Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Explaining China's Trade Imbalance Puzzle

Yi Wen

Working Paper 2011-018A
http://research.stlouisfed.org/wp/2011/2011-018.pdf

August 2011

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

Explaining China's Trade Imbalance Puzzle*

Yi Wen
Federal Reserve Bank of St. Louis
&
Tsinghua University

This Version: October 12, 2011

Abstract

The current global-imbalance literature (which explains why capital flows from poor to rich countries) is unable to explain China's foreign asset positions because capital cannot flow out of China under capital controls. Hence, this literature has not succeeded in explaining China's large and persistent trade imbalances with the United States. A closely related but deeper puzzle that this literature fails to address is China's high household saving rate despite an astonishingly rapid income growth rate. This paper shows that a modified (and calibrated) Melitz (2003) model can qualitatively and quantitatively explain China's trade surplus and its massive accumulation of low-yield foreign reserves. The simple infinite-horizon model is also consistent with the stylized fact that high saving is the *consequence* of high growth (Modigliani and Cao, 2004), which the permanent income theory and global-imbalance literature fail to predict.

Keywords: Chinese Economy, Foreign Reserve, High Saving Rate Puzzle, Global Imbalance, US Trade Deficit.

JEL Codes: E21, F11, F30, F41, F43, O16.

*This is a substantially revised version of my earlier working papers (Wen, 2009a, 2011). I am grateful to B. Ravikumar for his detailed comments and suggestions. I also thank Franklin Allen, David Andolfatto, Kaiji Chen, Nicolas Coeurdacier, Silvio Contessi, Yannick Kalantzis, Nobu Kiyotaki, Narayana Kocherlakota, Patrick Pintus, Zheng Song, Mark Spiegel, Shang-Jin Wei, Xiaodong Zhu, and seminar participants at the Federal Reserve Bank of St. Louis, the 2010 Conference on Chinese Economy (Shanghai), the Marseille Conference on "Asset Prices, Credit and Macroeconomic Policies", the 2011 System Committee on International Economic Analysis at the Atlanta Fed, and the Columbia-Tsinghua Workshop in International Economics 2011 for comments. I also thank Mingyu Chen and Xin Wang for research assistance and Judy Ahlers for editorial assistance. The usual disclaimer applies. Correspondence: Yi Wen, Research Department, Federal Reserve Bank of St. Louis, St. Louis, MO, 63104. Phone: 314-444-8559. Fax: 314-444-8731. Email: yi.wen@stls.frb.org.

1 Introduction

Recent advances in international trade theory attempt to explain the patterns of trade at the country, industry, and firm levels.[1] However, it has little to offer on the question of why trade may be imbalanced between countries over a prolonged period. According to Deardorff (2010), standard trade theories cannot explain why developing countries (such as China) with comparative advantages in producing future goods have been running trade surpluses while industrial nations (such as the United States) with comparative advantages in producing present goods are running trade deficits. That is, from the viewpoint of the current trade theories, the steady increase in China's trade balance from a small deficit (-$1.1 billion) in 1978 (the beginning year of economic reform) to a gigantic surplus ($400 billion) in the first half of 2009 is a puzzle.[2]

The international finance approach to the trade imbalance puzzle is to model conditions under which China's domestic savings exceed domestic investment so that China should experience capital outflows. Between 1978 and 2009, China's foreign exchange reserves (mostly U.S. dollars) increased dramatically, from $2 billion to $2.4 trillion—a more than 1,000-fold expansion—making China the world's largest holder of foreign exchange reserves. Consequently, the accounting identity between the current account and the capital account implies that China automatically runs a trade surplus.[3]

Such an approach faces two fundamental challenges. First, why is China's astonishing 45% investment-to-GDP ratio not able to fully absorb its domestic savings? How can the notoriously high household saving rate in China be explained when household income has been growing at 10% per year and the real return to household saving is negative? Second, even if a sufficiently high saving rate in excess of the investment rate can be generated by a model, such a saving-investment imbalance may still run counter to China's reality: (i) If domestic savings generated from the model are in the form of fixed capital, then the model would predict outflows from China to developed countries in the form of foreign direct investment (FDI). However, the reality indicates the opposite: it is the industrial economies that have been sending FDI to China over the past few decades. (ii) If the excess domestic savings are in the form of local currency or demand deposits, such financial capital cannot leave China because China has capital controls and the renminbi (RMB) is not internationally convertible. To purchase U.S. government bonds, for example, Chinese savers must first have dollars in hand. Therefore, to explain China's trade surplus through an international-

[1] See, e.g., the seminal works of Bernard, Eaton, Jensen, and Kortum (BEJK, 2004), Chaney (2008), Melitz (2003), and the rapidly growing literature following these papers.

[2] The bulk of the surplus results from trade with the U.S. and takes the form of final consumption goods.

[3] See, Caballero, Farhi, and Gourinchas (2008); Mendoza, Quadrini, and Rios-Rull (2009); and Song, Storesletten, and Zilibotti (2011), among others.

finance approach, the "excess savings" must be in the form of consumption goods instead of assets. But this conclusion revisits the trade imbalance puzzle: Why is China exporting more goods than it is importing?

Resolving the trade imbalance puzzle thus requires not only starting with a basic trade framework with exporters as the main actors, but also find new ways to model the micro-incentives for the demand of low-yield liquid assets (such as dollars and U.S. government bonds). Conventional approaches to model the demand for low-yield assets, such as those with cash-in-advance constraints or money-in-the-utility assumptions, are inadequate since the optimal asset demand in such models is proportional to consumption, which is inconsistent with Chinese data: The consumption-to-income ratio in China has been declining while the foreign reserve-to-GDP ratio has been increasing over the past decades.

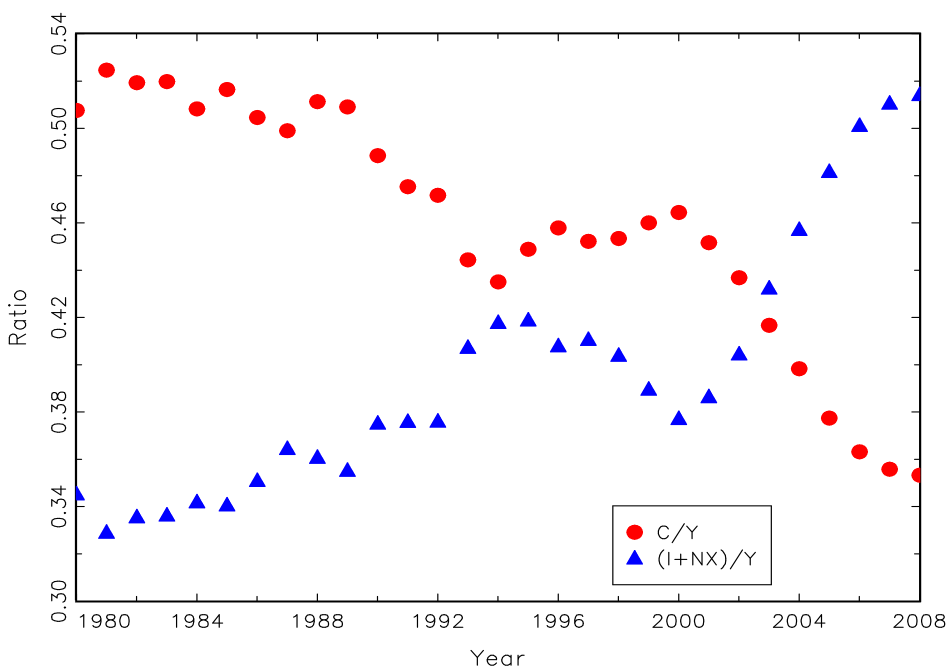

Figure 1. Aggregate Saving Rate (▲) and Consumption Ratio (●).

This paper develops a simple, analytically tractable model to qualitatively and quantitatively explain China's trade surplus and foreign-reserve buildups in recent decades. The model builds on the current trade framework with heterogeneous firms (see, e.g., Eaton and Kortum, 2002; Bernard et al., 2003; Melitz, 2003; and Ghironi and Melitz, 2005), but we start where the existing trade literature ends; namely, we take the existence of an export sector (with heterogeneous firms) as given and embed it into a monetary growth model featuring incomplete markets (e.g., Bewley, 1980; Lucas, 1980; Aiyagari, 1994) and long-run productivity growth.

Our theory predicts that when exporters (i) face uninsured risks that remain constant over time as income grows and (ii) are subject to borrowing constraints, their marginal propensity to save increases with income growth. That is, the more income they earn, the larger portion of the income will they save, in sharp contrast to the conventional wisdom based on Friedman's (1957) permanent income hypothesis (PIH). Indeed, during the 30 years of China's rapid economic growth, its private consumption-to-GDP ratio has fallen from roughly 50% in 1980 to 35% in 2008 (see the C/Y ratio in Figure 1), while government spending as a fraction of total national income has remained roughly constant at about 14%. Thus, Chinese consumers have reduced their propensity to consume significantly despite the rapidly rising per capita income. China's national saving rate (private investment plus net exports over GDP) has also increased steadily over the past 30 years, from 34% to 51% (see the (I+NX)/Y ratio in Figure 1).

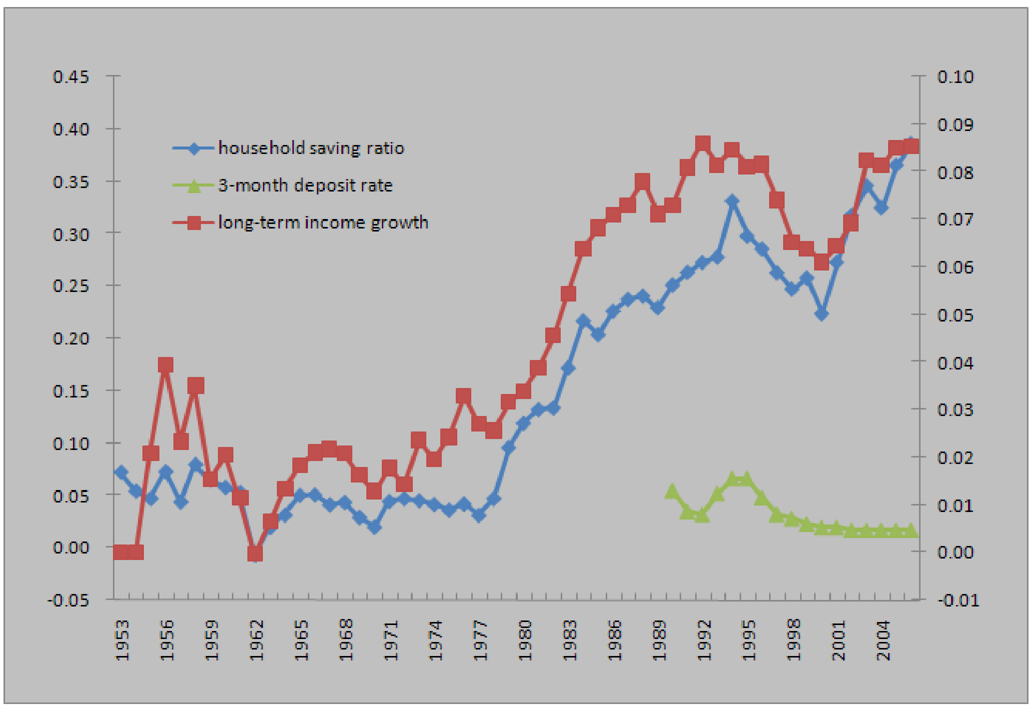

Figure 2. Household Saving Ratio (left scale) and Income Growth (right scale)

Our model is also consistent with a well-known stylized fact that is not well explained by the existing literature: Lagged income growth is a strong predictor of future saving rates—or high saving is the consequence of high growth (e.g., Modigliani and Cao, 2004). Figure 2 depicts the household saving ratio (blue diamonds, left axis) and the long-term growth rate of household income (red squares, right axis) for the 1953-2006 period. The household saving rate is defined as the ratio between net wealth changes and disposable income, and the long-term income-growth rate is defined as the average growth rate of household income of the past 14 years, following Modigliani and Cao

(2004). The figure shows that the household saving rate tracks the past average income growth rate very closely and has increased steadily since 1978. The average saving rate was about 5% before 1978 when the average income growth rate was about 2% per year. After the economic reform, the saving rate increased to over 35% after 1994 when the average rate of household income growth reached near 10% per year. The extraordinary Chinese saving rate and its positive association with income growth are not unique. Similar high saving rates have also been observed in other emerging economies during their rapid-growth periods, such as Japan in the 1960-70s, Hong Kong in the 1980s, and Taiwan and South Korea in the 1990s (see, e.g., Wang and Wen, 2011).

Our analysis suggests that given the elastic labor supply from China's rural areas and the rapid growth in export income (e.g., due to a comparative advantage in labor costs and an expanding world market for Chinese goods), financial frictions in China will naturally lead to massive trade surplus and foreign-reserve buildups. The faster the export income grows, the larger is the foreign reserve-to-GDP ratio. In fact, China's total imports-to-exports ratio has fallen during the fast-growth period, from 1.6 in 1985 to around 0.8 in 2008. That is, while exports have grown at a double-digit annual rate, total imports have failed to keep pace. As a result, China's trade surplus and foreign reserves have exploded. Therefore, the data suggest that Chinese households might have been saving an increasingly larger portion of their income (including dollars earned from international trade) to provide the safety net and self-insurance unavailable to them from markets.[4] Such a precautionary saving motive is also supported by the fact that the bulk of the household saving in China consists of bank deposits despite low interest rates.[5] The average real interest rate in China remained essentially at zero or negative values in the post-reform period. For example, the average nominal 3-month deposit rate was 3.3%, the average 1-year rate was 5.6% (see the line with green triangles in Figure 2, left axis), while the average inflation rate was about 6% in the 1991-2007 period.[6]

Many analysts believe that the steady increase in America's trade deficit with China is the consequence of a significantly undervalued Chinese currency. In other words, Chinese goods are too cheap relative to American goods. Hence, Americans can buy many Chinese goods while the Chinese can barely afford American goods. Indeed, some economists and politicians in the U.S. have alleged that the Chinese government has been manipulating its currency to deliberately achieve a large trade surplus and an excessive amount of foreign reserves. This paper argues that the trade

[4] Even though China has experienced impressive economic growth over the past 30 years since its economic reform and entering the global marketplace, its financial reform has not kept pace with its economic growth. Because of the lack of social safety nets and severely underdeveloped insurance and financial markets, Chinese workers must save excessively (including dollars earned from international trade) to insure themselves against idiosyncratic uncertainty, such as bad income shocks, unemployment risk, accidents, and unexpected spending for housing, education, health care, and so on.

[5] Wen (2009b) shows that in China and India the share of cash and bank deposits accounts for more than 90% of total household financial wealth.

[6] Data for the interest rates before 1990 are not available.

imbalance puzzle has little to do with the real exchange rate.

The rest of the paper is organized as follows. Section 2 presents a benchmark model with heterogeneous exporters (entrepreneurs) facing borrowing constraints and uninsured risks, characterizes general equilibrium of the model, and derives decision rules for consumption, saving, and production in closed form. Section 3 studies the relationship between the aggregate saving rate and the income growth rate along a balanced growth path and a stochastic growth path, respectively. Section 4 uses the model to predict China's trade surplus and foreign-reserve buildups. Section 5 explores the implications of the model for other dimensions. Section 6 and Appendix B (not for publication) discusses some robustness issues pertaining to the model structure and parameter values. Section 7 reviews the related literature. Section 8 concludes the paper with remarks for future research.

2 Benchmark Model

2.1 Caveats

Some important caveats are in order before we describe the model. A key assumption in this paper is that idiosyncratic risk facing entrepreneurs remains constant over time despite long-run growth—that is, the percentage changes in entrepreneur income (consumption) are stationary regardless of the rate of income (consumption) growth. This assumption implies that idiosyncratic income shocks have constant variances and are multiplicative to the income *level* but additive to the logarithm of income. This assumption is consistent not only with the data but also with a large body of the empirical and theoretical segments of the literature that measure and model idiosyncratic uncertainty.[7] In contrast, if idiosyncratic shocks are assumed to be additive to the income level (instead of its logarithm), then idiosyncratic risk will diminish to zero on a balanced growth path, rendering it irrelevant for the analysis.

Of course, a constant income risk implies that the variance of absolute income (i.e., the gap between actual income level and its exponential growth trend) is increasing over time. However, an increasing variance of the absolute income does not imply an increasing income risk because the correct measure of risk is the percentage of fluctuations in income. A multiplicative shock to the income trend simply means that the variance of logged income (i.e., the gap between logged

[7] A large body of the empirical literature finds that the idiosyncratic component of logged household income (consumption) is stationary over time (see, e.g., DeBacker, Heim, Panousi, and Vidangos, 2010, Gottschalk and Moffitt, 1994, 2009; Moffitt and Gottschalk, 1995, 2008; Haider, 2001, Heathcote, Perri, and Violante, 2010; Sabelhaus and Song, 2009; Shin and Solon, 2008; among others). For example, using a large panel of tax returns from the Internal Revenue Service, DeBacker et al. (2010) find that the variance (volatility) of logged household income (after controlling for time trend) has remained stationary over time in the United States for the period 1988-2006; namely, the standard deviation of percentage changes in income has not diminished but instead has remained constant for American households.

income and its time trend) is constant over time.[8] The question is this: Will rational entrepreneurs (households) change their marginal propensity to save in a borrowing constrained economy when the growth rate of their income changes but the income risk remains constant (that is, the fluctuations in income remain constant relative to trend regardless of the growth rate)? It is precisely this lack of such knowledge that has created the big puzzle regarding the high household saving rate in China despite its well-known 10% per year income growth rate over the past few decades.[9]

Another caveat is that uncertainty in households' spending needs (such as unexpected expenditures for health care, education, housing, and unpredictable spending shocks related to personal accidents and property damages) may be an equally or even more important source of idiosyncratic risk than labor income risk, especially in developing countries where insurance and credit markets are poorly developed. Wages—the main source of household labor income for the majority of the working population—are sticky and highly predictable by individuals. On the other hand, unexpected expenditures for health care, education, housing, and other spending needs in daily life tend to grow more rapidly than wage income (because wage income tends to lag inflation in developing countries while spending costs track inflation very closely and may even outgrow inflation). Also, costs in health care and housing are large relative to monthly or annual income and fluctuate significantly more than wage income.

As an example, the majority of the Chinese population is not effectively covered by any form of health care insurance system (Wang, Xu, and Xu, 2007). Data from the 1998 China National Health Services Survey indicate that medical costs increased by 625% for each clinic visit and 511% for each hospital admission during the 1990-1998 period (Liu, Rao, and Hsiao, 2003). The average cost per hospital admission (over-night stay) in 1998 was as high as 2,891 yuan, which is equivalent to 42.5% of annual GDP per capita and 71% of annual consumption per person in that year. These authors also estimate that the rapid increase in out-of-pocket medical spending in China raised the number of rural households living below the poverty line by more than 44%. So medical expenditure has become an important source of transient poverty in rural China. This explains why Chamon and Prasad (2010) found empirically that the high saving rates of Chinese households across all demographic groups "are best explained by the rising private burden of expenditures on housing, education, and heath care" (p. 93).[10] This also explains why consumption expenditures

[8] In other words, an x dollar increase (or decrease) in absolute income provides no information about income risk unless it is translated into percentage terms.

[9] Given the income growth trend, a higher income risk would have led to a higher household saving rate under borrowing constraints. However, this assumption of increasing risk is not needed in this paper to explain China's high saving rate even though adding it to our model would further enhance our results. Instead, we show that even if the degree of idiosyncratic risk remains constant relative to income trend, a higher income growth will lead to a higher saving rate. This is in contrast to the model of Chamon, Liu, and Prasad (2010), who explain the high household saving rate in China by assuming that the degree of idiosyncratic risk increases over time not only absolutely but also relative to the trend of income growth.

[10] Quantitatively speaking, Chamon and Presad (2010) found that health risk is the most important driving force of household savings because expenditures on housing and education involve less uncertainty and tend to cancel the

in developing countries are more volatile than income (see, e.g., Mark and Gopinath, 2007). These considerations lead us to assume in this paper two sources of idiosyncratic uncertainty: one stemming from entrepreneur income and the other from the subsistence level of household consumption (or preference shocks).[11]

2.2 Model Setup

This is a small open-economy model. The home country (e.g., China) is denoted by H and the rest of the world by F. For simplicity, we assume (i) complete international specialization in that home-produced goods are for export only and home residents consume only foreign-produced goods, and (ii) F is large enough that the price of tradable goods is not affected by H's exports and imports. Let P_t^* denote the nominal price of goods sold in country F in terms of foreign currency (dollars), which is also the price that H households pay for imported goods from abroad. So trade involves the same goods and the law of one price holds. The foreign inflation rate, $\pi_t = \frac{P_t^* - P_{t-1}^*}{P_{t-1}^*}$, is assumed to be constant over time.[12]

As in Melitz (2003), there is a continuum of heterogeneous entrepreneurs in country H indexed by $i \in [0,1]$. Each entrepreneur operates a linear production technology, $Y_t(i) = A_t v_t(i) N_t(i)$, where $N_t(i)$ is labor demand,[13] $v_t(i)$ is an idiosyncratic shock to firm-level productivity, and A_t represents aggregate productivity that grows over time according to

$$A_t = A_0 \left(1 + \bar{g}\right)^t . \tag{1}$$

To make the model analytically tractable, however, we change the Melitz production function to

$$Y_t(i) = A_t N_t(i) + v_t(i) A_t, \tag{2}$$

where the idiosyncratic shock $v(i)$ can be reinterpreted as cost shocks if taking negative values. Notice that $v(i)$ is multiplicative to the aggregate growth trend (A_t) so that idiosyncratic risk remains constant (instead of diminishing) over time as the economy grows.

Entrepreneurs choose optimal paths of consumption, saving, and production plans subject to financial constraints. Labor mobility within the country and perfect competition among firms imply that the real wage in the labor market is given by $W_t = A_t$.[14] Since the economy has a balanced growth path, we can transform the model into a stationary economy by scaling (normalizing) all

savings for such purposes across cohorts (see Wang and Wen, 2011).

[11] We follow the existing literature by using preference shocks as a shortcut for consumption demand shocks.

[12] The results are not sensitive to this assumption.

[13] Lu (2010) documents that Chinese exporters are predominantly labor-intensive firms and they sell the bulk of their output to foreign markets.

[14] In this paper, "entrepreneurs" and "households" are used interchangeably.

endogenous variables except hours worked, $N_t(i)$, by the growth factor, $(1+\bar{g})^t$. All normalized variables are denoted by lowercase letters (e.g., $x_t \equiv \frac{X_t}{(1+\bar{g})^t}$). Note that the rescaled real wage is given by $w_t \equiv A_0$.

There is an international reserve currency (called dollars) in the model that can also serve as the means of payment for tradable goods. However, we do not impose the standard cash-in-advance constraint or the money-in-utility assumption to induce money demand. Instead, we motivate money demand by precautionary saving motives as in Bewley (1980). Thus, if households opt to hold foreign currency, it is purely for precautionary saving reasons by carrying it as a store of value. This modeling strategy allows us to combine precautionary saving behavior with money demand without making other assumptions about why people hold money or low-yield liquid assets.

Nominal (dollar) income earned from exports can be either saved or spent. Foreign reserves are assumed to be kept by entrepreneurs (households) instead of by the government. Alternatively, we can allow entrepreneurs to sell dollars to the government and use the proceeds to purchase local government bonds. In this way, all foreign reserves will then be held by the government in country H instead of by households, but the results are similar.[15] For simplicity, we assume that foreign reserves earn a zero nominal interest rate; hence, the real rate of return to saving is the inverse of the inflation rate in country F.[16]

Entrepreneurs are borrowing and short-sale constrained, as in Bewley (1980). That is, they cannot hold negative amounts of any assets. Each entrepreneur is subject to two types of idiosyncratic shocks, one stemming from income and another from spending needs (preferences). These two idiosyncratic shocks are assumed to be i.i.d. and independent of each other. The household income is given by $[N_t(i) + v_t(i)]\, A_t = N_t(i)W_t + v_t(i)A_t$. The additive term $v_t(i)A_t$ captures income outside regular wage earnings and is called "bonus income" in this paper—which reflects the entrepreneur's cost shocks or operating losses (gains) unrelated to labor productivity. Notice that the random shock to bonus income can take negative values. Empirical studies by Wang (2011) show that nonwage income is an important component of household disposable income in both rural and urban China and this component is far more variable than wage earnings ($N_t(i)W_t$).

The simplest way to model fluctuations in household spending needs (or subsistence consumption) is to introduce an idiosyncratic random variable $\theta(i)$ into the utility function, $U(C_t(i) - \theta_t(i)W_t)$, where $\theta_t(i)$ denotes shocks to the subsistence consumption level and is idiosyncratic across households. This random variable captures idiosyncratic uncertainty in consumption de-

[15] Caballero, Farhi, and Gourinchas (2008, p. 361) point out that "most of these reserves are indirectly held by the local private sector through (quasi-collateralized) low-return sterilization bonds in a context with only limited capital account openness."

[16] A low interest rate goes against our results. The results would be enhanced if we let the government bear the inflation risk by holding foreign reserves or paying positive interest on households' foreign currency accounts.

mand, such as unexpected expenditures related to accidents, illness, housing, education, and so on. Notice that the subsistence spending shock is also multiplied by the growth trend (the real wage $W_t = A_0 (1+\bar{g})^t$) to capture the notion of constant risk. The intuition is that basic spending needs rise with economic growth. For example, costs in health care and education increase with income growth. Also, richer people or higher-earning income groups tend to buy bigger houses in more expensive areas, send their children to more expensive (e.g., private) schools, visit hospitals with better facilities, or buy medical insurance with more comprehensive coverage. Such spending needs rise with income and are often viewed as necessary instead of optional.

Alternatively, we can also assume that the preference shock $\theta(i)$ is multiplicative to household utility, $\theta_t(i)U(C_t(i))$. Because household consumption level grows over time, this alternative modeling strategy avoids the need of scaling the shocks by growth trend A_t. Although the results do not depend on whether the preference shocks are additive or multiplicative, a multiplicative preference shock is more tractable than additive preference shocks; therefore, this paper assumes multiplicative preference shocks.[17]

A salient feature of China's labor market is the existence of abundant cheap labor in rural areas. This feature implies that labor supply in China is highly elastic. However, there also exist tremendous frictions in China's labor market (i.e., the existence of the so called Hukou system that restricts farmers to live and work freely in the cities unless they have signed labor contracts with companies). To capture these characteristics of the Chinese labor market in the simplest possible way within a neoclassical framework, two additional assumptions are made: (i) The utility function is quasi-linear (as in the indivisible labor models of Hansen, 1985; Rogerson, 1986; and Lagos and Wright, 2005); (ii) hours worked are predetermined (quasi-fixed) within each period—i.e., they are determined before observing any idiosyncratic shocks in each period. Together these two assumptions imply that the elasticity of labor supply is high across periods (or at the aggregate level along the extensive margin) but low within periods (or at the micro level along the intensive margin). These implications are also qualitatively consistent with empirical estimates of labor supply elasticities at macro and micro levels in developed countries (Chang and Kim, 2006). In addition, these assumptions (especially the quasi-linear utility function) greatly facilitate analytical tractability of our heterogeneous-agent model.[18]

Let $M_t(i)$ denote the stock of nominal money balances held by entrepreneur i by the end of period $t-1$, $C_t(i)$ real consumption for imported goods in period t, and $N_t(i)$ hours worked in period

[17] It is well known in the literature that analytical tractability is extremely hard to obtain in models with heterogeneous agents and borrowing constraints even in the very simple endowment model of Bewley (1980) with a single idiosyncratic shock.

[18] Technically speaking, because of quasi-linear preference, the assumption of predetermined hours supply ensures that entrepreneurs cannot use labor income to fully insure themselves against idiosyncratic risks.

t. Applying the aforementioned transformation, we have $m(i) \equiv \frac{M_t(i)/P^*_{t-1}}{(1+\bar{g})^t}$, $m'(i) \equiv \frac{M_{t+1}(i)/P^*_t}{(1+\bar{g})^{t+1}}$, and $c_t(i) \equiv \frac{C_t(i)}{(1+\bar{g})^t}$. Entrepreneur i's problem is to solve

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \theta_t(i) \log c_t(i) - aN_t(i) \right\} \tag{3}$$

subject to

$$c(i) + (1+\bar{g})\, m'(i) \leq \frac{1}{1+\pi} m(i) + w\left[N(i) + v(i)\right] \tag{4}$$

$$m'(i) \geq 0, \tag{5}$$

and $N(i) \in \left(0, \bar{N}\right)$, where $m' \equiv m_{t+1}$. The cumulative density function of $\theta(i)$ is denoted by $\Pr\left[x \leq \theta\right] = F(\theta)$ with support $\left[\underline{\theta}, \overline{\theta}\right]$, where $F(\cdot)$ is differentiable. The distribution of $v(i)$ follows

$$v_t(i) = \begin{cases} \xi, & \text{with probability } p \\ -\xi, & \text{with probability } 1-p \end{cases}, \tag{6}$$

where $\xi \geq 0$ controls the size of the income shock.

Equation (4) is the budget constraint, which states that total income earned from exporting goods to F can be used to finance purchases of foreign-produced goods and the accumulation of foreign assets (foreign reserves $m' - m$), subject to the borrowing constraint (5). Without loss of generality, we assume $\alpha = 1$ in the utility function. Note the following implications of the model:

(i) If there exist only borrowing constraints (but with idiosyncratic uncertainty), since $\beta < 1+\bar{g}$, entrepreneurs would set consumption equal to total export income (i.e., no need to save). Hence, trade would always be balanced and there would be no accumulation of foreign reserves.

(ii) If there exist only idiosyncratic risks (but without borrowing constraints), entrepreneurs would set consumption equal to permanent income by borrowing heavily from outside against their growing future income. Since agents are much richer in the future than at the present, country H would opt to run a trade deficit with country F, as predicted by the PIH and Deardorff (2010). However, with both borrowing constraints and idiosyncratic risks, the outcome is completely different, as shown below.

2.3 Characterization of General Equilibrium

A general equilibrium is defined as a balanced growth path characterized by the following conditions:

(i) A sequence of decision rules for each entrepreneur i, $\{c_t(i), m_{t+1}(i), N_t(i)\}_{t=0}^{\infty}$, such that given the sequence of prices $\{P_t^*, W_t\}_{t=0}^{\infty}$, these decision rules maximize each entrepreneur's lifetime utility subject to constraints (4) and (5).

(ii) A sequence of demand function for labor, $\{N_t\}_{t=0}^{\infty}$, such that given the sequence of prices $\{P_t^*, W_t\}_{t=0}^{\infty}$, the demand function maximizes firms' profits.

(iii) The law of large numbers holds and all resource constraints are respected:

$$\int N_t(i)di = N_t \tag{7}$$

$$\int C_t(i)di + \frac{\int M_{t+1}(i)di - \int M_t(i)di}{P^*} = Y_t, \tag{8}$$

where equation (7) represents the labor market-clearing condition, and equation (8) represents a balanced budget in the tradable-goods sector. Because this is a small open economy, there is no market-clearing condition for foreign currency. Hence, equation (8) states that all revenues generated from exports ($P^* Y_t$) are used to finance either imports ($P^* \int C_t(i)di$) or the accumulation of foreign reserves. In other words, the (nominal) trade deficit is represented by net increase in foreign reserves, $M_{t+1} - M_t$.

(iv) The transversality condition holds: $\lim_{t\to\infty} \beta^t \frac{\frac{1}{P_t^*}\int M_{t+1}(i)di}{W_t} = 0$.

Proposition 1 *Denoting* $x \equiv \frac{1}{1+\pi} m(i) + wN(i)$ *as cash in hand net of the bonus income* $v(i)w$*, the decision rules of consumption, asset demand, and cash in hand for entrepreneur* i *are given by*

$$c(i) = \begin{cases} \min\left\{\frac{\theta(i)}{\theta_1^*}, 1\right\}[x - \xi w], & \text{with probability } 1-p \\ \min\left\{\frac{\theta(i)}{\theta_2^*}, 1\right\}[x + \xi w], & \text{with probability } p \end{cases} \tag{9}$$

$$(1+\bar{g})\, m'(i) = \begin{cases} \max\left\{\frac{\theta_1^* - \theta(i)}{\theta_1^*}, 0\right\}[x - \xi w], & \text{with probability } 1-p \\ \max\left\{\frac{\theta_2^* - \theta(i)}{\theta_2^*}, 0\right\}[x + \xi w], & \text{with probability } p \end{cases} \tag{10}$$

$$x = \theta_1^* \frac{(1+\bar{g})(1+\pi)}{\beta} w + \xi w, \tag{11}$$

where the cutoff variables $\{\theta_1^*, \theta_2^*\}$ *are determined by the following two equations:*

$$1 + \bar{g} = \frac{\beta}{1+\pi} R(\theta_1^*) \tag{12}$$

$$\theta_2^* = \theta_1^* + 2\xi R(\theta_1^*)^{-1}, \tag{13}$$

where the function $R(\cdot)$ *satisfies*

$$R(\theta_1^*) \equiv (1-p)\left[F\left(\theta_1^*\right) + \int_{\theta \geq \theta_1^*} \frac{\theta}{\theta_1^*} dF(\theta)\right] + p\left[F(\theta_2^*) + \int_{\theta \geq \theta_2^*} \frac{\theta}{\theta_2^*} dF(\theta)\right] > 1. \tag{14}$$

Proof. See Appendix A.I. ■

The decision rules for consumption and saving in Proposition 1 are quite intuitive. Optimal consumption is a concave function of a target wealth adjusted for bonus incomes, $x_t \pm \xi w$, with the marginal propensity to consume given by the function, $\min\left\{\frac{\theta}{\theta^*}, 1\right\}$. When the urge to consume is low ($\theta(i) < \theta^*$), the marginal propensity to consume the bonus-adjusted income is less than 1; when the urge to consume is high ($\theta(i) \geq \theta^*$), the marginal propensity to consume equals 1 and the individual does not save in this period. Therefore, saving is a buffer stock: The entrepreneur saves for a rainy day in the case of low consumption demand ($m'(i) > 0$ if $\theta(i) < \theta^*$), anticipating that consumption demand may be high in the future. The optimal saving rate also depends on income shock: It is higher if $v(i) = \xi$ and lower if $v(i) = -\xi$ because the cutoff $\theta_2^* > \theta_1^*$. These properties of the saving behavior, which relates the saving rate to a bonus-income-adjusted wealth target ($x_t \pm \xi w_t$), have been conjectured by some Japanese economists based on empirical observations of the high Japanese household saving rates in the 1970s (see, e.g., Ishikawa and Ueda, 1984, and the literature therein) but have never been proved analytically in theory; they are only numerically indicated by the buffer-stock saving literature (see, e.g., Deaton, 1991; Aiyagari, 1994, and Carroll, 1992, 1997).

Notice that the optimal cash in hand (net worth) x is independent of i (i.e., identical across entrepreneurs regardless of income and preference shocks). The intuition for this is that (i) x is predetermined before the realizations of $\theta(i)$ and $v(i)$, and (ii) the labor supply $N(i)$ can adjust elastically to target any level of cash in hand before idiosyncratic shocks are realized. That is, since all entrepreneurs face the same distribution of idiosyncratic shocks at the beginning of each period and must choose hours worked in advance, the quasi-linear utility function makes it feasible and optimal that entrepreneurs set their labor supply to target the same level of cash in hand regardless of the individual's history of asset holdings. That is, x is optimal ex ante given the distributions of $\theta(i)$ and $v(i)$ and the macroeconomic environment (e.g., the real wage, real interest rate, the inflation rate, and so on), regardless of initial wealth at the beginning of each period, $\frac{m(i)}{P^*}$. This property is key to obtaining closed-form solutions but the main insights of this paper do not hinge critically on this property.

Also notice that $R(\cdot) > 1$ because $F(\theta_j^*) + \int_{\theta \geq \theta_j^*} \frac{\theta}{\theta_j^*} dF(\theta) > 1$ for $j = 1, 2$. It captures the liquidity value (premium) of cash under borrowing constraints and uninsured risks. Hence, the effective rate of return to saving is determined by the real interest rate compounded by the liquidity premium R. The liquidity premium is decreasing in the cutoff θ_1^*: $\frac{dR}{d\theta_1^*} < 0$. That is, with a higher cutoff, the liquidity constraint is less likely to bind, so the liquidity value of savings is lower.

Thus, the left-hand side (LHS) of equation (12) is the shadow marginal cost of saving: The opportunity cost of not consuming a rapidly rising income is proportional to the income growth rate. The right-hand side (RHS) of the equation measures the effective rate of return to saving, including the real interest rate (β) and the liquidity premium (R). Hence, optimal saving of an asset is determined by equating the marginal cost with the marginal benefit, taking into account the liquidity premium of the asset. In equilibrium, the liquidity premium R is thus an increasing function of income growth $\bar{g}$.

The main intuition is that uninsured risk and borrowing constraints induce precautionary savings even if the real interest rate is negative ($\frac{\beta}{1+\pi} \leq \beta < 1$). Agents would want to maintain a stable buffer stock of savings relative to trend income because of the need for self-insurance when the degree of risk does not diminish with economic growth. Since income is a flow and savings a stock, when income grows, the stock-to-flow ratio would decline if the saving rate remain unchanged—which would hinder the buffer-stock function of savings and reduce the extent of self-insurance (since the degree of idiosyncratic uncertainty remains constant relative to growth trend). Thus, the liquidity premium R would increase dramatically unless the saving rate rises. In other words, a higher liquidity premium or shadow return to saving (due to income growth) would induce a higher saving rate along the balanced growth path (the steady state).

Using letters without index i to denote aggregate variables and by the law of large numbers, aggregate (or average) consumption and saving of entrepreneurs are given, respectively, by

$$c = (1-p)\, D(\theta_1^*)\, (x - \xi w) + pD(\theta_2^*)\, (x + \xi w) \tag{15}$$

$$(1+\bar{g})\, m' = (1-p)\, H(\theta_1^*)\, (x - \xi w) + pH(\theta_2^*)\, (x + \xi w)\,, \tag{16}$$

where the functions $\{D(\cdot), H(\cdot)\}$ are defined by

$$D(\theta_j^*) = \int_{\theta<\theta_j^*} \frac{\theta}{\theta_j^*} dF(\theta) + \int_{\theta \geq \theta_j^*} dF(\theta) \in (0,1)\,, \quad \text{for } j = 1,2 \tag{17}$$

$$H(\theta_j^*) = \int_{\theta<\theta_j^*} \frac{\theta_j^* - \theta}{\theta_j^*} dF(\theta) \in (0,1)\,, \quad \text{for } j = 1,2. \tag{18}$$

Note $D(\cdot) + H(\cdot) = 1$ because $D(\cdot)$ is the average marginal propensity to consume from cash in hand ($x \pm \xi w$) and $H(\cdot)$ is the marginal propensity to save. The equilibrium path of the model is characterized by the sequence $\{c, m', x, \theta_1^*, \theta_2^*\}$, which can be solved uniquely and explicitly from equations (11)-(16) once the distributions for $\theta(i)$ and $v(i)$ are specified.

3 Saving Behavior

3.1 Saving Along a Constant Balanced Growth Path

Clearly, the cutoffs $\{\theta_1^*, \theta_2^*\}$ determine the aggregate saving-to-income ratio. A higher cutoff implies a larger fraction of savers in the population versus non-savers since $\frac{\partial H_j}{\partial \theta_j^*} > 0$ and $\frac{\partial D_j}{\partial \theta_j^*} < 0$. More precisely, the saving rate τ in the economy is defined as the ratio of net changes in asset position to disposable income: $\tau_t = \frac{M_{t+1} - M_t}{P_t^* Y_t} = \frac{(M_{t+1} - M_t)/P_t^*}{(N_t + \bar{v}) W_t}$, where $\bar{v}W = (2p-1)\,\xi W$ denotes the expected bonus income.

Proposition 2 *Along a balanced growth path, the aggregate household saving rate is given by*

$$\tau = \frac{g\left[(1-p)\,\theta_1^* H_1 + p\theta_2^* H_2\right]}{(1+g)\left[(1-p)\,\theta_1^* + p\theta_2^*\right] - \frac{1}{1+\pi}\left[(1-p)\,\theta_1^* H_1 + p\theta_2^* H_2\right]}. \tag{19}$$

Proof. See Appendix A.II. ■

Notice that in the absence of income shocks ($\xi = 0$), we have $\theta_1^* = \theta_2^*$ and $H_1 = H_2$, so the saving rate simplifies to

$$\tau = \frac{\bar{g} H(\bar{g})}{1 + \bar{g} - \frac{1}{1+\pi} H(\bar{g})}, \tag{20}$$

as originally derived by Wen (2009a). This saving function is remarkably similar to that implied by the life-cycle model of Modigliani and Brumberg (1954).[19] However, the underling mechanisms are completely different. The life-cycle model relies on an aggregation (composition) effect across different cohorts to generate the positive relationship between income growth and the aggregate saving rate. To the extent that the economy is growing, workers' savings will increase relative to retirees' dissavings (as if the population is expanding); thus, the measured aggregate saving rate will increase with growth. This life-cycle mechanism, however, has not fared well empirically. For example, careful analysis of the implications of the life-cycle hypothesis has led Hayashi (1986) to reject it as a plausible theory of Japan's high saving behavior during its high growth period.[20]

[19] For example, in a two-period life-cycle model where the income in the retirement period is zero, the aggregate saving rate is given by $\frac{\beta}{1+\beta} \frac{\bar{g}}{1+\bar{g}}$, which is zero if $\bar{g} = 0$ and increases monotonically with $\bar{g}$. For a recent literature review on the life-cycle model and the growth-saving puzzle, see Modigliani and Cao (2004).

[20] The life-cycle hypothesis has also gained little support from Chinese data (see, e.g., Horioka and Wan, 2007; Chamon and Presad, 2010). The empirical failure of the life-cycle model in explaining the positive growth-saving relationship has pushed the literature to search for alternative explanations, such as habit formation (Carroll and Weil, 1994), high TFP growth (Chen, Imrohoroglu, and Imrohoroglu, 2006), and rising income and health care risks (Chamon and Presad, 2010; Chamon, Liu, and Presad, 2010).

Proposition 3 *The saving rate τ is a hump-shaped (inverted-U) function of income growth, increasing in $\bar{g}$ if $\bar{g} < \bar{g}^*$ and decreasing in $\bar{g}$ if $\bar{g} > \bar{g}^*$, where the threshold $\bar{g}^* \in (0, \infty)$ is strictly positive and bounded from above.*

Proof. See Appendix A.III. ■

For example, when $\bar{g} = 0$, we have $\tau = 0$ and $\frac{d\tau}{d\bar{g}} > 0$, as in a typical life-cycle model. By continuity, the saving rate increases with income growth for small values of $\bar{g}$. This proposition shows that higher income growth can lead to a higher saving rate instead of a higher marginal propensity to consume, in sharp contrast to the prediction of the PIH. The PIH predicts that forward-looking consumers should increase their marginal propensity to consume when they expect income to be permanently higher in the future. However, with uninsured uncertainty and borrowing constraints, this prediction is no longer necessarily correct when the growth rate of income is below a threshold level ($\bar{g}^*$).

The PIH is based on two critical assumptions: (i) Agents are able to consume their future income by borrowing, and (ii) agents do not face any uninsured risk. However, with borrowing constraints and uninsured risk, people are not able to consume their future income and they need to keep a buffer stock as self-insurance against idiosyncratic shocks.[21] The key insight of Proposition 2 is that under borrowing constraints and uninsured risk, the optimal saving rate will increase with income growth for empirically plausible growth rates, consistent with much of the empirical evidence.[22]

It is important to reiterate the intuition. Since saving provides liquidity, it yields a liquidity premium R (shadow rate of return). Because the degree of risk remains constant relative to the growth trend, savings can serve effectively as a buffer stock if and only if the stock-to-income ratio remains sufficiently high. However, since income is a flow, a higher income growth rate will lead to a lower stock-to-income ratio if the saving rate remains unchanged. Consequently, the liquidity premium will increase with $\bar{g}$, and a higher liquidity premium will induce a higher saving rate.

On the other hand, if the growth rate is sufficiently high, then the opportunity cost of not consuming the rapidly growing income outweighs the benefits of precautionary savings, causing the optimal saving rate to decline with growth (as predicted by the PIH). This hump-shaped concave saving function is dictated by a bounded liquidity premium R. Note that the function $R(\cdot)$ is bounded above by $\frac{E\theta}{\underline{\theta}} > 1$. To see this, recall that $\theta_2^* = \theta_1^* + 2\xi R^{-1} > \theta_1^* \geq \underline{\theta}$ (where $\underline{\theta}$ is the lower bound of the support in the distribution) and that R is decreasing in θ_1^*. So the maximum value

[21] That is, with uninsured risks, having a binding borrowing constraint by setting $s_{t+1}(i) = 0$ for all t is not optimal.

[22] See, for example, Modigliani (1970), Carroll and Weil (1994), Modigliani and Cao (2004), Horioka and Wang (2007), among many others.

of R is given by

$$R(\underline{\theta}) = (1-p)\,\frac{E\theta}{\underline{\theta}} + p\left[F(\theta_2^*) + \int_{\theta \geq \theta_2^*} \frac{\theta}{\theta_2^*} dF(\theta)\right] < (1-p)\,\frac{E\theta}{\underline{\theta}} + p\int_{\theta \geq \underline{\theta}} \frac{\theta}{\underline{\theta}} dF(\theta) + pF(\theta_2^*) < \frac{E\theta}{\underline{\theta}} + p. \tag{21}$$

Thus, there exists a maximum value of the growth rate $\bar{g}_{\max} = \frac{\beta}{1+\pi}R(\underline{\theta}) - 1 > \bar{g}^*$ such that if $\bar{g} \geq \bar{g}_{\max}$, R can no longer increase. In this case, the borrowing constraint (5) becomes binding for all households and nobody saves. Hence, the saving function $\tau(\bar{g})$ must be hump-shaped, increasing with $\bar{g}$ first and then decreasing with $\bar{g}$ for $\bar{g} \geq \bar{g}^* > 0$, and approaching zero as $\bar{g} \to \bar{g}_{\max}$.

Calibration. For a quantitative picture of the saving behaviors in the model, consider the following calibration exercise. The parameter a in the utility function has no effect on the saving rate, so we normalize $a = 1$. Let the time period t be a year and set $\beta = 0.96$. For tractability, we assume θ follows the Pareto distribution,

$$F(\theta) = 1 - \theta^{-\sigma}, \tag{22}$$

with $\sigma > 1$ and $\theta \in (1, \infty)$. A value of $\theta = \infty$ may indicate a life-threatening medical need, but the probability of such events is infinitely small or zero.[23] The results remain robust to alternative distributions, such as lognormal and uniform distributions. With Pareto distribution, we have $R = (1-p)\left[1 + \frac{1}{\sigma-1}\theta_1^{*-\sigma}\right] + p\left[1 + \frac{1}{\sigma-1}\theta_2^{*-\sigma}\right]$, $D(\theta_j^*) = \frac{\sigma}{\sigma-1}\theta_j^{*-1} - \frac{1}{\sigma-1}\theta_j^{*-\sigma}$, and $H(\theta_j^*) = 1 - D(\theta_j^*)$.

We assume $p = 0.5$, so that the average bonus income is zero for all households. This also implies that the income shock is symmetric around a zero mean. We set $\xi = 2$ to match the income inequality in China. However, the results are not sensitive to the value of ξ, suggesting that income shocks are not essential for our results (Figure 3). The parameter $\sigma = 1.3$ in the Pareto distribution is estimated by the method of moments (to be discussed in the next section). The parameter values are summarized in Table 1.

With the calibrated parameter values, the relationship between the aggregate household saving rate (τ) and the growth rate ($\bar{g}$) is graphed in Figure 3 (the line with circles). It shows (quantitatively) the two important predictions of the model discussed previously. First, a higher income growth leads to a higher saving rate even if the real rate of return to saving is negative ($\frac{\beta}{1+\pi} - 1 < 0$). For example, when the income growth rate is 1% per year, the saving rate is 6.5%. When the income growth rises to 10% per year, the saving rate increases to 22%, a more than 15 percentage-point increase.

[23] The absence of an upper bound on preference shock is assumed only for simplicity; an upper bound can be incorporated in the analysis without qualitatively changing the main results.

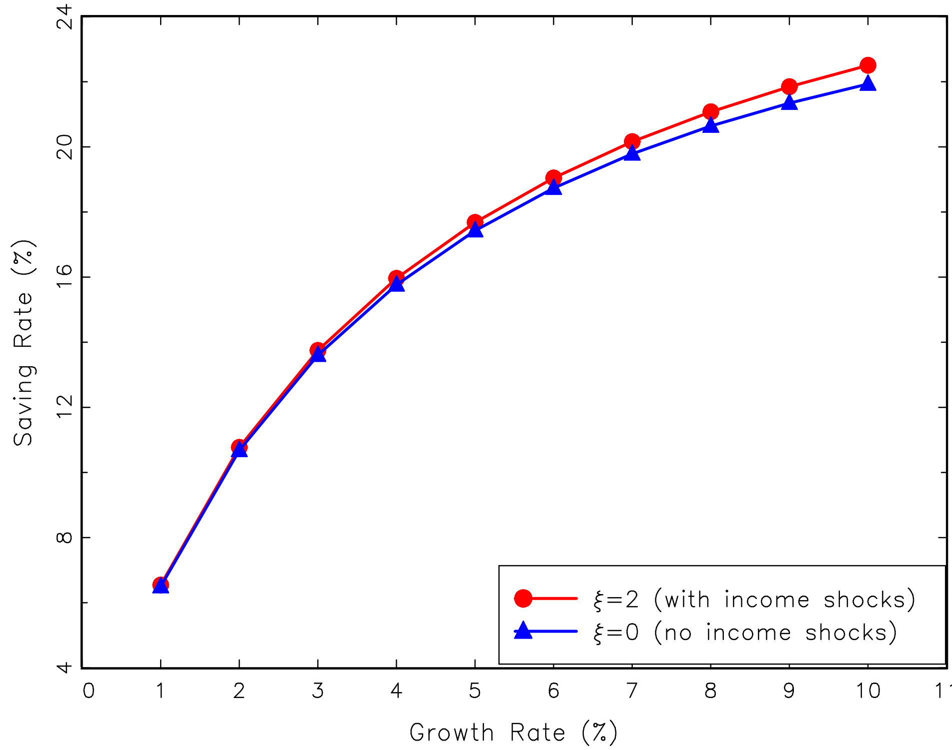

Figure 3. Saving Rate as a Function of Growth.

Second, the saving function is hump-shaped (i.e., inverted-U-shaped). Under the current calibration, the function will reach a maximum at the growth rate $\bar{g}^* = 24$ percent per year. The saving rate becomes flat beyond $\bar{g}^* = 0.24$ and declines slowly with $\bar{g}$. The saving rate will eventually approach zero as $\bar{g}$ increases to $g_{\max}$. This suggests that unless the growth rate exceeds 24% per year, the saving rate will always be positively associated with growth. Moreover, even if the growth rate is greater than 24% per year, the saving rate can still remain at very high levels. For example, the saving rate remains at 23% even when the income growth rate is as high as 40% per year.[24]

Such an inverted-U-shaped saving function differentiates the infinite-horizon, incomplete-market model from the life-cycle model of Modigliani and Brumberg (1954). Since the life-cycle model relies exclusively on an aggregation effect to generate the positive growth-to-saving relationship, its implied aggregate saving rate will increase without bound as the growth rate approaches infinity, in sharp contrast to the implication of this infinite-horizon model.

[24] When idiosyncratic risks in the model are gradually reduced to zero, the hump of the saving function will shift to the left and eventually disappear. For example, suppose the variances of income and preference shocks are tiny (with $\xi = 0.00001$ and $\sigma = 15$), the saving rate not only is very low (about 0.003% when income growth rate is 1% per year), but also declines rapidly to zero when income growth reaches 2.8% per year. That is, there is no need to save when uninsured risks are minimal. In this case, trade is always balanced. It can also be shown that the home country will run trade deficits with the rest of the world if there are no borrowing constraints (but with positive growth), consistent with the prediction of the PIH. Therefore, our model includes the PIH as a special (limiting) case.

If the foreign asset (m) is not liquid (i.e., cannot be adjusted instantaneously to buffer idiosyncratic shocks), then its rate of return must be higher than the discounted growth rate ($\geq \frac{1+\bar{g}}{\beta}$) to induce entrepreneurs to hold it. For example, if the growth rate is 10% per year and the discounting factor is $\beta = 0.96$, then any illiquid asset with a rate of return less than 14% per year will have a zero demand and, in such a case, trade is always balanced. This explains why China's foreign reserves are predominantly low-yield liquid foreign bonds.

The line with triangles in Figure 3 represents the case without income shocks (i.e., $\xi = 0$). It shows that income risk is not quantitatively important in generating the high saving rates in our model. This result is similar to that obtained by Krusell and Smith (1998) who showed in a neoclassical growth model with incomplete markets that labor income risk is not as important as preference shocks in explaining the income and wealth inequalities in the United States.

Proposition 4 *Without income shocks ($\xi = 0$), hours worked for each entrepreneur ($N_t(i)$) are strictly bounded within the open interval $\left(0, \bar{N}\right)$, where $\bar{N} = \beta \left(\frac{\sigma}{\sigma-1}\right)^2$ if $\bar{g} < \bar{g}_{\max}$ and $\bar{N} = a$ if $\bar{g} \geq \bar{g}_{\max}$.*

Proof. See Appendix A.IV. ■

With income shocks ($\xi > 0$), hours worked are still bounded above by a fixed number. However, whether $N(i) \geq 0$ is satisfied depends on the variance of income shocks and the other parameter values of the model. Numerical analysis indicates that under our current calibration in Table 1, $N_t(i)$ may hit the lower bound 0 or become negative occasionally but the probability (frequency) is too small to quantitatively affect our results.

3.2 Saving Along a Stochastic Growth Path

China's economic growth rate may not stay at 10% forever. If the high growth rate is transitory instead of permanent, then the corresponding saving rate would be much larger in the calibrated model. The intuition is simple: If entrepreneurs are already willing to increase their saving when facing a permanently higher income growth rate, the incentive to save is even stronger (because of consumption-smoothing motives) when the higher income growth is transitory.

This can be illustrated easily by impulse response analysis. Our model permits closed-form solutions even if the income growth rate is stochastic; hence, dynamic impulse response functions can be expressed analytically. Suppose aggregate technology grows according to the stochastic process, $A_t = (1 + g_t)A_{t-1}$, where the stochastic growth rate g_t satisfies the law of motion

$$\log\left(\frac{1+g_t}{1+\bar{g}}\right) = \rho_g \log\left(\frac{1+g_{t-1}}{1+\bar{g}}\right) + \varepsilon_t, \tag{23}$$

where $\bar{g} \geq 0$ is the mean and ε_t is an i.i.d. normally distributed process with zero mean. To compute the stochastic equilibrium path of the model, we rescale all variables (except N_t) by the level of technology A_{t-1}; we use A_{t-1}, instead of A_t, as the scaling factor so that the transformed stock variable m_t remains as a state variable that does not respond to changes in A_t in period t.[25] Using lowercase letters to denote the transformed variables, $x_t \equiv \frac{X_t}{A_{t-1}}$, the production function becomes $y_t = (1 + g_t) N_t$, the real wage becomes $w_t = 1 + g_t$, and the aggregate household resource constraint becomes

$$c_t + (1 + g_t) m_{t+1} = m_t + (1 + g_t) N_t. \tag{24}$$

Once the equilibrium decision rules and the dynamic paths of the stationary model are solved, we can then uncover the original level variables by the inverse transformation: $X_t \equiv x_t A_{t-1}$.

Proposition 5 *In a stochastic dynamic equilibrium, the year-to-year changes in foreign reserves are given by*

$$\begin{aligned} & M_{t+1} - M_t \qquad (25) \\ = \; & P_t^* Y_t \left[1 - \frac{(1-p)\, D(\theta_{1t}^*)\theta_{1t}^* R(\theta_{1t}^*) + pD(\theta_{2t}^*)\theta_{2t}^* R(\theta_{1t}^*)}{\theta_{1t}^* R(\theta_{1t}^*) + \xi - \frac{1}{(1+g_t)} \left[(1-p) H(\theta_{1t-1}^*)\theta_{1t-1}^* R(\theta_{1t-1}^*) + pH(\theta_{2t-1}^*)\theta_{2t-1}^* R(\theta_{2t-1}^*) \right]} \right], \end{aligned}$$

which is proportional to total nominal exports $P_t^* Y_t$ with a time-varying saving rate τ_t.

Proof. See Appendix A.V. ■

For example, suppose the stochastic growth rate g_t is i.i.d. with $\rho_g = 0$ and the steady-state growth rate is 5% per year. The impulse response function of the saving rate to an unexpected 5-percentage-point increase in income growth is plotted in the left panel in Figure 4, where the unexpected change in growth rate takes place in the 11th period. It shows that an unexpected jump in the rate of income growth from 5% to 10% raises the saving rate from 21% to 33%, a dramatic 12-percentage-point increase. In contrast, if this jump in income growth were permanent, the saving rate at the new steady state (with 10% income growth per year) would be just 26% (as shown in Figure 3 and the right panel in Figure 4). So the additional saving rate brought by the higher income growth is only 6 percentage points under a permanent growth shift instead of the 12 percentage points under a transitory change in the growth rate. Notice that the transitory change in the saving rate would be negative (instead of positive) in a standard PIH model without financial frictions, because a temporarily higher income growth is equivalent to a permanent increase in the income level.

[25] The particular methods of transformation do not affect the dynamics of the original variables.

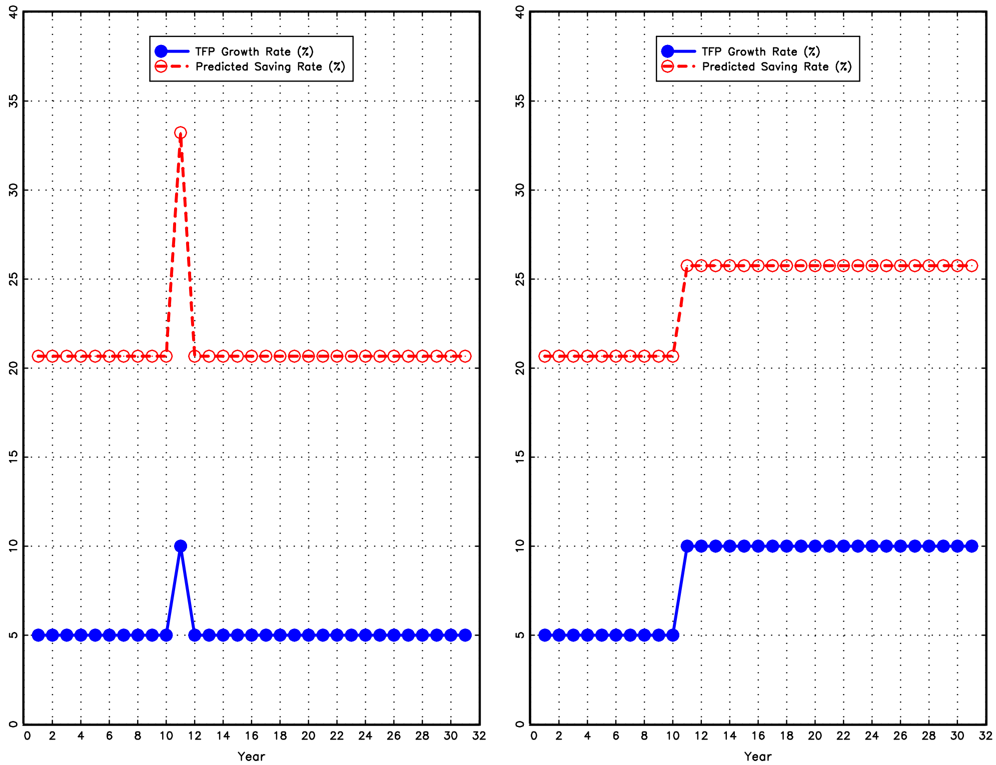

Figure 4. Impulse Responses of Saving Rate to Growth Rate Shock.

4 Predicting China's Trade Imbalance and Foreign Reserves

Data show that between 1978 and 2009, China's current account surplus increased dramatically, reaching \$426 billion (USD) in 2008. The bulk of the increase in the current account is due to a rapidly rising trade surplus. Associated with the rising current account is the massive buildup of China's foreign reserves. In our model, trade surplus is determined by entrepreneurs' precautionary saving in the tradable-goods sector. Because of uninsured risks and borrowing constraints, a substantial fraction of income earned from exports is saved, which leads to imbalances between exports and imports. Most importantly, the precautionary saving rate rises with the growth rate of income. When the growth rate of income is stochastic, so is the saving rate.

To calibrate the stochastic process $\{g_t\}$ in the model, the data for aggregate exports, the price deflator, and hours worked in the tradable-goods sector are needed. The growth rate of nominal exports is given by

$$\frac{P_t^* Y_t}{P_{t-1}^* Y_{t-1}} = (1 + \pi_t)(1 + g_t) \frac{N_t}{N_{t-1}}, \tag{26}$$

where $1 + \pi_t \equiv \frac{P_t^*}{P_{t-1}^*}$. Since data for the price deflator $(1 + \pi_t)$ and hours worked (N_t) in the tradable-goods sector are not available and since TFP growth typically mimics output growth, we follow the methodology in Durdu, Mendoza, and Terrones (2007) by approximating the growth rate of technology in the tradable-goods sector by the growth rate of total exports adjusted by a constant inflation rate.[26] With the estimated growth rate process $\{g_t\}$ in hand and assuming that g_t follows equation (23), then the mean growth $\bar{g}$, the autocorrelation ρ_g, and the variance σ_g^2 can all be estimated.

Notice that when g_t is i.i.d. (i.e., $\rho_g = 0$), the cutoffs $\{\theta_{1t}^*, \theta_{2t}^*\}$ are constant and the implied saving rate (τ_t) based on equation (25) will be highly contemporaneously correlated with g_t because $\frac{\partial R(\theta^*)}{\partial \theta^*} < 0$. On the other hand, if g_t is serially correlated (as in the data), then the cutoffs are time varying and the implied saving rate will be positively correlated with not only the current g_t but also the lagged g_t because high growth in the last period also tends to induce high saving in the current period.

Using Chinese data for nominal exports (measured in USD) for the 1978-2009 period, the estimated mean nominal growth rate is $\bar{g} = 0.17$. With an average inflation rate of 4% in the U.S., the real growth rate of the export income is about 13% per year. This estimate of real average TFP growth is also consistent with the empirical studies on Chinese manufacturing firms by Brandt, Biesebroeck, and Zhang (2009).[27] Data suggest that the AR(1) coefficient $\rho_g = 0.18$ and the variance of the residual $\sigma_\varepsilon^2 = 0.012$.

Table 1. Parameter Values

$\hat{\beta}$	$\hat{\sigma}$	$\hat{g} - \pi$	$\hat{\rho}_g$	$\hat{\sigma}_\varepsilon^2$	p	ξ	π
0.96	1.3	0.13	0.17	0.012	0.5	2.0	0.04

We ask whether the model is able to explain both the business cycle features and the growth trends of China's trade surplus and foreign exchange reserves. To this end, we estimate the key structural parameters by minimizing the distance between the time series implied by the model and the data. The parameter ξ cannot be estimated precisely, suggesting that income uncertainty is not as important as expenditure uncertainty in explaining the data. That is, setting $\xi = 0$ gives essentially identical results in our model (e.g., see Figure 3), indicating that bonus income shocks are not essential for our results. However, since we also want the model to be consistent with the income and consumption distributions in China, we set $\xi = 2$ to match the Gini coefficient of household income in China.

Given that the technology innovation ε_t may not exactly follow lognormal distribution, the parameters $\left\{\bar{g}, \rho_g, \sigma_e^2\right\}$ are reestimated using the method of moments. That is, we take only the

[26] We assume a 4% annual inflation rate but the results are not sensitive to this value.

[27] Brandt et al. (2009, Figure 2) report that the average TFP growth rate for manufacturing firms in China is about 9.6% for 1999-2006 and 13-14% for 2002-2006. Given that exporting firms are among the most productive enterprises in China, our estimate of a 13% real output growth rate is credible.

values of $\pi = 0.04$, $p = 0.5$, and $\xi = 2$ as given and estimate $\{\beta, \sigma, \bar{g}, \rho_g, \sigma_\varepsilon^2\}$ so that the model can best match the time-series data of total exports and the changes in foreign reserves in China. The estimated parameter values are reported in Table 1 and denoted by circumflex. It shows that the estimated values under the method of moments, $\{\hat{g} = 0.17, \hat{\rho}_g = 0.17, \hat{\sigma}_\varepsilon^2 = 0.012\}$, are nearly identical to their original values in the raw data. The estimated value of $\hat{\beta} = 0.96$ is also identical to standard calibrations in the literature.

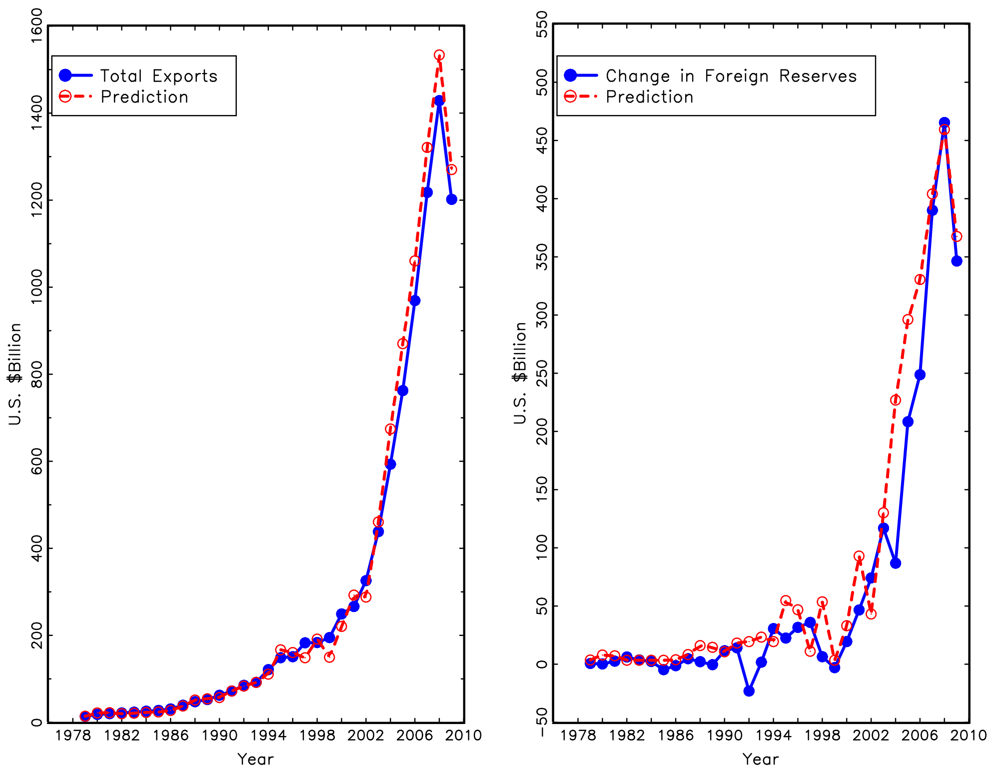

Figure 5. Predictions of Total Exports (left) and Changes in Foreign Reserves (right).

Under these parameter values, the predicted total exports income ($P_t^* A_t N_t$) and year-to-year changes in foreign reserves ($M_{t+1} - M_t$) in the model are shown in the left and right panels in Figure 5, respectively, where the solid lines in each panel represent data and dashed lines are predictions.[28] Despite having only a single aggregate shock, the model explains more than 90% of the data (e.g., the R^2 implied by ordinary least squares is 0.99 for total exports and 0.93 for exchange reserves). In particular, the model tracks the surge in total exports and foreign reserves since 2002 quite well (China joined the WTO in December 2001), mainly because stochastic changes in productivity growth can generate large swings in the saving rate along a transitional path. The model also tracks well the 2009 slump due to the recent U.S. financial crisis.[29]

[28] Labor (N_t) is predicted using equation (51) in Appendix A.V.

[29] The prediction for net exports is similar to that of foreign-reserve changes, as the two variables are equivalent in

The model can track China's trade surplus and foreign reserves well because income growth drives saving in the model. Hence, the saving rate (changes in foreign reserves or net exports) in the model can closely trace both the large acceleration in export income after China joined the WTO and the sharp decline in export income during the 2008 financial crisis. Both events have significantly changed the export sector's saving rate in the same direction. This result is reminiscent of the findings in Chen et al (2006), who show that aggregate income growth was the fundamental driving force behind movements in Japan's aggregate saving rate. However, their results are obtained under the assumption that high TFP growth leads to a higher real interest rate, which in turn stimulates household savings. In contrast, we generate the positive link between entrepreneurial saving and income growth despite the constant (negative) real interest rate.

In addition to matching the time series of trade surplus and foreign reserves, the model is also consistent with measures of income and expenditure inequality in China. For example, a large body of the empirical literature reports that the Gini coefficient for income distribution in China is around 0.4-0.45 (see, e.g., Benjamin, Brandt, and Giles, 2006). There are fewer studies on the distribution of consumption expenditures in China. One exception is the work by Xing, Fan, Luo, and Zhang (2009). Based on data from three villages in southwest China, these authors report a consumption Gini coefficient around 0.33-0.38 and an income Gini coefficient around 0.42. The model-implied consumption Gini is 0.36 and income Gini is 0.42.

Despite the good match between our model and the data, we cannot claim that rapid income growth and inefficient financial system are the only explanatory factors for China's trade surplus and excessive foreign reserves. The quantitative results simply suggest that these two factors are important in understanding China's high saving rate and trade imbalance without resorting to a linked exchange rate and undervalued home currency. Our story does not eliminate or exclude other contributing factors. For example, in addition to precautionary saving motives, a competitive saving behavior due to an imbalanced sex ratio in China could also be partially responsible for the rising household saving rate in China (Wei and Zhang, 2011). The export sector's corporate savings may have also played a role in China's trade surplus and foreign-reserve buildups (Sandri, 2010; and Buera and Shin, 2010). However, the high corporate saving rate in China is mostly absorbed by firms' high investment rate; rather, it is the large pool of household savings that has been unable to find its way to productive firms due to an inefficient state-owned banking system. This explains why China's trade surplus is positively correlated with household saving rate but not with corporate saving rate (see Jin, 2011).[30]

the model and behave similarly in the data.

[30] Appendix B shows that introducing fixed capital investment and a nontradable-goods sector into the model does not change the main results.

5 Explaining the Paradoxical Growth-Saving Causal Relations

The empirical literature documents two puzzling and mutually conflicting facts about household saving behaviors: (i) Lagged income growth is a significant and the single most important factor in explaining the high household saving rate (e.g., Modigliani and Cao, 2004) and (ii) future income growth is often negatively correlated with the current household saving rate (e.g., Kraay, 2000). The first fact is inconsistent with the PIH that predicts not only that a higher income growth should lead to a lower saving rate, but also that households are forward looking. However, the data indicate not only that households save more when their incomes grow faster, but also that they are backward looking. The second fact seems to contradict the first and supports the PIH because it indicates that households reduce current saving when they expect to be richer in the future. The greater puzzle is that the same data set yields these two opposite implications.

Both facts are predicted by our model. Under uninsured risk and borrowing constraints, optimal consumption policy appears myopic, depending only on the current level of a target wealth. However, the target wealth (x_t) depends on past savings (s_t), which in turn depend on the past income growth rate (g_{t-1}). In addition, because consumption (c_t) and the current savings (m_{t+1}) are both proportional to the target wealth x_t, the changes in savings ($m_{t+1} - m_t$) are then determined by both current and lagged income growth instead of future income growth. This implies that movements in the saving rate will lag movements in the income growth rate. Therefore, past income growth is a good predictor of the current saving rate but not vice versa. This also implies that if income growth is serially correlated over the business cycle, then higher future income growth will appear to indicate a lower current saving rate because saving lags income growth.

To quantitatively illustrate the lead-lag relations between income growth and the saving rate, we first generate artificial time series for $\{g_t\}_{t=1}^T$ and $\{\tau_t\}_{t=1}^T$ from the model. The sample size is $T = 10,000$. The growth rate is generated according to the lognormal law of motion in equation (23) and the saving rate is generated based on the saving function in equation (25). The parameters are based on Table 1. Based on the simulated time series, we then estimate the following equation by ordinary least squares (OLS):

$$\tau_t = \gamma_0 + \gamma g_{t-J} + e_{\tau t} \ , \quad J = \{1, 2, -1, -2\} \, . \tag{27}$$

The estimated values of γ are reported in Table 2 (standard error SD is in parentheses).

Table 2. OLS Estimation of Equation (27)

	$\rho_g = 0.17$				$\rho_g = 0.5$			
J	1	2	-1	-2	1	2	-1	-2
γ	0.788	0.133	0.023	-0.000	1.430	0.716	-0.307	-0.145
SD	(0.000)	(0.002)	(0.002)	(0.002)	(0.005)	(0.006)	(0.007)	(0.007)
R^2	0.998	0.030	0.001	0.000	0.461	0.114	0.021	0.005

The left panel of Table 2 shows that lagged income growth (either with a one-period lag or a two-period lag) significantly predicts the current saving rate, consistent with the findings of Modigliani and Cao (2004). The OLS coefficient is 0.788 with an extremely high R^2 of 0.988 when the lag length $J = 1$. In the case of a two-period lag ($J = 2$), the coefficient is substantially reduced to 0.133 but still highly significant with a SD of 0.002. In sharp contrast, future income growth with a one-period lead ($J = -1$) has only a very weak explanatory power on current saving rate (the coefficient is 0.023). Future income growth with a two-period lead ($J = -2$) has zero power in explaining the current saving rate and the coefficient tends to be negative.

This paradoxical pattern of growth-saving relations implied by the model can be further enhanced if we increase the serial correlation of the income growth rate from $\rho_g = 0.17$ to $\rho_g = 0.5$, for example. The right panel in Table 2 shows that when $\rho_g = 0.5$, past income growth remains significant in predicting future saving. More importantly, future income growth rates (either with a one-period lead or a two-period lead) are all negatively (and significantly) associated with the current saving rate, consistent with the empirical findings of Kraay (2000).

Therefore, a negative association between future income growth and current saving does not necessarily imply that the PIH holds. On the other hand, a positive association between past income growth and future saving does not necessarily imply that households are not rational either. Our model suggests that these seemingly paradoxical relationships can arise simultaneously—not because people are irrational but because they face constraints (incomplete financial markets and borrowing limits) not taken into account by the conventional PIH framework.

6 Caveats about Parameter Values (*Not For Publication*)

To capture China's salient features of labor market and also to achieve analytical tractability, we have made two crucial assumptions in the model: (i) quasi-linear utility with predetermined hours and (ii) i.i.d. idiosyncratic shocks. A quasi-linear utility with predetermined supply of hours implies that forward-looking entrepreneurs can adjust their labor supply to target an optimal level of net worth only in advance based on the distribution of idiosyncratic shocks or anticipated changes in income and spending needs. As a result, when the idiosyncratic shocks are i.i.d., all households will choose the same level of targeted net worth (or cash on hand, x_t) in the beginning of each period before idiosyncratic shocks are realized, regardless of their initial asset holdings ($m_{t-1}(i)$). This property is key for yielding closed-form solutions at both the household and the aggregate levels, so the model can be solved by pencil and paper.[31]

[31] Comparing to financial frictions, the assumption of an elastic labor supply (at the aggregate level) is not as important for our main results although it is technically crucial for obtaining closed-form solutions.

If the idiosyncratic shocks are serially correlated, closed-form solutions are no longer possible, yet we expect our results to be quantitatively similar (or even enhanced rather than weakened). In particular, the precautionary saving rate may be even higher with serially correlated shocks, everything else equal. Suppose shocks are highly persistent, a positive shock to spending needs (or a negative shock to bonus income) would signal a long period of high demand for buffer-stock saving relative to income. This would raise the optimal saving rate. The converse is true under a negative preference shock (or a positive income shock), which would reduce the saving rate. However, because the utility function is concave in consumption, agents are thus risk averse. This implies that a persistent decline in consumption in a bad state outweighs a persistent increase in consumption in a good state, especially when people are borrowing constrained in a bad state. This asymmetric implication for welfare would result in a higher steady-state saving rate so that households have an adequate buffer against persistent shocks.

These caveats notwithstanding, it is well known that it is impossible for a model with i.i.d. idiosyncratic shocks to simultaneously match both the income distribution across households and the overtime volatility of income for each household. Cross-sectional variance and cross-time variance are different statistics but they happen to be identical under i.i.d shocks. Typically, the Gini coefficient is about 2 to 3 times larger than the overtime variance of income (or consumption) in developed countries. For example, the U.S. household income Gini is about 0.52 to 0.55 while the variance of household income over time is about 0.23 (DeBacker et al., 2010). Therefore, while our model matches the cross-household dispersions (inequalities) in income and consumption, it may have overstated the true risks facing Chinese households. However, the following discussions explain why this is not the case for China.

i) The Gini coefficient in China reported in the existing empirical literature may be significantly understated for several reasons. China's income Gini is reported to be 0.45 in 2004, which is significantly smaller than that in the United States in 2006 (0.55 for pretax household income and 0.52 for post-tax household income).[32] According to Flannery (2009), China's top 400 richest people are worth $314 billion, which is just one-fourth the total net worth of their American counterparts. On the other hand, the poorest group of people in the U.S. is likely to be 50 times richer than their Chinese counterparts because the U.S. average hourly wage was $19 in 2010 while the hourly wage in southern China cities was only about 75 cents an hour (*New York Times*, June 7, 2010), this is a 25 fold gap. In addition, the income level in urban China is at least 3 times higher than that in rural area. Thus, the reason for possible underestimations of Gini coefficients in China is the lack of sophisticated survey data with sufficiently large household samples that cover both rich and poor Chinese households.[33]

[32] See, DeBacker et al. (2010).

[33] A recent study by Wang (2010) using rarely available government survey data in the countryside of east China

ii) Measuring income risk is far more difficult than measuring income inequality in China because the former requires long time-series data to track family income changes. In particular, the existing short panel data in particular cannot provide accurate information about the true income risk in China. For a country such as China with a well-known volatile history, at least 50 to 75 years of time-series data are needed to correctly gauge the volatility of household income. China has experienced dramatic social and political turbulences and revolutions in the past 100 years and the most recent ones—the Great Famine in 1959-61 and the Cultural Revolution in 1966-76—took place only 40 to 50 years ago. A correct measure of income risk must take these extraordinary events into account because they reflect people's real experiences about income risk and sociopolitical risk in life. Unfortunately, the best-available panel data in China start in the early 1990s (such as the data used by Chamon and Prasad, 2010; and Chamon, Liu, and Prasad, 2010), which cover only the most peaceful time in recent Chinese history, and even these best-available panel data cover only a very small number (a few hundreds) of households in China. Such panel data are thus hardly representative.

Alternatively, income risk can be associated with career opportunities or the probability of changes in one's income class or social status. Enormous economic opportunities in modern China and the lack of legal and institutional protections of property rights mean that (i) ordinary people can become suddenly rich and (ii) rich people can become suddenly poor (due to either business failures or political reasons). For example, the business failure rate in China is about 23 times larger that that in the U.S. In the period 1948-97, the average business failure rate in the U.S. was 0.6% (or 60 firms exit per year per 10,000 firms).[34] In China in the period 1998-2006, the business failure rate was 14% (or 1,400 firms exit per 10,000 firms per year).[35]

iii) Aside from income risk, a perhaps more important source of risk is expenditure uncertainty, such as unexpected spending for housing, education, and health care, or unpredictable expenditures related to accidents, property damages, and volatile fluctuations in consumption goods prices. An ideal proxy of spending risk would be the frequency of illness and the associated costs or accessibility of medical services, but such data are either unavailable or highly inadequate in China. However, the following statistics may provide partial information to gauge expenditure uncertainty or spending risk in China relative to that in the United States.

a) The majority of Chinese people are not effectively covered by any health care insurance system. Medical insurance coverage in rural China has been essentially nonexistent since economic reform (Wang, Xu, and Xu, 2007). Even in urban China, only 39% of the urban population had

estimates that in 2009 the household income Gini is above 0.74 for family business income, 0.6 for wage income, 0.93 for property income, and 0.78 for transfer income. These estimates seem closer to truth than the majority of estimates in the existing literature.

[34] The data are from the Department of Commerce (Census Bureau) and Dun & Bradstreet, Inc.

[35] See Brandt and Zhang (2009).

some limited form of medical insurance in 2003 (Xu, Wang, Collins, and Tang, 2007).

b) Medical costs have increased dramatically in China since the economic reform and privatization of the health care system, far faster than the growth rate of income and general price level. The degree of uncertainty involved in medicare costs is also extremely high.

c) The WHO reports that approximately 2 million people in the world died prematurely each year from water and air pollution, and more than one-third of that figure comes from China (Platt, 2007).

d) The risk related to car accidents in China is 24 to 35 times higher than in the U.S. In 2005, the U.S. had about 247 millions cars on the road and about 42,000 people were killed; the ratio is 0.00017 death per car. In 2009, China had 62 millions cars on the road and 250,000 people were killed; the ratio is 0.004 death per car.[36] Therefore, traffic accident-related risk is 23.5 times higher in China than in the U.S. Even this is an underestimation because only a very small percentage of the Chinese population owns automobiles. If the ownership rate in China were as large as the U.S., the accident rate would have been much higher. For example, China's Guandong province has the highest car ownership in China because it is economically more advanced than other provinces. In 2006, Guandong province had about 1.5 millions cars on the road and about 9,000 people were killed. The death accident rate is 0.06 person per car, which is 35 times higher than in the U.S.[37]

e) The risk of work-related injuries in China is two orders of magnitude higher than in the U.S. For example, according to an official report from the U.S. Department of Labor (Solis and Main, 2011), the average annual incidence rate of fatal injuries in the U.S. mining industry is 0.026% (or 2.6 individuals per 1000 persons for the period 2005-09). According to Xing, Liu, Sun, and Zhang (1996), the comparable incidence rate in China (for the period 1981-94) is about 15%. Alternatively, if the accident rate is measured by the number of fatal injuries (death) per millions of tons of coal output, the value is 0.02% in the U.S. but 4% in China. Therefore, the risk of work-related injuries is 200 to 500 times higher in China than in the U.S.

These data and information regarding income risk and expenditure risk can easily place the true degree of idiosyncratic uncertainty in China at a level a couple of orders of magnitude higher than that in the U.S. A conservative back-of-the-envelope estimate suggests that both income and expenditure risk in China are at least 25 to 30 times higher than in the U.S. In addition, China lacks well-developed insurance markets, social safety nets, and bankruptcy laws, so household and entrepreneurs must bear most of the risks themselves through self-insurance and buffer-stock savings.

[36] Data source: http://www.car-accidents.com.

[37] The official number is 9,000 people killed (Wang, 2009). However, many analysts have repeatedly pointed out that the actual number in China is at least three times larger than official government reports (see, e.g.,). Therefore, the true accident rate in Guangdong province may be more than 100 times higher than in the U.S.

These back-of-the-envelope estimates of uninsured risks suggest that it is unlikely that our model parameters have overstated the true degree of idiosyncratic risks in China. For example, DeBacker et al. (2010) report that the overtime SD of U.S. male earnings (in logarithms) is around 0.48. If we suppose that risk can be transmitted linearly in the model, multiplying this number by 25 suggests that the true volatility of household income in China should be about $\sigma_{\log y} = 25 \times 0.48 = 12$. In general, the variance of household consumption (in logarithms) is about half of the variance of household income (see, e.g., Heathcote, Storesletten, and Violante, 2009), so the implied consumption volatility of Chinese households should be about $\sigma_{\log c} = 6$. Alternatively, using PSID and CES data, Guvenen and Smith (2010) found that the estimated SD of the idiosyncratic transitory component in household log income is $\sigma_{\log y} = 0.3$.[38] Multiplying this number by 25 gives $\sigma_{\log y} = 7.5$. If consumption is half as volatile as income, the implied consumption volatility in China would be $\sigma_{\log c} = 3.75$. On the other hand, based on parameter calibrations in Table 1 and the simulated time-series samples used in Table 2, the implied SD of log consumption in our model is $\sigma_{\log c} = 0.6$.[39] This value is only one-tenth to one-sixth of the back-of-the-envelope estimates of consumption risk in China.

Finally, Appendices B and C (available upon request) and Wen (2011) provide further robustness analyses to our results. Appendix B shows that our results continue to hold when capital investment and a nontradable-goods sector are introduced into the benchmark model. In particular, it shows that similar parameter values can imply a 42% investment rate (investment-to-GDP ratio) and a 46% national saving rate. Thus, China's domestic investment rate is unable to fully absorb its domestic savings despite the fact that it is one of the highest in the world.

Appendix C shows that capital controls and a linked exchange rate are not essential for China's trade surplus and foreign-reserve buildups. Rather, Heckscher-Ohlin-Ricardian comparative advantages in labor costs and productivity growth (i.e., a rapidly growing export income) and precautionary saving due to an inefficient financial system and underdeveloped insurance markets hold the key. This implies that even if China appreciates the RMB dramatically against the dollar (or adopts a flexible exchange rate), its trade with the U.S. and the rest of the world will remain imbalanced, everything else equal.[40] This point is important because large trade surpluses and

[38] These authors assume that the transitory component of household income (in logarithms) follows an AR(1) process, $y_t = \rho y_{t-1} + e_t$. The estimated value of persistence is $\rho = 0.75$ and the estimated SD of the innovation is $\sigma_e = 0.2$. Hence, the SD of y is given by $\sigma_{\log y} = \sqrt{\frac{\sigma_e^2}{1-\rho^2}} = 0.3$.

[39] Because disposable income may take negative values in our model (due to negative bonus shocks), we are not able to compute the variance of log income in the model.

[40] This prediction is consistent with data. The value of the Chinese yuan (RMB) has increased significantly in recent years—more than 21% against the U.S. dollar since 2005. Nevertheless, U.S. trade deficits with China have continued to increase since then. As of 2010, China overtook Japan to become the second-largest trading partner (country) of the U.S., next only to Canada. Similarly, Japan's massive trade surplus with the U.S. remains even after dramatic appreciations of the yen against the dollar in the 1980s under U.S. pressure.

foreign-reserve buildups are not unique to the Chinese economy. Other economies where capital controls and a linked exchange rate are not necessarily universal features of emerging economies. Japan, Hong Kong, Taiwan, Singapore, and South Korea also exhibited high household saving rates and persistent trade surpluses during their respective rapid economic growth periods.[41] Therefore, it is desirable to show that our results do not hinge on the preconditions of capital controls or a linked exchange rate, even though our modeling strategies respect such status quo in China as they impose disciplines on the assumptions—or impose restrictions on the realisticness of the working mechanisms—in the model (see discussions in the Introduction).

7 Relation to the Literature

The heterogeneous productivity structure of the exporting firms in the model is based on Melitz (2003).[42] Incorporating tradable assets and borrowing constraints into a setting with heterogeneous exporters, where forward-looking entrepreneurs make production, consumption, and saving decisions in a dynamic environment with both idiosyncratic and aggregate uncertainties, necessarily increases the technical complexity of the Melitz model. In order to reduce this additional complexity, we abstract from some of the firm-level dynamics, such as entry and exist, fixed export costs, and monopolistic competition, that are typically modeled in the recent trade literature. The primary forces explaining the existence of imbalanced trade as an equilibrium phenomenon are nevertheless quite intuitive. Large uninsured risks and severe borrowing constraints can generate a strong precautionary demand for liquid assets (in both the tradable and nontradable-goods sectors), which leads to current account surplus. Because the marginal propensity to save in such an environment depends positively on income growth, the more firms are able to export, the larger is the trade imbalance.

This paper also relates to the literature on global financial imbalances, most notably works by Caballero, Farhi, and Gourinchas (2008); Mendoza, Quadrini, and Rios-Rull (2009); Ju and Wei (2010); and Song, Storesletten, and Zilibotti (2011). Caballero et al. (2008) attempt to explain the global imbalances in financial asset allocations as an equilibrium outcome of an economic environment in which various regions of the world differ in their capacity to generate financial assets from real investments. They argue that fast growth in emerging economies, coupled with their inability to generate sufficient local store-of-value instruments, would increase their demand for saving instruments from the developed world. This leads to a rise in capital flows toward the United States and an increase in the importance of American assets in global portfolios.

Mendoza et al. (2009) argue that persistent global imbalances and their portfolio composition

[41] Some of these economies later adopted a flexible exchange rate regime, yet their trade surplus remained.
[42] Also see Eaton and Kortum (2002), BEJM (2004), and Ghironi and Melitz (2005).

could be the result of international financial integration among countries with heterogeneous domestic financial markets. In particular, countries with more-advanced financial markets attract financial capital from less-developed countries by providing safe return assets while maintaining positive net holdings of risky equity and FDI.

Ju and Wei (2010) document a very interesting and important stylized fact: Financial capital tends to flow from poor to rich countries, whereas fixed capital tends to flow from developed to developing countries. They study how domestic institutions can explain this pattern of two-way international capital flows. They argue that an inefficient financial system and poor corporate governance may be bypassed by two-way capital flows in which domestic savings leave the country in the form of financial capital outflows but domestic investment occurs through inward FDI from countries with more-efficient financial systems and better corporate governance.

Song, Storesletten, and Zilibotti (2011) provide a model of financial friction and resource reallocation to explain China's growth experience and economic transition, as well as its foreign-reserve buildup. They argue that, because private firms have higher productivity but limited access to external financing while state-owned enterprises (SOEs) have lower productivity but better access to credit markets, resources are only gradually reallocated from inefficient firms to efficient firms during the transition, generating sustained high returns to capital. On the other hand, the downsizing of SOEs during the transition reduces aggregate demand for capital and forces the excess domestic savings to be invested abroad, generating a foreign surplus. While their model is consistent with the rising foreign surplus in China and provides an excellent framework for understanding China's economic transition through resource reallocation, it takes the saving rate in China as given and generates capital outflows through a declining domestic capital demand by SOEs.[43]

Aforementioned papers all emphasize financial frictions in creating the financial imbalance between domestic saving and domestic investment and rely on the accounting identity between capital and current accounts to explain trade surpluses. However, by focusing on the saving-investment imbalance alone, this literature falls short in explaining the large trade surplus in China because capital controls in China make financial-capital outflows impossible, as discussed in the introduction of this paper.[44] In addition, this literature either is unable to explain China's high household saving

[43] Models that can generate domestic savings in excess of domestic capital demand also include Sandri (2008), Buera and Shin (2010), Chamon, Liu, and Presad (2010), and Carroll and Jeanne (2009), among others. Sandri (2008) and Buera and Shin (2010) argue that financial frictions and productivity growth in a developing country can cause a larger increase in saving than in investment because borrowing constraints induce entrepreneurs not only to self-finance investment but also to accumulate precautionary wealth outside their business enterprises. Using an overlapping generations framework of precautionary savings due to increasing income (unemployment) risk, Chamon, Liu, and Presad (2010) and Carroll and Jeanne (2009) show that income growth and an associated acceleration in income (unemployment) risk can lead to a large buildup of domestic savings relative to domestic capital demand, and hence outflows of capital from developing countries to developed ones. In sharp contrast, this paper does not rely on the assumption of increasing risk to explain the positive growth-saving relationship. Income (consumption) risk is assumed to be constant in our model.

[44] Namely, Chinese firms and households cannot purchase foreign assets with RMB, and the only way to obtain large amount of dollars is through exports.

rate and the positive growth-to-saving causality found in the data, or fails to take into account the rapid income growth in China when deriving the saving-investment relations.[45]

Building on and complementing the existing literature, this paper considers several important empirical issues not addressed previously and uses these issues to discipline our modeling strategies. First, we aim to explain the positive relationship between high saving and high growth under low interest rates without relying on the Modigliani-Brumberg (1954) demographic mechanism.[46] We believe that this positive growth-to-saving effect is critical in understanding China's extraordinary path of trade surplus and foreign-reserve buildups in recent decades, especially since joining the WTO. Second, we would like our model to generate financial capital outflows instead of fixed capital outflows, consistent with the empirical facts documented by Ju and Wei (2010). Third, we take into account capital controls in China so that financial capital outflows are the result of imbalanced trade rather than the other way around. In addition, our model deals explicitly with nominal foreign reserves, which requires properly modeling the micro-incentives behind the demand for foreign currency or low-yield liquid foreign assets. This modeling strategy not only enables us to discuss China's foreign reserves in nominal terms but also opens the possibility of addressing other interesting issues related to the nominal exchange rate. Instead of assuming cash-in-advance constraints or money-in-the-utility function as in the standard international finance and monetary literature, this paper combines Bewley's (1980) store-of-value function of money demand with a buffer-stock precautionary saving model to generate demand for money and nominal foreign reserves.

Precautionary saving is a well-acknowledged factor contributing to high saving rates (e.g., Aiyagari, 1994). However, this paper shows that precautionary saving per se is not sufficient for quantitatively explaining the excessively high saving rates of China and other emerging economies despite large uninsured risks. Besides providing an analytically tractable model, a novel contribution of this paper is it shows that rapid income growth can significantly enhance the rate of precautionary saving in an infinite horizon model despite low interest rates and constant idiosyncratic risk.

Finally, this paper's concept in explaining high saving through high income growth in an infinite-horizon model is related to the work of Chen, Imrohoroglu, and Imrohoroglu (2006). These authors use a standard neoclassical growth model to offer a quantitative account of the time path of Japan's high national saving rate in the postwar period. Their analysis reveals that TFP growth is the main force driving Japan's high saving rate in the 1960-70s. However, their model requires a high real interest rate (implied by the high marginal product of capital) to induce households to save. For example, in their model the required real interest rate for households to save would be as high as

[45] High household-income growth may generate low household saving in these models.

[46] The cohort effect on the positive growth-to-saving relationship in a life-cycle model is rejected empirically by many studies; see Hayashi (1986), Kraay (2000), Horioka and Wan (2007), and Chamon and Presad (2010), among others.

14 percent a year when the TFP growth rate is 10% per year and the time discounting factor is 0.96. But a real interest rate this high is rarely observed in Japanese or Chinese data (see an earlier version of this paper (Wen, 2009a) for more discussions on this issue and the relevant data). In contrast, our model can generate a high saving rate in a zero or even negative real interest rate environment, which is consistent with the real-life scenario in China.[47]

8 Conclusion and Remarks for Future Research

This paper provides a simple, analytically tractable framework to understand China's massive trade imbalances and foreign-reserve buildups. Despite its simplicity, the model performs quite well in qualitatively and quantitatively explaining the Chinese data. Our framework is disciplined by two sets of important empirical facts: (i) China has a high income growth rate and a depressed (negative) real interest rate, and (ii) China has capital controls and its large financial capital outflow is the consequence rather than the source of its current account surplus. These facts were the motivation to (i) start with a trade model (that explicitly includes an export sector) and (ii) enrich the model with the demand for low-yield liquid assets in a heterogeneous-agent framework with borrowing constraints and uninsured risks.

Our analysis shows that China's gargantuan foreign reserves can be a natural consequence of rapid economic growth in conjunction with an inefficient financial system (or lack of timely financial reform) that has hampered Chinese entrepreneurs from consuming their rapidly growing future income. Specifically, because of large uninsured risks and severe borrowing constraints, the saving rate remains an increasing function of income growth even at relatively high growth rates. Given the high growth rate of income, Chinese entrepreneurs (exporters) opt to save a substantial fraction (more than a quarter) of their income earned from trade, which leads to the massive buildup of China's foreign reserves. Hence, China's trade imbalance puzzle is resolved without appealing to a hypothetically distorted exchange rate.[48]

The fundamental determinants of the exchange rate include not only excess demand for tradable

[47]See Appendix B for a model with capital investment that shows that a high national saving rate and investment rate can be generated from high TFP growth despite a negative real interest rate.

[48]Even a layman can understand the following arithmetic: Suppose the real exchange rate between Chinese goods and American goods is 1:1—for example, 1 Chinese orange for 1 American apple. Whenever China gives up N oranges, it receives N American apples in return. Trade is therefore always balanced between the two countries because the total value of Chinese exports (N oranges) always equals the value of imports (N apples). Suppose the Chinese government is able to manipulate the real exchange rate so that Chinese workers must give up 100 oranges for 1 American apple. Despite the extremely cheap Chinese products with a real exchange rate of 100:1, trade between the two nations is still balanced—for every 100 oranges China exports to the U.S., it receives 1 U.S. apple, so the total value of Chinese exports ($100 \times N$ oranges) still equals the total value of Chinese imports (N American apples). Therefore, trade can always be balanced regardless of the exchange rate. However, imbalanced trade could occur in the following situation: Suppose Chinese workers gave up 100 oranges and received 1 U.S. dollar as payment, with which they could buy 1 American apple. But instead of spending the entire dollar on American apples, Chinese workers bought only half an American apple and kept the remaining 50¢ as savings. In this case, China would incur a trade surplus of 0.5 American apple, equivalent to lending 50 oranges to American consumers by holding 50¢ as IOUs. The question is why Chinese workers opt to do that? This question is answered in this paper.

goods but also excess demand for tradable assets.[49] Even though the home country in our model runs a massive current account surplus, the supply of dollars in the local exchange market ($P_t^* Y_t$) always equals the total demand of dollars ($P_t^* C_t + M_{t+1} - M_t$) in our model. Hence, there is no pressure for the home currency to appreciate. Therefore, trade imbalances between China and the rest of the world need not be attributed to a linked exchange rate or a undervalued home currency. This conclusion holds true even if entrepreneurs in our model do not want to use dollars as a saving device because they can always opt to exchange the amount ($M_{t+1} - M_t$) in each period with their government for home currency. In this case, the government would become the holder of foreign reserves without affecting our results. Also, the government should have no fear of inflation even without sterilization because households will save, instead of spend, the home currency they exchanged with the government.[50]

Some important issues about China's patterns of trade and reserve buildups remain untouched by this paper. First, this paper focuses on the source of China's foreign reserves but does not explicitly explain the optimal portfolio aspect of the reserves. Currently, the bulk of China's foreign reserves are in U.S. government bonds. Why does the Chinese government not diversify its portfolio better by investing in high-yield assets? It is noteworthy that foreign reserves held by the government are owned by the private sectors since they stem from the private sectors' precautionary savings but are bought (borrowed) by the government through sterilization.[51] Therefore, the government is effectively holding the reserves on behalf of the private sectors and must stand ready whenever the private sectors want to buy back the dollars sold to the government. This implies that the reserves must be predominantly in the form of liquid assets.

Second, China does not necessarily incur a trade surplus with all of its major trading partners. For example, in sharp contrast to the case of Sino-U.S. trade, China has been running large trade deficits with Japan and South Korea in the past decade (see, e.g., Dekle, Eaton, and Kortum, 2007 and 2008). This is puzzling from the viewpoint of the global imbalance literature (e.g., Mendoza et al. 2009) because Japan and Korea both have a far more advanced financial system than China, yet China is not lending goods or having net capital outflows to these major trading partners despite its rapidly growing trading volumes with them. A noteworthy point is that for every dollar

[49] This simple point is often ignored by trade theories and by a large body of the existing literature arguing for the appreciation of the RMB (see, e.g., Evenett, 2010; and Krugman, 2010).

[50] Our analysis suggests that gradual financial development in China will ultimately reduce household saving rate and eliminate China's trade imbalances with the developed world. Recent household data suggest that this mechanism may be already taking effect.

[51] The Chinese government buys dollars from residents by issuing bonds to retrieve the local currency. This practice (sterilization) enables the government to absorb dollars without increasing the supply of local currency when trade surplus increases. Officially, the government is also obligated to buy dollars from the private sector to maintain a fixed exchange rate. Thus, sterilization is equivalent (in outcome) to a situation where the Chinese government meets the savings demands of its domestic residents by selling them Chinese government bonds and using the proceeds to purchase foreign (especially U.S.) bonds. If the private sectors want to increase spending on foreign goods, in principle they can exchange dollars back from the government by selling bonds. In this sense, the Chinese government is functioning like a bank, enabling savers to deposit and invest their foreign income.

Chinese exporters make in trading with the U.S., only one fourth is saved; the remaining 75¢ is not spent entirely on U.S. goods because Chinese entrepreneurs also need dollars to import raw materials and intermediate goods from other countries to produce final consumption goods. Japan and South Korea are the major suppliers of China's intermediate goods and assembly parts. As a result, Japan and South Korea have successfully transformed themselves from America's final-goods suppliers into China's intermediate-goods suppliers through a Heckscher-Ohlin-Ricardian mechanism of comparative advantage: They ship intermediate goods and assembly parts to China and re-export them to the U.S. in the form of final goods. This roundabout export process can significantly reduce labor and raw material costs.[52] Therefore, theories of both international trade (that deal with excess demand of tradable goods) and international finance (that deal with excess demand of tradable assets) must be combined to fully understand the patterns of the imbalanced world trade. Hopefully, extending the simple incomplete-market model developed in this paper to a multi-country general-equilibrium framework with tradable intermediate goods and assets, as well as differentiated labor costs will explain the pattern of trade among the U.S., China, and China's Asian trading partners (such as Japan, Korea, and Taiwan). Explicitly dealing with these issues is beyond the scope of this paper, so they are left to future research.

[52] According to Gaulier, Lemoine, and Ünal-Kesenci (2007), "Assembly and processing of imported inputs for re-export account for about half of China's foreign trade. These activities have been the most dynamic part of China's exports since the early 1990s and have allowed for their rapid diversification from textile to electronics. Assembly and processing is responsible for China's entire trade surplus with the US and Europe."

References

[1] Aiyagari, R., 1994, Uninsured idiosyncratic risk and aggregate saving, Quarterly Journal of Economics 109(3), 659-684.

[2] Benjamin, D., L. Brandt, and J. Giles, 2006, Inequality and Growth in Rural China: Does Higher Inequality Impede Growth? Working Paper IZA DP No. 2344.

[3] Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, 2003, Plants and Productivity in International Trade, American Economic Review 93, 1268-1290.

[4] Bewley, T., 1980, The optimum quantity of money, in Kareken, J., and Wallace, N., eds., Models of Monetary Economics. Minneapolis, MN: Federal Reserve Bank of Minneapolis, 1980, 169-210.

[5] Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, 2009, Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing, NBER Working Paper 15152.

[6] Buera, F. and Y. Shin, 2010, Productivity growth and capital flows: The dynamics of reform, Working Paper, Washington University in St. Louis.

[7] Caballero, Ricardo, Emmanuel Farhi, and Pierre-Olivier Gourinchas, 2008, An equilibrium model of global imbalances and low interest rates, American Economic Review 98(1), 358-393.

[8] Carroll, C and D. Weil, 1994, Saving and Growth: A Reinterpretation, Carnegie-Rochester Conference Series on Public Policy 40, 133-192.

[9] Carroll, C. and O. Jeanne, 2009, A tractable model of precautionary reserves, net foreign assets, or sovereign wealth funds, Peterson Institute for International Economics Working Paper Series WP 09-10.

[10] Chamon, M. and E. Prasad, 2010, Why are saving rates of urban households in China rising? American Economic Journal: Macroeconomics 2(1), 93-130.

[11] Chamon, M., K. Liu, and E. Prasad, 2010, Income Uncertainty and Household Savings in China, NBER and IZA Discussion Paper No. 5331.

[12] Chaney, T., 2008, Distorted gravity: Heterogeneous firms, market structure, and the geography of international trade, American Economic Review 98, 1707-1721.

[13] Chang, Y. and S. Kim, 2006, From Individual to aggregate labor supply: A quantitative analysis based on a heterogeneous agent macroeconomy, International Economic Review 47(1), 1–27.

[14] Chen, K., A. Imrohoroglu, and S. Imrohoroglu, 2006, The Japanese saving rate, American Economic Review 96(5), 1850-1858.

[15] China Statistical Year Book (2008, 2009).

[16] Deardorff, Alan, 2010, A trade theorist's take on global imbalance, in Evenett, Simon J. (ed.), The US-Sino Currency Dispute: New Insights from Economics, Politics and Law. London: VoxEU.Org Publications, Centre for Economic Policy Research, 43-47.

[17] DeBacker, J., B. Heim, V. Panousi, and I. Vidangos, 2010, Rising Inequality: Transitory or Permanent? New Evidence from a Panel of U.S. Tax Returns 1987-2006, Working Paper.

[18] Dekle, Robert, Jonathan Eaton, Samuel Kortum, 2007, Unbalanced trade, American Economic Review: Papers and Proceedings 97, 351-355.

[19] Dekle, Robert, Jonathan Eaton, Samuel Kortum, 2008, Global rebalancing with gravity: Measuring the burden of adjustment, IMF Stuff Papers 55(3), 511-540.

[20] Eaton, Jonathan and Kortum, Samuel, 2002, Technology, Geography, and Trade, Econometrica 70(5), 1741-1780.

[21] Evenett, Simon J. (ed.), 2010, The US-Sino Currency Dispute: New Insights from Economics, Politics and Law. London: VoxEU.Org Publications, Centre for Economic Policy Research.

[22] Flannery, Russell, 2009, China's 400 Richest, Forbes.com, November 4, 2009; http://www.forbes.com/2009/11/04/china-richest-billionaires-buffett-baidu-china-billionaires-09-wealth_intro.html.

[23] Friedman, M., 1957, A Theory of the Consumption Function. Princeton, NJ: Princeton University Press.

[24] Gaulier, Guillaume, Françpise Lemoine, and Deniz Ümal-Kesemci, 2007, China's emergence and the reorganization of trade flows in Asia, China Economic Review, 18(3), 209-243.

[25] Ghironi, F. and M. Melitz, 2005. International Trade and Macroeconomic Dynamics with Heterogeneous Firms, Quarterly Journal of Economics, 120(3), 865-915.

[26] Gottschalk, P. and R. Moffitt, 1994, The growth of earnings instability in the U.S. labor market, Brookings Papers on Economic Activity, 217-272.

[27] Gottschalk, P. and R. Moffitt, 2009, The rising instability of U.S. earnings, Journal of Economic Perspectives 23, 3-24.

[28] Haider, S., 2001, Earnings instability and earnings inequality of males in the United State, 1967-1991, Journal of Labor Economics 19, 799-836.

[29] Hayashi, F., 1986, Why Is Japan's Saving Rate So Apparently High? NBER Macroeconomics Annual 1, 147-210.

[30] Heathcote, J., K. Storesletten, and G. Violante, 2009, Consumption and Labor Supply with Partial Insurance: An Analytical Framework, Federal Reserve Bank of Minneapolis Research Department Staff Report 432.

[31] Horioka, C. Y. and J. Wan, 2007, The Determinants of Household Saving in China: A Dynamic Panel Analysis of Provincial Data, Journal of Money, Credit, and Banking 39(8), 2077–2096.

[32] Ishikawa, T. and K. Ueda, 1984, The bonus payment system and Japanese personal savings, in M. Aoki, ed., The Economic Analysis of the Japanese Firm. Amsterdam: North Holland, 133-192.

[33] Jin, Keyu, 2011, Credit Constraints and Growth in a Global Economy, Working Paper, London School of Economics.

[34] Ju, Jiandong and Shang-Jin Wei, 2010, Domestic institutions and the bypass effect of financial globalization, NBER Working Paper 13148.

[35] Kraay, Aart, 2000, Household Saving in China, World Bank Economic Review 14(3), 545-570.

[36] Krugman, Paul, 2010, Taking on China, New York Times, 15 March 2010.

[37] Lagos, R. and R.Wright, 2005, A Unified Framework for Monetary Theory and Policy Analysis, Journal of Political Economy 113(3), 463-484.

[38] Liu, Yuanli, Keqin Rao, and William C. Hsiao, 2003, Medical Expenditure and Rural Impoverishment in China, Journal of Health, Population, and Nutrition 21(3), 216-222.

[39] Lu, D., 2010, Exceptional exporter performance? Evidence from Chinese manufacturing firms, Working Paper, University of Chicago.

[40] Lucas, R.E. Jr., 1980, Equilibrium in a pure currency economy, Economic Inquiry 18(2), 203-220.

[41] Mark, A. and G. Gopinath, 2007, Emerging market business cycles: The cycle is the trend, Journal of Political Economy 115(1), 69-102.

[42] Melitz, Marc, 2003, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, Econometrica 71, 1695-1725.

[43] Mendoza, Enrique G., Vincenzo Quadrini, and José-Victor Rios-Rull, 2009, Financial integration, financial development, and global imbalances, Journal of Political Economy 117, 371-416.

[44] Modigliani, F., 1970, The life-cycle hypothesis of saving and intercountry differences in the saving ratio, in "Induction, Growth, and Trade: Essays in Honour of Sir Roy Harrod" (Oxford: Clarendon Press) 197-225.

[45] Modigliani, F. and R. Brumberg, 1954, Utility analysis and the consumption function: An attempt at integration, in Post-Keynesian Economics. Kenneth Kurihara, ed.

[46] Modigliani, Franco and Richard H. Brumberg, 1954, Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data, in Kenneth K. Kurihara, ed.: Post-Keynesian Economics. New Brunswick, NJ: Rutgers University Press, pp 388-436.

[47] Modigliani, F. and S. Cao, 2004, The Chinese saving puzzle and the life-cycle hypothesis, Journal of Economic Literature 42(1), 145-170.

[48] Moffitt, R. and P. Gottschhalk, 1995, Trends in the covariance structure of earnings in the US.: 1969-1987, Working Paper.

[49] Moffitt, R. and P. Gottschhalk, 2008, Trends in the transitory variance of male earnings in the U.S.: 1970-2004, Working Paper.

[50] Platt, Kevin Holden, 2007, Chinese Air Pollution Deadliest in World, National Geographic News, July 9, 2007.

[51] Ricardo, D., 1817, On the Principles of Political Economy and Taxation, P. Sraffa, ed., Cambridge, Cambridge University Press, 1951.

[52] Sabelhaus, J. and J. Song, 2009, Earnings volatility across groups and time, Working Paper.

[53] Shin, D. and G. Solon, 2008, Trends in Men's earnings volatility: What does the panel study of income dynamics show? NBER Working Paper 14075.

[54] Sandri, Damiano, 2010, Growth and capital flows with risky entrepreneurship, IMF Working Paper WP/10/36.

[55] Solis, Hilda L. and Joseph A. Main, 2011, Injury Experience in Coal Mining, 2009, U.S. Department of Labor, IR 1348, 2011.

[56] Song, Z., K. Storesletten, and F. Zilibotti, 2011, Growing like China, American Economic Review 101, 202-241.

[57] Song, Z. and D.T. Yang, 2011, Life Cycle Earnings and the Household Saving Puzzle in a Fast-Growing Economy, Working Paper.

[58] Wang, X. and Y. Wen, 2011, Can rising housing prices explain China's high household saving rate? Federal Reserve Bank of St. Louis *Review* 93(2), 67-88.

[59] Wang, Hong-Ming, 2009, Present Situation of Road Traffic Accident in China and Its Characteristics, China Safety Science Journal 19(10), 121-126.

[60] Wang, H., T. Xu and J. Xu, 2007, Factors Contributing to High Costs and Inequality in China's Health Care System, Journal of the American Medical Association 298(16), 1928-1930.

[61] Wang, O., 2011, An empirical analysis of rural household income distribution, (in Chinese), working paper.

[62] Wei, S.-J. and X. Zhang, 2011, The competitive saving motive: Evidence from rising sex ratios and savings rates in China, Journal of Political Economy, 119(3), 551-564.

[63] Wen, Yi, 2009a, Saving and growth under borrowing constraints: Explaining the high saving rate puzzle, Federal Reserve Bank of St. Louis Working Paper 2009-045C.

[64] Wen, Yi, 2009b, Liquidity and welfare in a heterogeneous-agent economy, Federal Reserve Bank of St. Louis Working Paper 2009-019B.

[65] Wen, Yi, 2011, Making sense of China's excessive foreign reserves, Federal Reserve Bank of St. Louis Working Paper 2011-006A.

[66] Xing, J.J., W.D. Liu, X.J. Sun, J.W. Zhang, 1996, Fatigue and accidencts of underground coal miners in China, China Safety Science Journal 6(5), 31-34.

[67] Xu, Ling, Yan Wang, Charles D. Collins, and Shenglan Tang, 2007, Urban health insurance reform and coverage in China using data from National Health Services Surveys in 1998 and 2003, BMC Health Services Research 2007, 7-37.

Appendix A

I. Proof of Proposition 1.

Proof. Denote $x \equiv \frac{1}{1+\pi} m(i) + wN(i)$ as cash in hand net of bonus income, and $\{\lambda_t(i), \pi_t(i)\}$ as the Lagrangian multipliers associated with equations (4) and (5), respectively. The first-order conditions for $\{c_t(i), N_t(i), m_{t+1}(i)\}$ are given, respectively, by

$$\frac{\theta_t(i)}{c_t(i)} = \lambda_t(i) \tag{28}$$

$$a = w_t \int \lambda_t(i) dF(\theta) \tag{29}$$

$$(1+\bar{g})\,\lambda_t(i) = \frac{\beta}{1+\pi} E_t \int \lambda_{t+1}(i) dF(\theta) + \pi_t(i) = \frac{\beta}{(1+\pi)} E_t \frac{a}{w'} + \pi(i), \tag{30}$$

where equation (29) reflects the fact that the labor supply must be determined before the realization of $\theta_t(i)$ and $v_t(i)$ in each period. The optimal decision rules are characterized by a cutoff strategy. There exist two cutoffs $\{\theta_1^*, \theta_2^*\}$ for preference shocks because income shock $v_t(i)$ takes two possible values. In anticipation that the cutoffs are independent of i, consider two possible cases as follow:

Case A. $m'(i) \geq 0$, $\pi(i) = 0$. In this case, the urge to consume is low, or the income shock is high, or both. It is then optimal to save to prevent possible borrowing constraints in the future when the urge to consume may be high or income may be low. Equation (30) then implies that the shadow value of good

$$\lambda(i) = \frac{\beta}{(1+\bar{g})\,(1+\pi)} E_t \frac{a}{w_{t+1}} \equiv \bar{\lambda}_t. \tag{31}$$

Equation (28) then implies that $c_t(i) = \theta(i)\bar{\lambda}^{-1}$. The household budget constraint then implies $(1+\bar{g})\, m'(i) = x(i) + v(i)w - \theta(i)\bar{\lambda}_t^{-1}$. The requirement $m'(i) \geq 0$ then implies

$$\theta(i) \leq [x(i) + v(i)w]\,\bar{\lambda}_t \equiv \theta^*, \tag{32}$$

which defines the cutoff θ_t^*. However, there are two sub-cases to consider.

Case A1: $v_t(i) = -\xi$. In this sub-case the cutoff is given by

$$\theta_1^* \equiv [x(i) - \xi w]\,\bar{\lambda}_t, \tag{33}$$

and this sub-case happens with probability $1-p$.

Case A2: $v_t(i) = \xi$. In this sub-case the cutoff is given by

$$\theta_2^* \equiv [x(i) + \xi w]\,\bar{\lambda}_t \tag{34}$$

and this sub-case happens with probability p. Because x_t is determined before the realization of $v_t(i)$, it is hence independent of $v_t(i)$. Equations (33) and (34) then together imply

$$\theta_{2t}^* = \theta_{1t}^* + 2\xi w_t \bar{\lambda}_t > \theta_1^*. \tag{35}$$

This implies that the agent adopts a lower cutoff for preference shocks when income is low and a higher cutoff when income is high.

Case B. $\pi(i) > 0$ and $m'(i) = 0$. In this case, the urge to consume is high, or income is low, or both. It is then optimal *not* to save. By the household budget constraint, we have $c(i) = x(i) + v(i)w$. There are again two sub-cases.

Case B1: $v_t(i) = -\xi$, equation (33) then implies $c(i) = \theta_1^* \bar{\lambda}_t^{-1}$.

Case B2: $v(i) = \xi$, equation (34) then implies $c(i) = \theta_2^* \bar{\lambda}_t^{-1}$. Equation (28) then implies that

$$\lambda_t(i) = \begin{cases} \frac{\theta(i)}{\theta_1^*}\bar{\lambda}_t & \text{if } v(i) = -\xi \\ \frac{\theta(i)}{\theta_2^*}\bar{\lambda}_t. & \text{if } v(i) = \xi \end{cases}. \tag{36}$$

Clearly, if $\theta(i) > \theta_1^*$, equation (30) implies $\pi(i) = (1+g)\,\bar{\lambda}\left[\frac{\theta(i)}{\theta_1^*} - 1\right] > 0$; and if $\theta(i) > \theta_2^*$, equation (30) implies $\pi(i) = (1+g)\,\bar{\lambda}\left[\frac{\theta(i)}{\theta_2^*} - 1\right] > 0$. In either case, $m'(i) = 0$.

The above analyses imply that $\lambda_t(i)$ takes three possible values as shown in equations (31) and (36), depending on the joint realizations of $\theta_t(i)$ and $v_t(i)$ in period t. Hence, the expected value, $\int \lambda_t(i)dF(\theta)$, can be expressed analytically. That is, equation (29) implies

$$\frac{1}{w} = \left[\frac{\beta}{(1+\bar{g})\,(1+\pi)} E\frac{1}{w'}\right] R(\theta^*), \tag{37}$$

where $R(\theta^*) \equiv (1-p)\left[\int_{\theta\le\theta_1^*} dF(\theta) + \int_{\theta<\theta_1^*} \frac{\theta}{\theta_1^*} dF(\theta^*)\right] + p\left[\int_{\theta\le\theta_2^*} dF(\theta) + \int_{\theta<\theta_2^*} \frac{\theta}{\theta_2^*} dF(\theta^*)\right]$. As a result, the optimal cutoffs, $\{\theta_{1t}^*, \theta_{2t}^*\}$, are determined jointly by equations (37) and (35). Equation (37) implies that the cutoffs are independent of i because $\theta(i)$ and $v(i)$ are both i.i.d. Hence, equations (33) and (34) in turn imply that the optimal cash in hand (x_t) is also independent of i and can be written either as $x_t = \bar{\lambda}_t^{-1}\theta_{1t}^* + \xi w_t$ or as $x_t = \bar{\lambda}_t^{-1}\theta_{2t}^* - \xi w_t$. Given the values of $\lambda(i)$, the decision rules of consumption are given by $\frac{\theta(i)}{\lambda(i)}$ and the decision rules of saving are given by $(1+g)\,m'(i) = x - c(i)$. ■

II. Proof of Proposition 2.

Proof. In the steady state, the aggregate variables satisfy $m' = m, x - \xi w = \theta_1^* wR$, and $x + \xi w = \theta_2^* wR$. So equation (16) can be rewritten as

$$(1+g)\, m = \left[(1-p)\, \theta_1^* H_1 + p\theta_2^* H_2\right] wR. \tag{38}$$

Since $wN = x - \frac{1}{1+\pi} m' = \theta_1^* wR + \xi w$, the disposable income can be written as

$$\wp = \theta_1^* wR + 2p\xi w - \frac{1}{(1+g)\,(1+\pi)} \left[(1-p)\, \theta_1^* H_1 + p\theta_2^* H_2\right] wR. \tag{39}$$

Hence, the saving rate is given by

$$\tau \equiv \frac{gm}{\wp} = \frac{g\left[(1-p)\, \theta_1^* H_1 + p\theta_2^* H_2\right]}{(1+g)\left(\theta_1^* + 2p\xi R^{-1}\right) - \frac{1}{(1+\pi)}\left[(1-p)\, \theta_1^* H_1 + p\theta_2^* H_2\right]}. \tag{40}$$

Since $2\xi R^{-1} = \theta_2^* - \theta_1^*$, the saving rate can be rewritten as equation (19). ■

III. Proof of Proposition 3.

Proof. Denoting $\bar{H}(g) \equiv \left[(1-p)\, \theta_1^* H_1 + p\theta_2^* H_2\right]$ and $G(g) \equiv \left(\theta_1^* + 2p\xi R^{-1}\right) = (1-p)\, \theta_1^* + p\theta_2^*$. Differentiating the saving rate with respect to $\bar{g}$ gives

$$\begin{aligned}
\frac{d\tau}{d\bar{g}} &= \frac{\left(\bar{H} + g\frac{\partial \bar{H}}{\partial g}\right)\left[(1+\bar{g})\, G - \frac{1}{1+\pi}\bar{H}\right] - g\bar{H}\left[G + (1+g)\,\frac{\partial G}{\partial g} - \frac{1}{1+\pi}\frac{\partial \bar{H}}{\partial g}\right]}{\left[(1+\bar{g})\, G - \frac{1}{1+\pi}\bar{H}\right]^2} \\
&= \frac{(1+\bar{g})\, G\left(\bar{H} + g\frac{\partial \bar{H}}{\partial g}\right) - \left(G + (1+g)\,\frac{\partial G}{\partial g}\right) g\bar{H} + \frac{1}{1+\pi} g\bar{H}\frac{\partial \bar{H}}{\partial g} - \frac{1}{1+\pi}\bar{H}\left(\bar{H} + g\frac{\partial \bar{H}}{\partial g}\right)}{\left[(1+\bar{g})\, G - \frac{1}{1+\pi}\bar{H}\right]^2} \\
&= \frac{(1+\bar{g})\, g\left[G\frac{\partial \bar{H}}{\partial g} - \frac{\partial G}{\partial g}\bar{H}\right] + (1+g)\, G\bar{H} - gG\bar{H} + \frac{1}{1+\pi}\bar{H}\left[g\frac{\partial \bar{H}}{\partial g} - \bar{H} - g\frac{\partial \bar{H}}{\partial g}\right]}{\left[(1+\bar{g})\, G - \frac{1}{1+\pi}\bar{H}\right]^2} \\
&= \frac{(1+\bar{g})\, g\left[G\frac{\partial \bar{H}}{\partial g} - \frac{\partial G}{\partial g}\bar{H}\right] + \bar{H}\left(G - \frac{1}{1+\pi}\bar{H}\right)}{\left[(1+\bar{g})\, G - \frac{1}{1+\pi}\bar{H}\right]^2}.
\end{aligned} \tag{41}$$

Notice that $G = (1-p)\, \theta_1^* + p\theta_2^* > \bar{H} = \left[(1-p)\, \theta_1^* H_1 + p\theta_2^* H_2\right]$ because the marginal propensity to save $H_j < 1$. Thus the second term in the numerator $\bar{H}\left(G - \frac{1}{1+\pi}\bar{H}\right) > 0$. However, the first term in the numerator, $\left[G\frac{\partial \bar{H}}{\partial g} - \frac{\partial G}{\partial g}\bar{H}\right]$, is negative because $\theta_j^* H_j$ decreases more than H_j when g

increases. This implies that $\frac{d\tau}{d\bar{g}} > 0$ if $\bar{g} = 0$ and $\frac{d\tau}{d\bar{g}} < 0$ if $\bar{g} \to \infty$. Continuity implies there exists a threshold $\bar{g}^* \in (0, \infty)$ that renders $\frac{d\tau}{d\bar{g}} = 0$ when $\bar{g} = \bar{g}^*$. So it must be true that $\frac{d\tau}{d\bar{g}} > 0$ if $\bar{g} < \bar{g}^*$ and $\frac{d\tau}{d\bar{g}} < 0$ if $\bar{g} > \bar{g}^*$. ■

IV. Proof of Proposition 4.

Proof. When $\xi = 0$, there exists a unique cutoff θ^*. So equations (11) and (12) imply that the optimal cash in hand is given by $\theta^* w R(\theta^*) = x \equiv m(i) + wN(i)$. Using the saving decision rule in equation (10), the condition that $N(i) > 0$ then implies $(1+\bar{g})\,\theta^* w R(\theta^*) > (1+\bar{g})\, m(i) = wR(\theta^*) \max\{\theta^* - \theta(i), 0\}$. Since the maximum value of $\max\{\theta^* - \theta(i), 0\}$ is $\theta^* - 1$ (recall the support of $\theta(i) \in [1, \infty]$), the above inequality is satisfied if

$$(1+\bar{g})\,\theta^* R(\theta^*) > R(\theta^*)\,[\theta^* - 1]\,, \tag{42}$$

which is clearly true for any parameter values in the model. Hence, regardless of the realization of preference shocks, the labor supply of each entrepreneur is strictly positive. On the other hand, the condition that $N(i) < \bar{N}$ implies $(1+\bar{g})\,\theta^* w R(\theta^*) < \max\{\theta^* - \theta(i), 0\} + w\bar{N}$. This inequality can be ensured if the function $\max\{\theta^* - \theta(i), 0\}$ takes the smallest value, which is 0. Hence, the above inequality is satisfied as long as

$$(1+\bar{g})\,\theta^* R(\theta^*) < \bar{N}. \tag{43}$$

Based on equation (12), the above inequality is equivalent to the condition $\frac{(1+\bar{g})^2}{\beta}\theta^* < \bar{N}$. Under Pareto distribution, we have $R(\theta^*) = 1 + \frac{1}{\sigma-1}\theta^{*-\sigma}$, so the cutoff can be solved in closed-form as $\theta^* = \left[(\sigma-1)\left(\frac{1+\bar{g}}{\beta} - 1\right)\right]^{-\frac{1}{\sigma}}$ and the above inequality becomes

$$\frac{(1+\bar{g})^2}{\beta}\left[(\sigma-1)\left(\frac{1+\bar{g}}{\beta} - 1\right)\right]^{-\frac{1}{\sigma}} < \bar{N}. \tag{44}$$

Given that $\sigma > 1$, the LHS of the above inequality is increasing in $\bar{g}$. Since there exists a maximum value of the growth rate $\bar{g}_{\max} = \frac{\beta\sigma}{\sigma-1} - 1$ such that if $\bar{g} \geq \bar{g}_{\max}$, we have $\theta^* = 1$ and $m(i) = 0$ for all i, and in this case $wN(i) = c(i)$ and $N(i) = a$. Therefore, hours worked by each entrepreneur is strictly bounded above by a fixed number $\bar{N} = \frac{(1+\bar{g}_{\max})^2}{\beta} = \beta\left(\frac{\sigma}{\sigma-1}\right)^2$ for $\bar{g} < \bar{g}_{\max}$ or by the fixed number $\bar{N} = a$ for $\bar{g} \geq \bar{g}_{\max}$. ■

V. Proof of Proposition 5.

Proof. When technology A_t is stochastic and follows the law of motion in equation (23), the general equilibrium path of the rescaled benchmark model can be characterized by the set of variables, $\{c_t, m_{t+1}, \theta_t^*, x_t, N_t,\}$, which can be solved uniquely by the following system of equations:

$$c_t = (1-p)\, D(\theta_{1t}^*)\, (x_t - \xi w_t) + pD(\theta_{2t}^*)\, (x_t + \xi w_t) \tag{45}$$

$$(1+g_t)\, m_{t+1} = (1-p)\, H(\theta_{1t}^*)\, (x_t - \xi w_t) + pH(\theta_{2t}^*)\, (x_t + \xi w_t) \tag{46}$$

$$\frac{1+g_t}{w_t} = \beta E_t \frac{1}{w_{t+1}} R(\theta_{1t}^*) \tag{47}$$

$$x_t \equiv m_t + w_t N_t + (2p-1)\, \xi w_t = \theta_{1t}^* R(\theta_{1t}^*) w_t + \xi w_t, \tag{48}$$

plus equation (24) and standard transversality conditions.

In general equilibrium, $w_t = 1+g_t$. So equation (47) becomes $1 = \beta R(\theta_{1t}^*) E_t \frac{1}{1+g_{t+1}}$, which solves implicitly for the cutoff θ_t^* and the cutoff $\theta_{2t}^* = \theta_{1t}^* + 2\xi R^{-1}(\theta_{1t}^*)$. Given the cutoffs, equations (45) and (48) solve for the optimal consumption path as

$$c_t = [(1-p)\, D(\theta_{1t}^*)\theta_{1t}^* R(\theta_{1t}^*) + pD(\theta_{2t}^*)\theta_{2t}^* R(\theta_{1t}^*)]\, (1+g_t)\,, \tag{49}$$

and equations (46) and (48) solve for the optimal foreign reserves as

$$(1+g_t)\, m_{t+1} = [(1-p)\, H(\theta_{1t}^*)\theta_{1t}^* R(\theta_{1t}^*) + pH(\theta_{2t}^*)\theta_{2t}^* R(\theta_{1t}^*)]\, (1+g_t)\,. \tag{50}$$

When $p = \frac{1}{2}$, equation (48) implies $w_t N_t = \theta_{1t}^* R(\theta_{1t}^*) w_t + \xi w_t - m_t$, which in turn implies employment in the tradable sector:

$$N_t = \theta_{1t}^* R(\theta_{1t}^*) + \xi - \frac{m_t}{(1+g_t)} = \theta_{1t}^* R(\theta_{1t}^*) + \xi - \frac{1}{2} \frac{H(\theta_{1t-1}^*)\theta_{1t-1}^* R(\theta_{1t-1}^*) + H(\theta_{2t-1}^*)\theta_{2t-1}^* R(\theta_{2t-1}^*)}{(1+g_t)}. \tag{51}$$

Therefore, year-to-year changes in foreign reserves in period t are given by

$$\begin{aligned} (1+g_t)\, m_{t+1} - m_t &= y_t \left[1 - \frac{c_t}{(1+g_t)\, N_t}\right] \qquad (52) \\ &= y_t \left[1 - \frac{\frac{1}{2}\left[(1-p)\, D(\theta_{1t}^*)\theta_{1t}^* R(\theta_{1t}^*) + pD(\theta_{2t}^*)\theta_{2t}^* R(\theta_{1t}^*)\right]}{\theta_{1t}^* R(\theta_{1t}^*) + \xi - \frac{1}{2(1+g_t)}\left[H(\theta_{1t-1}^*)\theta_{1t-1}^* R(\theta_{1t-1}^*) + H(\theta_{2t-1}^*)\theta_{2t-1}^* R(\theta_{2t-1}^*)\right]}\right]. \end{aligned}$$

Multiplying the lowercase variables on both sides of the above equation by A_{t-1} gives equation (25).

Using the law of motion in equation (23) and assuming lognormal distribution for ε_t with mean zero and variance σ_ε^2, equation (47) implies

$$1 = \beta R(\theta_{1t}^*) E_t \frac{1}{1+g_{t+1}} = \beta R(\theta_{1t}^*) \left(\frac{1+\bar{g}}{1+g_t} \right)^{\rho_g} \frac{e^{\frac{1}{2}\sigma_\varepsilon^2}}{1+\bar{g}}. \tag{53}$$

With θ following the Pareto distribution $F = 1 - \theta^{-\sigma}$, we have $R(\theta_{1t}^*) = (1-p)\left[1 + \frac{1}{\sigma-1}\theta_{1t}^{*-\sigma}\right] + p\left[1 + \frac{1}{\sigma-1}\theta_{2t}^{*-\sigma}\right]$, so the above equation solves for the cutoffs $\{\theta_{1t}^*, \theta_{2t}^*\}$ in conjunction with equation (13). ■

Appendix B (Not for Publication)

The benchmark model does not have capital. Therefore, that model cannot directly answer the broader question of why China's high investment rate is unable to completely absorb its domestic savings. Also, the benchmark model does not have nontradable goods. This appendix extends the benchmark model to a more general setting with capital and a nontradable-goods sector.

There are two production sectors in the home country—a domestic sector (sector 1) that produces nontradable goods and an export sector (sector 2) that produces tradable goods. Because of capital controls and an inconvertible home currency, residents of the home country cannot use foreign currency earned from the export sector to purchase domestic goods and assets, nor can they use income earned from the nontradable-goods sector to buy foreign goods and assets. In other words, nontradable goods are purchased by home currency (RMB) and tradable goods are purchased by foreign currency (USD). However, households in country H can bypass the capital controls through working in both nontradable and tradable sectors. Therefore, despite the capital control, households are able to adjust their baskets of consumption goods for tradable and nontradable goods by choosing an optimal mixture of hours worked in each sector.

Also because of capital controls, firms in the domestic sector must use income earned from domestic sales to rent capital from a domestic rental market, while firms in the export sector can use foreign income to rent capital from an international market with a constant world real interest rate $\bar{r}^w$. Fixed capital is not mobile across sectors but labor is. This setup of segregated capital markets not only captures the reality in China (such as large capital adjustment costs, irreversible investment, the lack of a resale capital goods market, and so on), but also allows the model to generate the Ballasa-Samuelson effect of technology growth on the real exchange rate even though the rate of productivity growth in both the nontradable-goods and tradable-goods sectors is the same (see Wen, 2009a, for more discussions on the determination of the exchange rate and the Ballasa-Samuelson effect).

Technology. Sector j $(j = 1, 2)$ has the following production technology:

$$Y_{jt} = K_{jt}^{\alpha} \left(A_t N_{jt}\right)^{1-\alpha}, \tag{54}$$

where A_t denotes a country's aggregate technology level with a stochastic growth rate specified in equation (23). The demands for capital and labor in sector j are given by

$$r_{jt} + \delta = \alpha \frac{Y_{jt}}{K_{jt}} = \alpha \left(\frac{N_{jt}}{K_{jt}}\right)^{1-\alpha} A_t^{1-\alpha} \tag{55}$$

$$W_{jt} = (1-\alpha) \frac{Y_{jt}}{N_{jt}} = (1-\alpha) \left(\frac{K_{jt}}{N_{jt}}\right)^{\alpha} A_t^{1-\alpha}, \tag{56}$$

where

$$r_{2t} = \bar{r}^w \tag{57}$$

is a constant world interest rate in the international rental market.

Labor mobility across sectors implies $\eta_{1t}W_{1t} = \eta_{2t}W_{2t}$, where η_{jt} denotes the marginal utility of income received from sector j. Hence, the real price of tradable goods in terms of nontradable goods—the real exchange rate in the economy—is given by $e_t \equiv \frac{\eta_{2t}}{\eta_{1t}} = \frac{W_{1t}}{W_{2t}}$. Equations (55) and (56) imply

$$e_t = \left(\frac{\bar{r}^w + \delta}{r_{1t} + \delta} \right)^{\frac{\alpha}{1-\alpha}}. \tag{58}$$

Hence, the real exchange rate is influenced by aggregate technology shocks through changes in the domestic real interest rate. In particular, a higher productivity growth leads to a higher domestic interest rate, which implies that nontradable goods become more expensive relative to tradable goods, so the real exchange rate appreciates (e_t decreases). This captures the Ballasa-Samuelson effect in an environment with identical productivity growth across tradable and nontradable-goods sectors.

Along a balanced growth path, the real wages $\{W_{1t}, W_{2t}\}$ and outputs $\{Y_{1t}, Y_{2t}\}$ in both sectors all grow at the rate of long-run productivity growth ($\bar{g}$) in the absence of aggregate uncertainty ($\varepsilon_t = 0$), while hours worked in both sectors are constant over time. To facilitate the analysis of a stochastic equilibrium path under aggregate uncertainty, we rescale all variables in the model by the level of technology A_{t-1} except for hours worked. Using lowercase letters to denote the transformed variables, $z_{jt} \equiv \frac{Z_{jt}}{A_{t-1}}$, the production functions become

$$y_{jt} = (1 + g_t)^{1-\alpha} k_{jt}^{\alpha} N_{jt}^{1-\alpha} \tag{59}$$

and the real wages become

$$w_{jt} = (1 - \alpha) \frac{y_{jt}}{N_{jt}} = (1 - \alpha) \left(\frac{k_{jt}}{N_{jt}} \right)^{\alpha} (1 + g_t)^{1-\alpha}. \tag{60}$$

Households. As in the benchmark model, there is a continuum of households in country H indexed by $i \in [0, 1]$. Each household has two members (husband and wife); one works in the nontradable-goods sector and the other works in the tradable-goods (exporting) sector. Each household consumes two types of goods: nontradable goods produced at home and foreign goods produced abroad.

Households put their domestic savings in banks and earn a real gross rate of return $1 + r_t^a$. Households exchange their foreign savings (in dollars) with the government for government bonds (sterilization). For simplicity and without loss of generality, assume that government bonds yield zero interest rate. Therefore, in the model it does not matter whether the households hold dollars or local government bonds. As documented by Wen (2009a), financial repression in China leads to a low and even negative real deposit rate for household savings. Despite this, however, the bulk of household wealth is kept in the form of bank deposits because of underdeveloped financial markets in China. On the other hand, firms must borrow funds from monopolistic state-owned banks at market interest rate r_t. To capture this reality, we assume that the real rate of return to household savings is zero ($r^a = 0$), and state-owned banks earn monopoly profits $(r_t - r^a)\, s_t$, which are returned in a lump sum to households. We will show that households in both sectors still save excessively despite the low real deposit rate, as in the benchmark model.

Let $s_t(i) \equiv \frac{S_t(i)}{A_{t-1}}$ denote the rescaled home asset and $m_t(i) \equiv \frac{M_t(i)}{A_{t-1}P_t^*}$ the rescaled real money balances for foreign currency held by household i. Household i's consumption for home goods is denoted as $c_{1t}(i) \equiv \frac{C_{1t}(i)}{A_{t-1}}$, imported goods as $c_{2t}(i) \equiv \frac{C_{2t}(i)}{A_{t-1}}$, and hours worked in sector j as $N_{jt}(i)$. For simplicity, assume $P_t^* = P_{t-1}^*$, so the inflation rate is zero. Household i's problem is to solve

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \theta_t(i) \left[\gamma_1 \log c_{1t}(i) + \gamma_2 \log c_{2t}(i) \right] - N_{1t}(i) - N_{2t}(i) \right\} \tag{61}$$

subject to

$$c_{1t}(i) + (1 + g_t)\, s_{t+1}(i) \leq x_{1t}(i) + v_t(i) w_t \tag{62}$$

$$s_{t+1}(i) \geq 0 \tag{63}$$

$$c_{2t}(i) + (1 + g_t)\, m_{t+1}(i) \leq x_{2t}(i) + v_t(i) w_t \tag{64}$$

$$m_{t+1}(i) \geq 0, \tag{65}$$

where $v_t(i)$ denotes idiosyncratic income shocks, $x_{1t}(i)$ is net wealth (cash in hand) in terms of income earned in sector 1:

$$x_{1t}(i) \equiv w_{1t} N_{1t}(i) + s_t(i) + \Pi_t, \tag{66}$$

and $x_{2t}(i)$ is cash in hand in sector 2:

$$x_{2t}(i) \equiv w_{2t} N_{2t}(i) + m_t(i), \tag{67}$$

where $\Pi_t = r_{1t} \int s_t(i) di$ denotes average profit income distributed from domestic banks. The parameter γ_j in the preference controls the relative equilibrium size of the domestic and export sectors.

Equation (62) is the budget constraint pertaining to domestic income, which state that total real wage income earned from the nontradable-goods sector can be used to finance consumption of nontradable goods (c_{1t}) and the accumulation of home assets ($(1+g)\,s' - s$) subject to the borrowing constraint (63). Analogously, equation (64) denotes the budget constraints pertaining to foreign income, which states that total real wage income earned from working in the tradable-goods sector can be used to finance purchases of foreign produced goods (c_{2t}) and the accumulation of foreign currency (real foreign reserves, $(1+g)\,m' - m_t$) subject to the borrowing constraint (65). For simplicity, we have assumed that income shocks ($v_t(i)$) are symmetric across the two sectors with distribution given by equation (6).

Proposition 6 *Denoting $s_{1t} \equiv s_t$ and $s_{2t} \equiv m_t$, the decision rules of consumption, savings, and cash in hand for household i are given by*

$$c_{jt}(i) = \begin{cases} \min\left\{\frac{\theta(i)}{\theta^*_{jl}}, 1\right\}[x_{jt} - \xi w_{jt}], & \textit{with probability } 1-p \\ \min\left\{\frac{\theta(i)}{\theta^*_{jh}}, 1\right\}[x_{jt} + \xi w_{jt}], & \textit{with probability } p \end{cases} \tag{68}$$

$$(1+g_t)\,s_{jt+1}(i) = \begin{cases} \max\left\{\frac{\theta^*_{jl}-\theta(i)}{\theta^*_{jl}}, 0\right\}[x_{jt} - \xi w_{jt}], & \textit{with probability } 1-p \\ \max\left\{\frac{\theta^*_{jh}-\theta(i)}{\theta^*_{jh}}, 0\right\}[x_{jt} + \xi w_{jt}], & \textit{with probability } p \end{cases} \tag{69}$$

$$x_{jt} = \gamma_j w_{jt} \theta^*_{jlt} R_t + \xi w_{jt}, \tag{70}$$

where the cutoff variables θ^*_{jlt} and θ^*_{jht} are determined by the following equations:

$$\frac{1+g_t}{w_{jt}} = \beta R(\theta_{jlt}) E_t \frac{1}{w_{jt+1}} \tag{71}$$

$$\theta^*_{jht} = \theta^*_{jlt} + 2\xi R^{-1}_{jt}, \tag{72}$$

where the liquidity premium function $R(\cdot)$ is given by $R_{jt} \equiv (1-p)\left[\int_{\theta<\theta^*_j} dF(\theta) + \int_{\theta\geq\theta^*_{jl}} \frac{\theta}{\theta^*_{jl}} dF(\theta)\right] + p\left[\int_{\theta<\theta^*_{jh}} dF(\theta) + \int_{\theta\geq\theta^*_{jh}} \frac{\theta}{\theta^*_{jh}} dF(\theta)\right]$.

Proof. Similar to the proof of Proposition 1. ■

These decision rules are similar to those in the benchmark model. However, the optimal cutoffs in each sector—determined by equations (71) and (72)—may differ across sector j because the real wage may differ across sectors (because of different capital markets and real interest rates).

Denoting any aggregate variable $z_{jt} \equiv \int z_{jt}(i)di$, market clearing for the domestic capital market implies $\int s_{1t}(i)di = k_{1t}$. Hence, the general equilibrium path of the model can be characterized by the sequences of 16 variables, $\left\{c_{jt}, s_{jt+1}, \theta^*_{jlt}, \theta^*_{jht}, x_{jt}, w_{jt}, y_{jt}, N_{jt}; j = 1, 2\right\}_{t=0}^{\infty}$, which can be solved by the following system of 16 equations (assuming $p = 0.5$ so that the average bonus income in each sector is zero, $\bar{v} = (2p - 1)\,\xi = 0$):

$$c_{jt} = (1 - p)\, D(\theta^*_{jlt})\, [x_{jt} - \xi w_{jt}] + pD(\theta^*_{jht})\, [x_{jt} + \xi w_{jt}] \tag{73}$$

$$(1 + g_t)\, s_{jt+1} = (1 - p)\, H(\theta^*_{jlt})\, [x_{jt} - \xi w_{jt}] + pH(\theta^*_{jht})\, [x_{jt} + \xi w_{jt}] \tag{74}$$

$$\frac{1 + g_t}{w_{jt}} = \beta E_t \frac{1}{w_{jt+1}} R(\theta^*_{jlt}) \tag{75}$$

$$\theta^*_{jht} = \theta^*_{jlt} + 2\xi R^{-1}_{jt} \tag{76}$$

$$x_{jt} = \gamma_j \theta^*_{jlt} R(\theta^*_{jlt}) w_{jt} + \xi w_{jt} \tag{77}$$

$$c_{1t} + (1 + g_t)\, k_{1t+1} - (1 - \delta)\, k_{1t} = y_{1t} \tag{78}$$

$$c_{2t} + (1 + g_t)\, m_{t+1} - m_t + (\bar{r}^w + \delta)\, k_{2t} = y_{2t} \tag{79}$$

where $\bar{r}^w + \delta = \alpha \frac{y_{2t}}{k_{2t}}$, plus the four equations (59) and (60) and standard transversality conditions. The profit income from financial intermediaries is given by

$$\Pi_t = (r_{1t} - r^a_t) \int s_{1t}(i) di = r_{1t} k_{2t}. \tag{80}$$

Hence, equation (78) is also the goods market-clearing condition for the nontradable-goods sector. Equation (79) is the household's budget constraint in the export sector, where $(\bar{r}^w + \delta)\, k_{2t}$ is rental payment for capital services. Thus, income from exports is used to finance imports of consumption goods (c_2), capital rental costs ($(\bar{r}^w + \delta)\, k_{2t}$), and foreign-reserve accumulation ($m_{t+1} - m_t$).

The model has a unique steady state. The eigenvalue method easily confirms that the steady state is a saddle, so the general equilibrium path implied by the above system of dynamic equations is unique near the steady state. To see the distortion effect of capital control, suppose $\xi = 0$, so that $\theta^*_{jlt} = \theta^*_{jht}$ in sector $j = 1, 2$. Equations (73) and (77) then imply

$$\frac{c_{1t}}{c_{2t}} = \frac{D\,(\theta^*_{1t})\, \theta^*_{1t} R(\theta^*_{1t})}{D\,(\theta^*_{2t})\, \theta^*_{2t} R(\theta^*_{2t})} \frac{\gamma_1 w_{1t}}{\gamma_2 w_{2t}} \equiv \varphi_t \frac{\gamma_1 w_{1t}}{\gamma_2 w_{2t}}, \tag{81}$$

where the coefficient $\varphi \neq 1$ measures efficiency loss (or deadweight loss) due to capital controls. The allocation would be efficient if $\varphi = 1$.

Defining $\wp_j$ as the total real disposable income in sector j, which includes wage income plus real capital gains (i.e., interest income, if any), we have

$$\wp_{1t} = W_{1t} N_{1t} + \Pi_t = X_{1t} - S_t \tag{82}$$

$$\wp_{2t} = W_{2t} N_{2t} = X_{2t} - \frac{M_t}{P_t^*}. \tag{83}$$

The saving rate for each type of income in the economy is defined as the ratio of net changes in asset position to disposable income in the respective sector:

$$\tau_1 = \frac{S_{t+1} - S_t}{\wp_{1t}} = \frac{(1 + g_t)\, s_{t+1} - s_t}{x_{1t} - s_t} \tag{84}$$

$$\tau_2 = \frac{(M_{t+1} - M_t)\,/P^*}{\wp_{1t}} = \frac{(1 + g)\, m_{t+1} - m_t}{x_{2t} - m_t}. \tag{85}$$

Proposition 7 *Denoting* $\bar{H}_j \equiv \frac{1}{2}\left[H(\theta_{jl}^*) + H(\theta_{jh}^*)\right]$*, the steady-state household saving rate in sector* j *is given by*

$$\tau_j = \frac{\bar{g}\bar{H}_j\left[\theta_{jl}^* R_j + \xi\right] + \frac{1}{2}\bar{g}\xi\left[H(\theta_{jh}^*) - H(\theta_{jl}^*)\right]}{\left[1 + \bar{g} - \bar{H}_j\right]\left[\theta_{jl}^* R_j + \xi\right] - \frac{1}{2}\xi\left[H(\theta_{jh}^*) - H(\theta_{jl}^*)\right]}. \tag{86}$$

Proof. Similar to the proof of Proposition 2. ■

By rearrangement, the saving rate in equation (86) can be shown to be identical to equation (19) in the benchmark model. So, as before, higher income growth can lead to a higher saving rate instead of a higher propensity to consume. The national saving rate (the ratio of investment and net exports to GDP) in the economy is given by $\frac{gk+egm}{y_1+eY_2}$, and the aggregate investment-to-GDP ratio is given by $\frac{gk}{y_1+ey_2}$. Because of trade surplus ($gm > 0$), the national saving rate exceeds domestic investment rate even if the investment rate is high. For example, under the following parameter values, $\beta = 0.96$, $\delta = 0.1$, $\gamma_1 = 0.8$, $\gamma_2 = 0.2$, $\sigma = 1.3$, $\xi = 2$, and $\bar{g} = 0.05$, we have $\frac{gk}{y_1+ey_2} = 0.4$ and $\frac{gk+egm}{y_1+eY_2} = 0.44$. The calibrated values of $\{\gamma_1, \gamma_2\}$ capture the relative sizes of the export and the non-export sectors in China since total exports account for about 20% of GDP. The implied national saving rate under these parameter values is 46% (not including government saving) while the household saving rate in the tradable goods sector is $\frac{gm}{y_2} = 22\%$. Of the 46% national saving rate, aggregate investment accounts for 42% of GDP and net exports accounts for 4% of GDP. These magnitudes are consistent with Chinese data.

Appendix C (Not For Publication)

This appendix shows that the previous results do not hinge on the assumptions of capital controls and a linked exchange rate. As in Appendix B, there are two production sectors in the home country, one producing nontradable goods for domestic consumption and the other producing tradable goods for exports. There are no capital controls and the nominal exchange rate is flexible. The real exchange rate (the relative price of the tradable goods in terms of the nontradable goods) is denoted by e_t. Households can choose to work in either sector and receive real wage W_{jt} in sector $j = 1, 2$. For simplicity and without loss of generality, there are no fixed capital and income shocks (i.e., $\alpha = 0$ and $\xi = 0$). The production technology in sector j is given by $Y_{jt} = A_t F(N_{jt})$, and the competitive real wage is given by $W_{jt} = \frac{dY_{jt}}{dN_{jt}}$. Applying the rescaling factor A_{t-1}, we have $y_{jt} = (1 + g_t) F(N_{jt})$ and $w_{jt} = (1 + g_t) F'(N_{jt})$.

Household i's problem is to solve

$$\max E \sum_{t=0}^{\infty} \beta^t \left\{ \theta_t(i) \left[\gamma_1 \log c_{1t}(i) + \gamma_2 \log c_{2t}(i) \right] - N_{1t}(i) - N_{2t}(i) \right\} \tag{87}$$

subject to

$$c_{1t}(i) + e_t c_{2t}(i) + (1 + g_t) \tilde{m}_{t+1}(i) \leq \tilde{m}_t(i) + w_{1t} N_{1t}(i) + e_t w_{2t} N_{2t}(i) + \Pi_{1t} + e_t \Pi_{2t} \tag{88}$$

$$\tilde{m}_{t+1}(i) \geq 0, \tag{89}$$

where $\tilde{m}_t$ denotes a portfolio of assets with a real gross rate of return equal to 1, and Π_{jt} denotes the profit income distributed from firms in sector j. Because currencies are fully convertible, the household faces only one budget constraint. The budget constraint implies that households can combine income received from either sector to finance consumption of both nontradable goods and tradable goods, as well as asset accumulations.

Proposition 8 *Denoting cash in hand as*

$$x_t \equiv \tilde{m}_t(i) + w_{1t} N_{1t}(i) + e_t w_{2t} N_{2t}(i) + \Pi_{1t} + e_t \Pi_{2t}, \tag{90}$$

the decision rules for consumption, imports, saving, and cash in hand are given, respectively, by

$$c_{1t}(i) = \frac{\gamma_1}{\gamma_1 + \gamma_2} \min \left\{ 1, \frac{\theta_t(i)}{\theta_t^*} \right\} x_t \tag{91}$$

$$e_t c_{2t}(i) = \frac{\gamma_2}{\gamma_1 + \gamma_2} \min \left\{ 1, \frac{\theta_t(i)}{\theta_t^*} \right\} x_t \tag{92}$$

$$(1 + g_t) \tilde{m}_{t+1}(i) = \max \left\{ \frac{\theta_t^* - \theta(i)}{\theta(i)}, 0 \right\} x_t \tag{93}$$

$$x_t = (\gamma_1 + \gamma_2) \theta_t^* R(\theta_t^*) w_{1t}, \tag{94}$$

where the cutoff and real exchange rate are determined by

$$\frac{1+g_t}{w_{1t}} = \beta E_t \frac{1}{w_{1t+1}} R(\theta_t^*) \tag{95}$$

$$e_t = \frac{w_{1t}}{w_{2t}}. \tag{96}$$

Proof. Similar to the proof in Proposition 1. ■

Equations (91)-(96) imply that

$$\frac{c_{1t}}{c_{2t}} = \frac{\gamma_1}{\gamma_2} e_t. \tag{97}$$

Compared with equation (81), lifting capital controls implies efficient allocation here. However, this efficiency gain has no effect on the household saving rate, as the following proposition shows.

Proposition 9 *The current account surplus is given by*

$$CA_t = (\gamma_1 + \gamma_2) \left[(1+g_t)\, \theta_t^* R(\theta_t^*) H(\theta_t^*) - \theta_{t-1}^* R(\theta_{t-1}^*) H(\theta_{t-1}^*) \right], \tag{98}$$

and the aggregate saving rate is given by

$$\tau_t = \frac{\left[(1+g_t)\, \theta_t^* R(\theta_t^*) H(\theta_t^*) - \theta_{t-1}^* R(\theta_{t-1}^*) H(\theta_{t-1}^*) \right]}{\left[(1+g_t)\, \theta_t^* R(\theta_t^*) - \theta_{t-1}^* R(\theta_{t-1}^*) H(\theta_{t-1}^*) \right]}. \tag{99}$$

Proof. Aggregating by the law of large numbers, the household decision rules (91)-(93) become

$$c_{1t} = \frac{\gamma_1}{\gamma_1 + \gamma_2} D(\theta_t^*) x_t \tag{100}$$

$$e_t c_{2t} = \frac{\gamma_2}{\gamma_1 + \gamma_2} D(\theta_t^*) x_t \tag{101}$$

$$(1+g_t)\, \tilde{m}_{t+1} = H(\theta_t^*) x_t \tag{102}$$

The budget constraint becomes

$$c_{1t} + e_t c_{2t} + (1+g_t)\, \tilde{m}_{t+1} = \tilde{m}_t + w_{1t} N_{1t} + e_t w_{2t} N_{2t} + \Pi_{1t} + e_t \Pi_{2t}. \tag{103}$$

Since households own firms, wage income plus profit income should equal the value of output in each sector:

$$w_{jt} N_{jt} + \Pi_{jt} = y_{jt}. \tag{104}$$

Hence, market clearing in the nontradable-goods sector implies

$$c_{1t} = w_{1t} N_{1t} + \Pi_{1t}. \tag{105}$$

Equation (103) then implies

$$e_t c_{2t} + (1 + g_t)\, \tilde{m}_{t+1} = \tilde{m}_t + e_t y_{2t}. \tag{106}$$

So the current account surplus is given by

$$CA_t = e_t y_{2t} - e_t c_{2t} = (1 + g_t)\, \tilde{m}_{t+1} - \tilde{m}_t. \tag{107}$$

Using equation (102), we have

$$\begin{aligned} CA_t &= H(\theta_t^*)x_t - \frac{1}{1+g_{t-1}} H(\theta_{t-1}^*)x_{t-1} \qquad (108) \\ &= (\gamma_1 + \gamma_2)\left[(1+g_t)\,\theta_t^* R(\theta_t^*)H(\theta_t^*) - \theta_{t-1}^* R(\theta_{t-1}^*)H(\theta_{t-1}^*)\right]. \end{aligned}$$

The aggregate saving rate is given by

$$\tau_t = \frac{y_{1t} + e_t y_{2t} - [c_{1t} + e_t c_{2t}]}{y_{1t} + e_t y_{2t}} = \frac{CA_t}{y_{1t} + e_t y_{2t}}, \tag{109}$$

where the denominator is given by

$$\begin{aligned} y_{1t} + e_t y_{2t} &= w_{1t}N_{1t} + \Pi_{1t} + e_t w_{2t} N_{2t} + e_t \Pi_{2t} \qquad (110) \\ &= x_t - \tilde{m}_t \\ &= x_t - \frac{1}{1+g_{t-1}} H(\theta_{t-1}^*)x_{t-1} \\ &= (\gamma_1 + \gamma_2)\left[(1+g_t)\,\theta_t^* R(\theta_t^*) - \theta_{t-1}^* R(\theta_{t-1}^*)H(\theta_{t-1}^*)\right]. \end{aligned}$$

So we have equation (99). ■

Notice that the saving rate is precisely the same as that in equation (25) in the benchmark model (by setting $\xi = 0$ and $\theta_1^* = \theta_2^*$) and is also identical to that in the general model in Appendix B (by setting $\alpha = 0$ in the production function). Clearly, both the current account and the national saving rate are independent of the exchange rate e_t. In the steady state, the current account surplus is given by

$$CA = g\theta R(\theta^*)H(\theta^*), \tag{111}$$

and the saving rate equals

$$\tau = \frac{\bar{g}H(\theta^*)}{1 + \bar{g} - H(\theta^*)}, \tag{112}$$

which is identical to equation (20) in the benchmark model or equation (86) in the general model (by setting $\xi = 0$ and $\alpha = 0$).