text{Var}(E_t \Delta q_{t+1})$ . But if the model doesn't have an explicit way that the long-run mechanisms differ from the short-run ones, the risk reversal doesn't occur. Models of investor underreaction, such as Bachetta and van Wincoop (2010), are able to separate short- and long-run behavior, but the results above indicated that the RER initially reacts *more* than what UIP would imply to an interest rate shock.

### 13.7 Farhi, Fraiberger, Gabaix, Ranciere and Verdelhan (2009): “Crash Risk in Currency Markets”

**Summary** It’s argued that a small probability of adverse movements in the SDF can explain a significant fraction of the excess return on carry trades. A complete markets model is developed in which the Home and Foreign log nominal SDFs are driven by (1) partially cross-country correlated Gaussian processes plus (2) a perfectly cross-country correlated Poisson disaster process with a very small arrival rate. Assuming that a disaster does not occur in the sample, the model delivers closed-form solutions for short-dated options (i.e., as the maturity tends to 0), interest rates, excess carry returns and risk reversals.

Expected carry returns can be written as the sum of a Gaussian risk premium and a disaster risk premium. A *hedged* carry trade is a zero-investment trade in which one borrows in the funding currency and uses that money to both buy insurance in the form of a put option and to buy the investment currency. The model implies that the expected return on a hedged carry trade can also be written as two components. However, both risk premia are reduced due to the presence of insurance: The disaster risk is completely eliminated (if one assumes that only the investment currency is subject to disaster risk), while the Gaussian risk is ameliorated, and the degree of amelioration depends on the put’s delta. Interest rates are increasing in disaster risk (as investors seek compensation to hold the local currency), and Home’s risk reversal (price of OTM put minus equally OTM call) is positive if its crash risk is higher than in Foreign. Putting the above together, the model predicts that (1) currency excess returns increase with interest rates, and (2) increases in risk reversals are associated with contemporaneous ER depreciations (due to higher perceived crash risk) but expected future positive returns (due to higher risk compensation).

Using GMM to estimate the model on monthly data, the authors find that about a quarter of the carry trade risk premium is due to priced disaster risk. The authors see the experience in the Fall of 2008 as an illustration of a rare disaster, and indeed carry trades suffered large losses over this crisis period.

**Empirics** The authors form currency portfolios to empirically test the predictions. Hedged portfolios have lower, but still positive, excess returns than unhedged portfolios. The implied volatilities of high-interest currencies are larger than those for low-interest currencies, indicating that option markets price a depreciation risk for high-interest currencies. Risk reversals increase monotonically with interest rates.

The model’s structure is taken to the data using GMM. The disaster risk premium is about 1–3% annually, whereas the Gaussian risk premium is 3–6%. A J-test leads to a p-value of 0.28.

There is a strong contemporaneous link between risk reversals and ERs, as predicted, but there is limited evidence of risk reversals providing extra ER predictability.

### 13.8 Lustig and Verdelhan (2007): “The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk”

**Summary** The UIP condition is violated in the data. For a U.S. investor, higher foreign interest rates predict higher excess returns (except in the case of very high inflation currencies). The authors show that this compensates the investor for U.S. consumption risk: high foreign interest rate currencies tend to depreciate against the dollar when U.S. consumption growth is low, while low foreign interest rate currencies appreciate. These statistics are obtained by assigning currencies to

8 continually rebalanced portfolios, sorted after interest rate differential with the U.S. dollar. The highest-interest portfolio earns a large positive excess return, while the lowest-interest portfolio earns a negative return. The very highest-interest portfolio is dominated by high-inflation currencies, and these actually have more modest excess returns (more in line with UIP). A consumption Euler equation SDF approach with EZ preferences over durable and non-durable consumption (Yogo, 2006) is used to price the returns, with estimates formed by Fama-MacBeth. The implied approximate linear factor model prices the returns well, explaining about 80% of the variation in returns. However, it suffers from the EPP, as it needs a CRRA of about 100, the same as the model gets from U.S. equity and bonds returns. The betas on C growth are small for the low-interest portfolios and high for the high-interest portfolios. The authors relate their findings to the Backus-Smith (1993) puzzle: *Conditional* on interest rates, the correlation between ERs and C growth is in fact high.

## A Appendix: Abbreviations

The following are some of the abbreviations used.

- B-S: Backus-Smith
- C: consumption
- CA: current account
- CB: central bank
- CD/C-D: Cobb-Douglas
- CRS: constant returns to scale
- C-S: Cauchy-Schwarz
- CU: currency union
- DIS: dynamic IS
- DRS: decreasing returns to scale
- D-S: Dixit-Stiglitz
- EoS: elasticity of substitution
- EPP: equity premium puzzle
- ER: exchange rate
- EZ: Epstein-Zin or Eurozone (depending on context)
- FOC: first-order condition
- FPP: forward premium puzzle
- GE: general equilibrium
- GHH: Greenwood-Hercowitz-Huffman
- H: hours
- HH: household
- HP: Hodrick-Prescott
- I: investment
- K: capital
- L: labor
- LCP: local currency pricing
- LOP: law of one price
- LRPT: long-run pass-through
- M: imports
- MC: marginal cost
- MP: monetary policy
- MPL: marginal product of labor
- MPK: marginal product of capital

- MRPT: medium-run pass-through
- MU: marginal utility
- N: labor
- NER: nominal exchange rate
- NERPT: nominal exchange rate pass-through
- NFA: net foreign assets
- NGDP: nominal GDP
- NGM: neo-classical growth model
- NK: New Keynesian
- NKPC: New Keynesian Phillips Curve
- NX: net exports
- OE: open economy
- OLG: overlapping generations
- PT: pass-through
- PCP: producer currency pricing
- PDV: present discounted value
- PPP: purchasing power parity
- RD: regression discontinuity
- RER: real exchange rate
- RMB: real money balances
- RP: risk premium
- RW: random walk
- S: saving
- SDF: stochastic discount factor
- SOE: small open economy
- TB: trade balance
- ToT: terms of trade
- UIP: uncovered interest parity
- X: exports
- Y: output
- ZCB: zero-coupon bond