Long-Run Money Demand *Redux*[*]

Luca Benati
University of Bern[†]

Robert E. Lucas, Jr.
University of Chicago[‡]

Juan-Pablo Nicolini
Federal Reserve Bank
of Minneapolis[§]

Warren Weber
University of South Carolina,
Bank of Canada, and
Federal Reserve Bank of Atlanta[¶]

Abstract

We explore the long-run demand for M1 based on a dataset comprising 32 countries since 1851. We report six main findings: (1) Evidence of cointegration between velocity and the short rate is widespread. For the United States we detect strong evidence based on three of the adjustments to the standard M1 aggregate originally proposed by Goldfeld and Sichel (1990). (2) Evidence of breaks or time-variation in cointegration relationships is weak to non-existent. (3) For several low-inflation countries the data prefer the specification in the *levels* of velocity and the short rate originally estimated by Selden (1956) and Latané (1960). This is especially clear for the United States. (4) There is *no* evidence of non-linearities at low interest rates. (5) If the data are generated by either a Selden-Latané or a semi-log specification, estimation of a log-log specification spuriously causes estimated elasticities to appear smaller at low interest rates. (6) Using the correct money demand specification has important implications for the ability to correctly estimate the welfare costs of inflation.

[*]We thank G. Cavaliere, J. Fernandez-Villaverde, H. Luetkepohl, A. Marcet, E. Nelson, and P. Teles and participants to seminars at SED 2017, University of Bern, and Minneapolis FED for useful discussions or comments. Special thanks to J. Ayres, N. Gorton, E. Werner, and J. Yano for invaluable help with the data, and to R. Judson for help in adjusting U.S. currency for the fraction held by foreigners. Usual disclaimers apply.

[†]Department of Economics, University of Bern, Schanzeneckstrasse 1, CH-3001, Bern, Switzerland. Email: luca.benati@vwi.unibe.ch

[‡]Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, Illinois 60637, United States. Email: relucas@uchicago.edu

[§]Federal Reserve Bank of Minneapolis, 90 Hennepin Avenue, Minneapolis, MN 55401, United States. Email: juanpa@minneapolisfed.org

[¶] Email: wew@webereconomics.com

1 Introduction

The idea that the quantity of money in an economy can be measured and analyzed with some accuracy, and that changes in this quantity can be related in systematic ways to changes in interest rates, output and prices has had a long but checkered history. The postwar work of Milton Friedman, Anna J. Schwartz, Karl Brunner, Allan Meltzer, and others led to a common vocabulary for different definitions of money, and well-documented data sets covering many countries over long time periods. Theoretical models proposed by William J. Baumol, James Tobin and others described well how changes in the money supply affect other variables, and their predictions conformed well to evidence, at least at the low frequencies. Yet, over recent decades many economists have come to the view that monetary aggregates convey no useful information, and have turned to macroeconomic models in which measures of money do not appear at all. One driver of this change was the alleged instability of the relationship between these series.

In this paper we reconsider the long-run demand for M1 based on a dataset comprising 32 countries, over periods that range in some cases to over 100 years. Evidence of cointegration between M1 velocity and the short rate is widespread, whereas evidence of breaks or time-variation in cointegration relationships is weak to non-existent. For the United States, in particular, we detect strong evidence of cointegration based on three of the alternative adjustments to the Federal Reserve's standard M1 aggregate which had originally been suggested by Stephen M. Goldfeld and Daniel M. Sichel (1990, pp. 314-315) in order to restore the stability of the long-run demand for M1, which had vanished around the mid-1980s. In most cases in which cointegration is not detected, we show *via* Monte Carlo that—conceptually in line with Robert F. Engle and Clive W.J. Granger (1987)—this is, in fact, what we *should* have expected to obtain if cointegration truly were in the data, due to the short sample length and/or the high persistence of the cointegration residual. For several low-inflation countries the data seem to prefer the specification in the *levels* of velocity and the short rate originally estimated by Richard Selden (1956) and Henry Allen Latané (1960), and forgotten ever since. This is especially clear for the United States. We further show that there is *no* evidence of non-linearities at very low interest rates: Rather, the data quite clearly suggest that the relationship between velocity and the short rate is the same both at very low, and at higher interest rates, at least, for the range of variation in interest rates experienced by countries such as the United States. We also show, however, that if the true money demand specification is of either the Selden-Latané, or the semi-log type, the popular practice of estimating a log-log specification *automatically* produces entirely spurious evidence of non-linearities at low interest rates (i.e., of a lower elasticity of money demand at rates approaching zero). Finally, we show, by example, that identifying which, among the three functional forms, is the correct one, has important implications for (e.g.) the ability to correctly estimate the welfare costs of inflation.

The paper is organized as follows. In Section 2 we develop a generalized version of the Baumol-Tobin model that will guide our empirical investigation. We set up the model and then work out its main predictions. We draw the conclusions described above in two steps. The first is described in Section 3, where we simply plot the implied predictions of a particular case of the model against the data for all countries we have. We also show low-frequency evidence, using the band-pass filter. We find this informal visual evidence quite remarkable. The second is described in Sections 4 to 6, where we discuss the methodology we use throughout the paper, we discuss the results from cointegration analysis, and we explore the possible presence of non-linearities at low interest rates. Section 7 provides examples of pitfalls of using the incorrect functional form. Section 8 concludes.

2 A Model of Money Demand

We begin by developing a simple model that will guide our empirical investigation. We study a labor-only, representative agent economy with uncertainty in which making transactions is costly. We let s_t be the state at time t, and let $s^t = \{s_0, s_1, ..., s_t\}$. The preferences of the representative agent are

$$E_0 \sum_{t=0}^{\infty} \beta^t U(x(s^t))$$

where $x(s^t)$ is his consumption given history up through date t, and the function U is differentiable, increasing and concave. The goods production technology is given by $y(s^t) = x(s^t) = z(s^t)l(s^t)$,where $l(s^t)$ is time devoted to the production of the consumption good and $z(s^t)$ is an exogenous stochastic process. The agent is endowed in each period with a unit of time, with $l(s^t)$ allocated to goods production and $1-l(s^t)$ used to carry out transactions.

We assume that households choose the number n of 'trips to the bank' in the manner of the classic Baumol-Tobin (BT) model. At the beginning of a period, a household begins with some nominal wealth that can be allocated to the transactional asset $M(s^t)$, or to non-transactional assets, risk-free government bonds or other state-contingent assets $Q(s^t, s_{t+1})$. During the first of the n subperiods, one member of the household uses money to buy consumption goods. During this same initial sub-period another member of the household produces and sells goods in exchange for money. At the end of the subperiod, producers transfer to the bank the proceeds from their transactions. The situation at the beginning of the second subperiod thus replicates exactly the situation at the beginning of the first. This process is repeated n times during the period. The choice of this variable n will be the only economically relevant decision made by households. Purchases over a period are then subject to a cash in advance constraint $P(s^t)x(s^t) \leq M(s^t)n(s^t)$.

BT assumed that the cost of carrying out these transactions increases linearly in the number of trips. We will consider this case here, and also allow for other forms for this cost function. Specifically, we describe the total cost of making transactions, measured in units of time, by a non-negative, increasing and smooth function $\theta(n(s^t), \nu(s^t))$, where $\nu(s^t)$ is an exogenous stochastic process. The variable $\nu(s^t)$ thus introduces some unobserved randomness into the model. This is essential to motivate the econometric analysis that is the core of the paper. It can be interpreted as changes over time in the technology to adjust portfolios available to households. We assume that $\theta(0, \nu(s^t)) = 0$, so the time involved in no trips to the banks is zero.

Equilibrium in the labor market and the equality of production and consumption imply

$$\begin{aligned} 1 &= l(s^t) + \theta(n(s^t), \nu(s^t)) \\ x(s^t) &= z(s^t)(1 - \theta(n(s^t), \nu(s^t))). \end{aligned}$$

The real wage is equal to $z(s^t)$ and the nominal wage is $z(s^t)P(s^t)$.

At the beginning of each period, an agent starts with nominal wealth $W(s^t)$, which can be allocated to $M(s^t)$, interest bearing bonds, $B(s^t)$, or state-contingent assets $Q(s^t, s_{t+1})$. Let $P^Q(s^t, s_{t+1})$ be the price of an Arrow-Debreu security, bought at t in state s^t, which pays off one unit of money in state s_{t+1}. The agent's allocation of these assets is then restricted by

$$M(s^t) + B(s^t) + \sum_{s_{t+1}} Q(s^t, s_{t+1}) P^Q(s^t, s_{t+1}) \leq W(s^t).$$

If we divide both sides by $P(s^t)$ and let $\widetilde{P}^Q(s^t, s_{t+1})$ denote the price of the state contingent asset divided by the probability of the state, we can write this constraint as

$$m(s^t) + b(s^t) + E\left[q(s^t, s_{t+1})\pi(s^t, s_{t+1})\widetilde{P}^Q(s^t, s_{t+1})\right] \leq \omega(s^t), \tag{1}$$

where lower case letters are real values and where $\pi(s^t, s_{t+1}) \equiv P(s^{t+1})/P(s^t)$ denotes the gross inflation rate between period t in sate s^t and period t+1 in state (s^t, s_{t+1}).

We treat the gross nominal return on short term bonds, $R(s^t)$, as an exogenous process determined by monetary policy.[1] This implies that the behavior of the growth rate of the money supply is restricted by other equilibrium conditions, as is well known and as we show in Appendix B.1.

So far, we have been silent with respect to what our measure of money, $M(s^t)$, accounts for. For the theoretical analysis, we allow for money to pay a nominal return, lower than the one paid by bonds, which we call $R^m(s^t)$. As we will show, this is an important aspect of the theory. We explain our choices for both the particular

[1] When policy is described as a sequence of interest rates, there may be indeterminacy of the price level. Real money balances will however be unique. In this paper we ignore issues regarding the determination of the price level.

monetary aggregate and its return in detail below, when discussing the empirical analysis.

We can now determine the agent's wealth next period, contingent on the actions taken in the current period and the realization of the exogenous shock s_{t+1}. In nominal units, this is

$$\begin{aligned} W(s^t, s_{t+1}) \leq\ & M(s^t)R^m(s^t) + B(s^t)R(s^t) + Q(s^t, s_{t+1}) \\ & + \left[1 - \theta(n(s^t), \nu(s^t))\right] z(s^t)P(s^t) + \tau(s^t, s_{t+1})P(s^{t+1}) - P(s^t)x(s^t), \end{aligned}$$

where $\tau(s^t, s_{t+1})$ is the real value of the monetary transfer the government makes to the representative agent. Dividing by the price level $P(s^{t+1})$, we obtain

$$\begin{aligned} \omega(s^t, s_{t+1}) \leq\ & \frac{m(s^t)R^m(s^t) + b(s^t)R(s^t)}{\pi(s^t, s_{t+1})} + q(s^t, s_{t+1}) \\ & + \frac{[1 - \theta(n(s^t), \nu(s^t))]\, z(s^t) - x(s^t)}{\pi(s^t, s_{t+1})} + \tau(s^t, s_{t+1}). \end{aligned} \tag{2a}$$

Finally, the cash in advance constraint can be written in real terms as

$$x(s^t) \leq m(s^t)n(s^t). \tag{3}$$

We now consider the decision problem of a single, atomistic agent who takes as given the prices $\widetilde{P}^Q(s^t, s_{t+1})$, the inflation rate $\pi(s^t, s_{t+1})$, the interest rate $R(s^t)$, the real wage $z(s^t)$ and the shock $\nu(s^t)$. Given the initial wealth $\omega(s^t)$, this agent chooses his consumption $x\,(s^t)$, the number of bank trips $n\,(s^t)$, and the assets $m\,(s^t)$, $b\,(s^t)$, and $q(s^t, s_{t+1})$ that he chooses to hold. These choices then determine the wealth $\omega(s^t, s_{t+1})$ that he carries into the next period conditional on s_{t+1}. These choices are restricted by equations (1), (2a), and (3).

The Bellman equation describing the decision problem is

$$\begin{aligned} V(\omega) =\ & \max_{x,n,m,b,q(s')} U(x) - \varepsilon\left[m + b + E\left[q(s')\pi(s')\widetilde{P}^Q(s')\right] - \omega\right] - \delta\left[x - mn\right] \\ & + \beta E\left[V\left(\frac{mR^m + bR + [1 - \theta(n)]\, z - x}{\pi(s')} + \tau(s') + q(s')\right)\right] \end{aligned}$$

where, for simplicity, we omitted the dependence of current variables on the state, and where s' denotes the future state.

As we show in Appendix B.2, the first order plus equilibrium conditions can be combined to yield a solution for the equilibrium number of portfolio adjustment, as follows

$$r^* \equiv (R - R^m) = n^2 \frac{\theta_n(n)}{1 - \theta(n)}, \tag{4}$$

which gives an extended squared root formula for the equilibrium value of n.[2]

Note first that, using just sub-indexes to indicate the dependency on the state, the solution for real money balances relative to output is

$$\frac{m(r_t, \nu_t)}{x(r_t^*, \nu_t)} = \frac{M(r_t^*, \nu_t)}{P(r_t^*, \nu_t)} = \frac{1}{n(r_t^*, \nu_t)},$$

which does not depend on z_t.

There are several empirical implications of this solution which do not depend on the particular functional form assumed for the function $\theta(n)$ we would like to discuss now. First, the theory implies an income elasticity equal to one. This is the specification we will study for much of the paper. In Appendix G, we allow for a more general specification which does not restrict the income elasticity to be one, and where we are able to test this unitary income elasticity implication. Second, as $\theta(n_t, \nu_t)$ is differentiable with a strictly positive derivative, some of its properties are inherited by the function $m(r_t^*, \nu_t)$. In particular, up to a linear approximation the stochastic properties of the money to output ratio, m_t/x_t are inherited from the stochastic properties of r_t^* and ν_t. This has testable implications, as long as ν_t is stationary as we will assume throughout the paper. Specifically, if r_t^* is stationary, so should be m_t/x_t, while if it is the case that r_t^* has a unit root, m_t/x_t should have a unit root too. As it turns out, for the specifications of the function $\theta(n_t, \nu_t)$ that we explore in the theory and use in the empirical section these properties hold exactly, not only in a linear approximation.

2.1 Analysis of the solution

We now consider three alternative functional forms for $\theta(n_t, \nu_t)$. They deliver approximations to functional forms which have been used in empirical work, and which we will explore in the following Sections.

The exponential case Consider first the function $\theta(n_t, v_t) = \gamma \nu_t n_t^\sigma$.In this case, equation (4) becomes

$$n_t^{\sigma+1} \frac{\sigma \gamma \nu_t}{1 - \gamma \nu_t n_t^\sigma} = r_t^*.$$

Note that $\gamma \nu_t n_t^\sigma$ is the cost of inflation in units of time, so it represents the welfare cost of inflation as a ratio of first-best output. This ratio is arbitrarily close to zero when the interest rate r_t^* is small. For moderate interest rates, the welfare cost is negligible. Even for relatively high interest rates, estimates of the welfare cost of inflation are hardly above 4%, so the denominator in the expression above would

[2] The squared root formula is the by now classic solution of the Baumol-Tobin formulation.

range from 1 to 0.96. We therefore use the approximation $1-\gamma\nu_t n_t^{\sigma}\simeq 1$ and write the solution as $n_t^{\sigma+1}\sigma\gamma\nu_t \simeq r_t^*$.Taking logs we then obtain

$$\ln\sigma\gamma+\ln\nu_t+(\sigma+1)\ln n_t=\ln r_t^* \tag{5}$$

which is the log-log function typically used in the literature. The BT case is the one obtained by assuming that the function $\theta(n(r_t^*))$ is linear, or $\sigma=1$, which implies an interest rate elasticity of $1/2$.

The Selden-Latané specification A less well know specification is obtained for the following cost function

$$\theta(n_t,\nu_t)=b\ln(\varepsilon+n_t)+\frac{e\varepsilon+\nu_t}{n_t+\varepsilon}-\left(b\ln\varepsilon+\frac{e\varepsilon+\nu_t}{\varepsilon}\right)$$

where the term $\left(b\ln\varepsilon+\frac{e\varepsilon+\nu_t}{\varepsilon}\right)$ guarantees that $\theta(0,\upsilon_t)=0$, and $b>e$ so the function is increasing. The function is concave, so it means that the marginal cost of making transactions is decreasing with the number of transactions (or, what is the same, decreasing with the nominal interest rate).

In this case the solution is given by

$$n_t^2\frac{\frac{1}{(n_t+\varepsilon)^2}\left[(\varepsilon+n_t)b-\varepsilon e-\nu_t\right]}{1-\theta(n_t,\nu_t)}=r_t^*$$

If, as before, we proceed with the approximation $1-\theta(n,\nu_t)\simeq 1$, we obtain

$$\frac{n_t^2}{(n_t+\varepsilon)^2}\left[(\varepsilon+n_t)b-\varepsilon e-\nu_t\right]\simeq r_t^*$$

Thus, for small values of ε, the solution can be approximated by

$$n_t b-\nu_t\simeq r_t^* \tag{6}$$

which implies a linear relationship between velocity and the interest rate.

This empirical specification was used by Richard Selden (1956) over half a century ago, and, to the very best of our knowledge, it has been used again in the literature only once, by Henry Allen Latané (1960). The main reason for considering this long-forgotten specification is that, as we will discuss in Section 5.1.2, for several low-inflation countries—first and foremost, the United States—the data seem to quite clearly prefer it over the traditional log-log one discussed above and the semi-log specification that we discuss next.[3]

[3] An alternative way of deriving the Selden-Latané specification is within the Sidrauski framework: As shown by Benati (2017), the Selden-Latané functional form corresponds to the case of log utility.

The semi-log Finally, consider the following specification

$$\theta(n_t,\nu_t) = -b\frac{\ln(\varepsilon+n_t)}{n_t+\varepsilon} - \frac{k+\nu_t}{n_t+\varepsilon} + \left(b\frac{\ln\varepsilon}{\varepsilon} + \frac{k+\nu_t}{\varepsilon}\right)$$

where again the term in the right-hand side implies $\theta(0,\nu_t)=0$.

In addition, we assume $k+\nu_t > b(1-\ln\varepsilon)$ for all ν_t, so that the function is always increasing in n_t. This function is also concave as the one before. The main difference between this function and the two studied above is that it asymptotes a constant (the term in parenthesis on the right hand side) as the number of trips grow arbitrarily large.

In this case, the solution is given by

$$\frac{n_t^2}{(n_t+\varepsilon)^2}\frac{[b(\ln(\varepsilon+n_t)-1)+k+\nu_t]}{1-\theta(n_t,\nu_t)} = r^*$$

If, as before, we ignore the term $1-\theta(n,\nu_t)$, and also consider relatively low values for ε, we obtain a linear relationship between the log of velocity and the interest rate, which corresponds to the well known semi-log specification.

3 A First Look at the Data

The functional forms considered in the previous Section deliver expressions that can be suitably taken to the data. The formal econometric analysis is presented in the following Sections. As a first descriptive step, in this Section we present the data and compare them to the theory. To do so, we focus on the particular case in which the function θ is linear in n, which corresponds to the BT case of the log-log specification in which the elasticity is constant and equal to 1/2.

Before doing that, we need to address the issue of how we map our theoretical construct M_t to the data. As the model makes clear, the choice of the natural aggregate comes associated to the discussion of the nominal return of that particular aggregate R^m, since real money balances in the model depend not on the interest rate on bonds, but rather on the spread between that rate and the rate paid by money. Since we do not have data on the interest rate paid by deposits, we choose to work with M1, which in most countries includes cash and checking accounts. We will proceed under the assumption that, in the countries we study, checking accounts do not pay interest. Although this is a questionable assumption, it is certainly more appropriate for M1 than for broader aggregates, which typically include interest-paying deposits.[4] As for cash, we follow Alvarez and Lippi (2009) and assume that it entails a *negative* return, associated with the risk of being lost or stolen. Alvarez and Lippi (2009)

[4]It is the case, for instance, that deposits did pay interest in the U.S. after Regulation Q was modified in the early 1980s. It is also the case that some deposits included in M_1 did pay interest in very high-inflation countries as Argentina and Brazil.

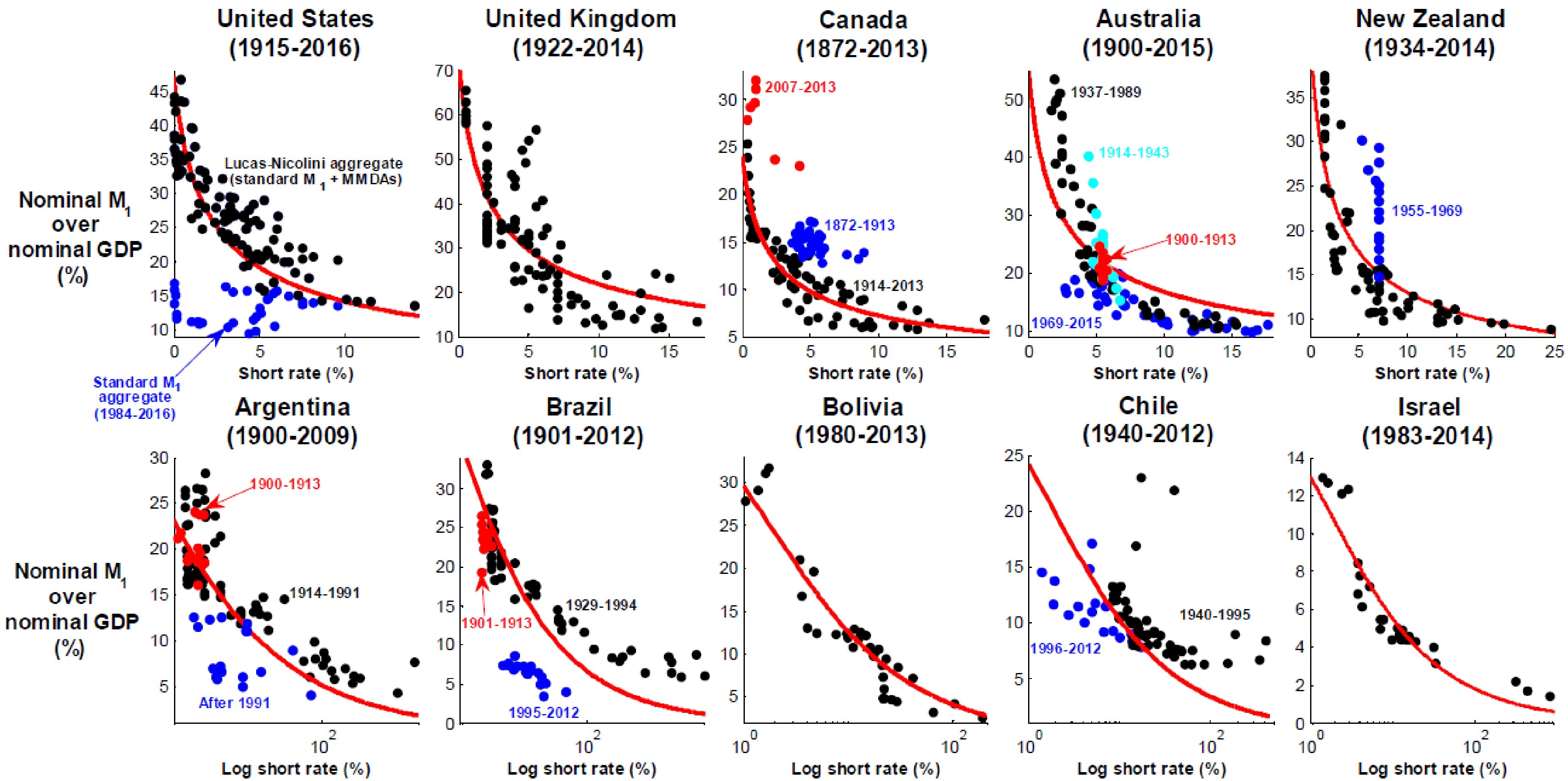

Figure 1a The raw data: short rate, ratio between nominal M_1 and nominal GDP,and calibrated Baumol-Tobin specification

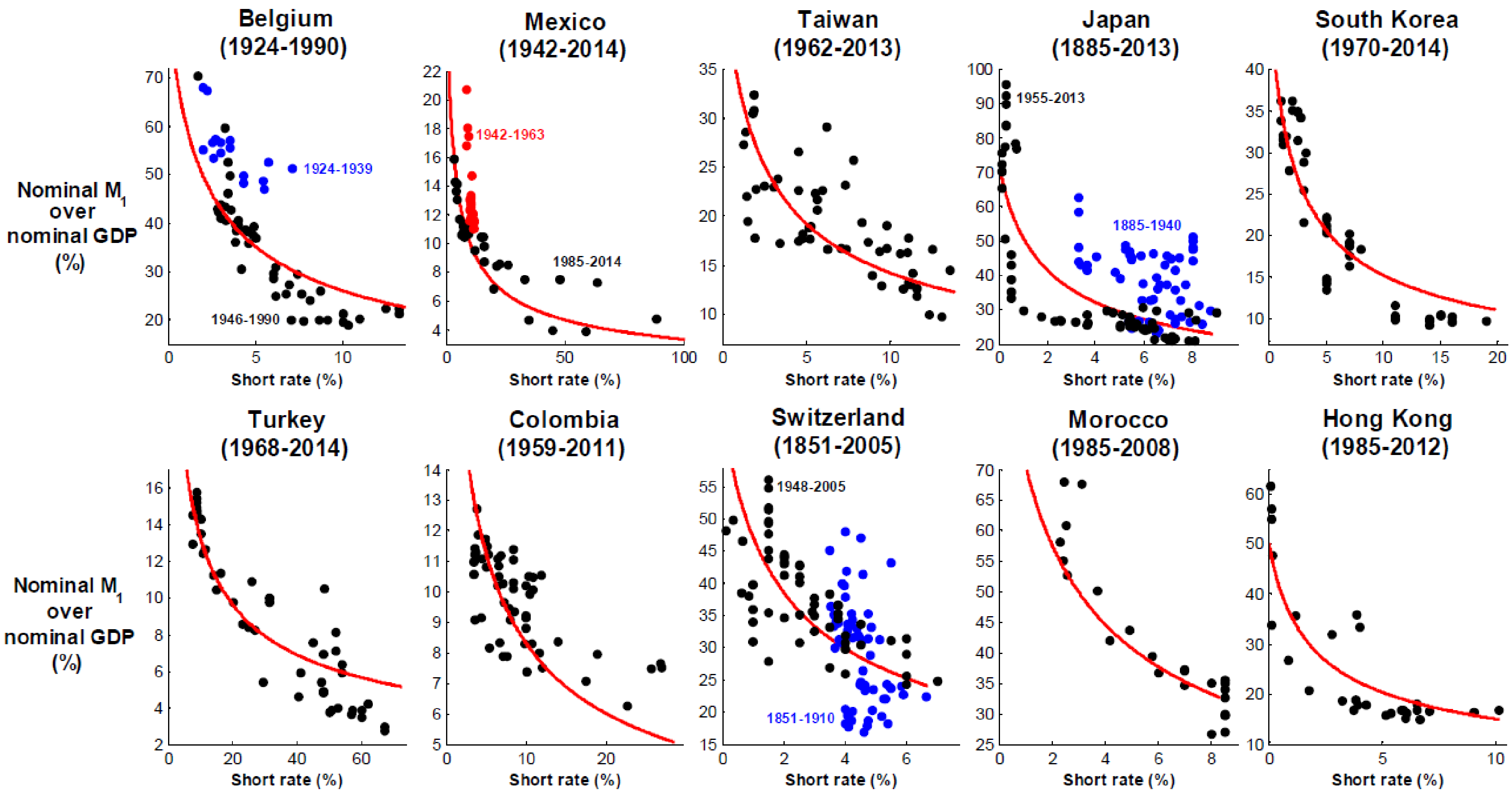

Figure 1b The raw data: short rate, ratio between nominal M_1 and nominal GDP,and calibrated Baumol-Tobin specification

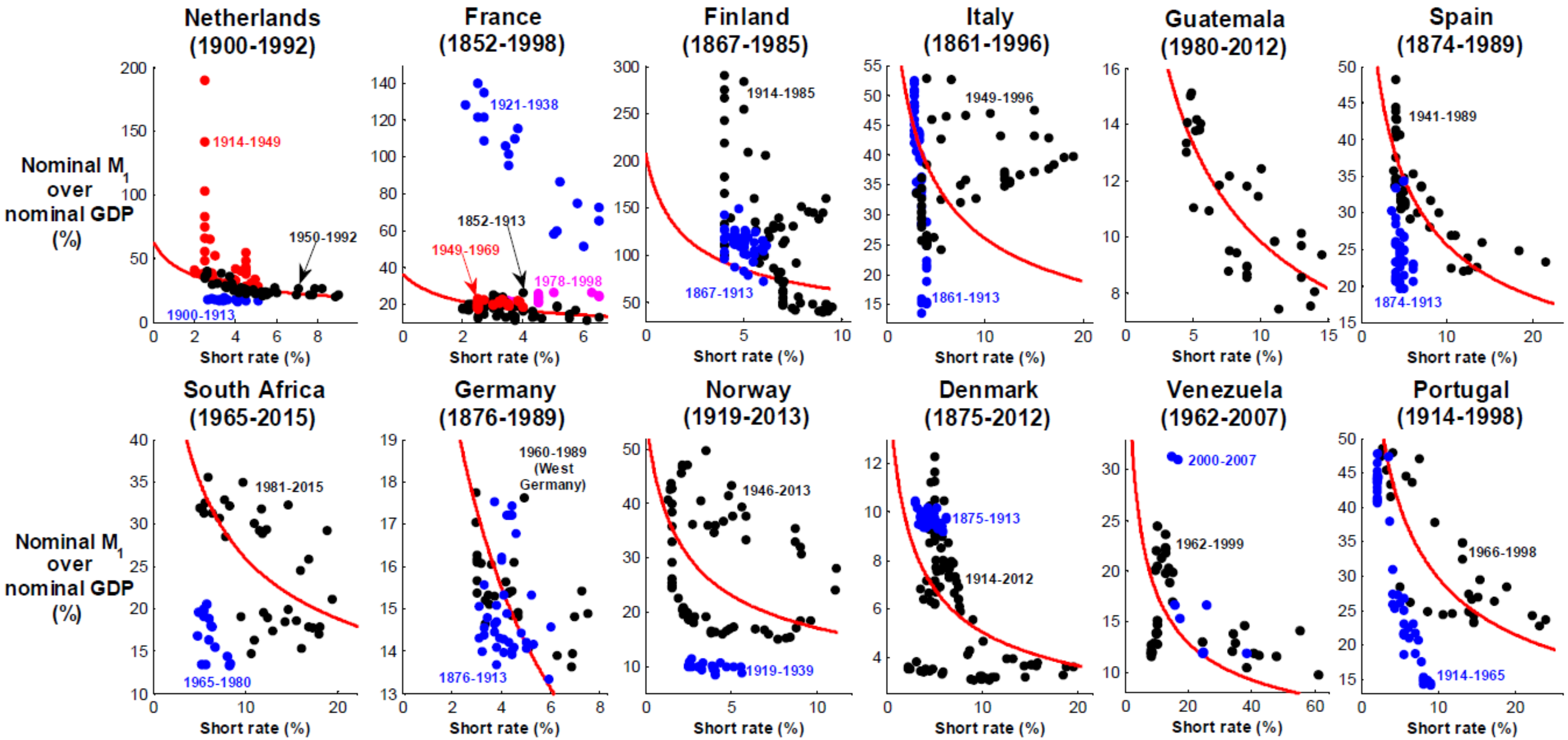

Figure 1c The raw data: short rate, ratio between nominal M_1 and nominal GDP,and calibrated Baumol-Tobin specification

estimate the cost of holding cash to be close to 2% using detailed individual data from Italy. In addition, and to simplify, we assume that cash is about half of total money so that $R^m = 0.99$. This is a very important assumption when considering the log-log specification, since it implies that real money balances have a satiation point when the interest rate on bonds is zero, as is the case for some countries in the sample. Indeed, if on the contrary we assume that $R^m = 1$, the log-log curve goes to infinity as $R \to 1$. As it can be seen in the evidence we show in this Section, this does not seem to be the case for countries that did experience several periods of almost zero interest rates, like the U.S. and Japan. This assumption also plays an important role in the formal econometric tests, as it often improves the performance of the empirical version of the log-log model.[5]

A *caveat* must be made explicit. Payments in this model are for household purchases of final goods, so they ignore other transactions where cash and deposits are used, as paying employees and suppliers of intermediate goods and to clear asset exchanges. We are implicitly treating all these payments—way larger than final goods payments—as proportional to final goods payments. This will require introducing a constant of proportionality as another free parameter in the model, that will be country specific. In other words, the theory we developed is not aimed at matching *levels* of M1 over GDP, but rather *changes* on this ratio as the interest rate changes. One way to see our descriptive exercise therefore is as using one free parameter per country, to allow for a country-specific intercept, while the slope will be given by the BT assumption of a linear technology, so that the elasticity is calibrated to 0.5.

Appendix B describes the data and their sources in detail. All of the series are standard, with the single exception that, for the United States, we consider three of the alternative adjustments to the Federal Reserve's standard M1 aggregate which had originally been suggested by Goldfeld and Sichel (1990, pp. 314-315) in order to restore the stability of the long-run demand for M1, which had vanished around the mid-1980s. Specifically, we augment the standard M1 aggregate with either (*i*) Money Market Deposits Accounts (MMDAs), as in Lucas and Nicolini (2015);[6] (*ii*) Money Market Mutual Funds (MMDFs); or (*iii*) both MMDAs and MMFAs. Finally, for reasons of robustness, for either of the three just-mentioned 'expanded' U.S. M1 aggregates we also consider an alternative version, in which currency has been adjusted along the lines of Judson (2017), in order to take into account of the fact that, since the early 1990s, there has been a sizeable expansion in the fraction of U.S. currency held by foreigners.[7] So, in the end, for the United States we consider *six* alternative

[5] For example, as we will see, for Switzerland for the period 1851-1906, the bootstrapped p-values for Johansen's trace and maximum eigenvalue tests of no cointegration between the logarithm of M_1 velocity and the short rate are 0.160 and 0.113 without the Alvarez-Lippi 1% correction to the short rate, but they fall to 0.094 and 0.057 with the correction.

[6] As discussed by Lucas and Nicolini (2015), the rationale for including MMDAs in M1 is that they perform an economic function similar to the more traditional 'checkable deposit' component of the Federal Reserve's official M1 series.

[7] The way the adjustment is performed is described in detail in Appendix B.

M1 aggregates. As we discuss below, adjusting, or not adjusting for the fraction of U.S. currency held by foreigners does not make a material difference to the results, which originates from the fact that the currency component of M1 is ultimately quite small compared to the deposits component.

Figures 1a to 1c show scatterplots of the short rate and of the ratio between nominal M1 and nominal GDP (i.e., the inverse of M1 velocity), together with the theoretical curve which corresponds to an approximation of equation (5), namely, the BT case so

$$\frac{M_t^j}{Y_t^j} = \frac{A^j}{(r_t^j + 1)^{1/2}} \tag{7}$$

where Y_t^j is nominal income at time t in country j and A^j is a country-specific constant. As mentioned above, we let $r_{t,j}^* = R_t^j - 0.99$, where R_t^j is the gross short term interest rate at time t in country j. In three cases in which we could not find a (sufficiently long) interest rate series,[8] we use inflation as a *proxy* for the opportunity cost of money.

The grouping of countries has been largely arbitrary. The first row of Figure 1a contains countries which belonged, at some point, to the Commonwealth. The second row contains countries which experienced very high inflation rates, and the interest rate (i.e., the horizontal) axis is in a logarithmic scale, due to the magnitudes reached by inflation and interest rates in these countries. In the second row of Figure 1a there are two countries, Argentina and Brazil, for which we highlight the most recent period (since 1991, and since 1995, respectively). These are the two countries in our sample which experienced recurrent periods of very high inflation that lasted over a decade. The blue squares correspond to the periods which followed the successful stabilization years, 1991 for Argentina and 1995 for Brazil. These points are highlighted since in both cases, the points following a successful stabilization lie below the theoretical curve that matches the previous period.

Figure 1b reports countries for which the theoretical curve is visually a still decent approximation to the data. The first row of Figure 1c shows countries for which the fit gets worse,[9] but still there seems to be some relationship between the theory and the data, while the second row of Figure 1c shows countries for which there seems to be no connection between theory and data.

In all of these figures, the data are shown with different colors and markers (dot, square, triangle, and star) under four circumstances: (i) data for the Gold Standard, up until 1913,[10] are always shown with a different color than that used for subsequent

[8] Specifically, Mexico, Chile (for the period 1941-2012), and Brazil (for the period 1934-2012).

[9] For the Netherlands, the two World Wars and their aftermaths had been characterized by an anomalous behaviour of velocity, which in some cases reached values ranging between 50 and almost 200. Because of this, in our econometric analysis we will uniquely focus on the period 1950-1992.

[10] Although we take the Gold Standard to have ended in August 1914, with the outbreak of WWI, in fact, marking the *exact* date of its end is all but impossible, as Richard Nixon's closing of the 'gold window' in August 1971 was the culmination of a decades-long unravelling process which had

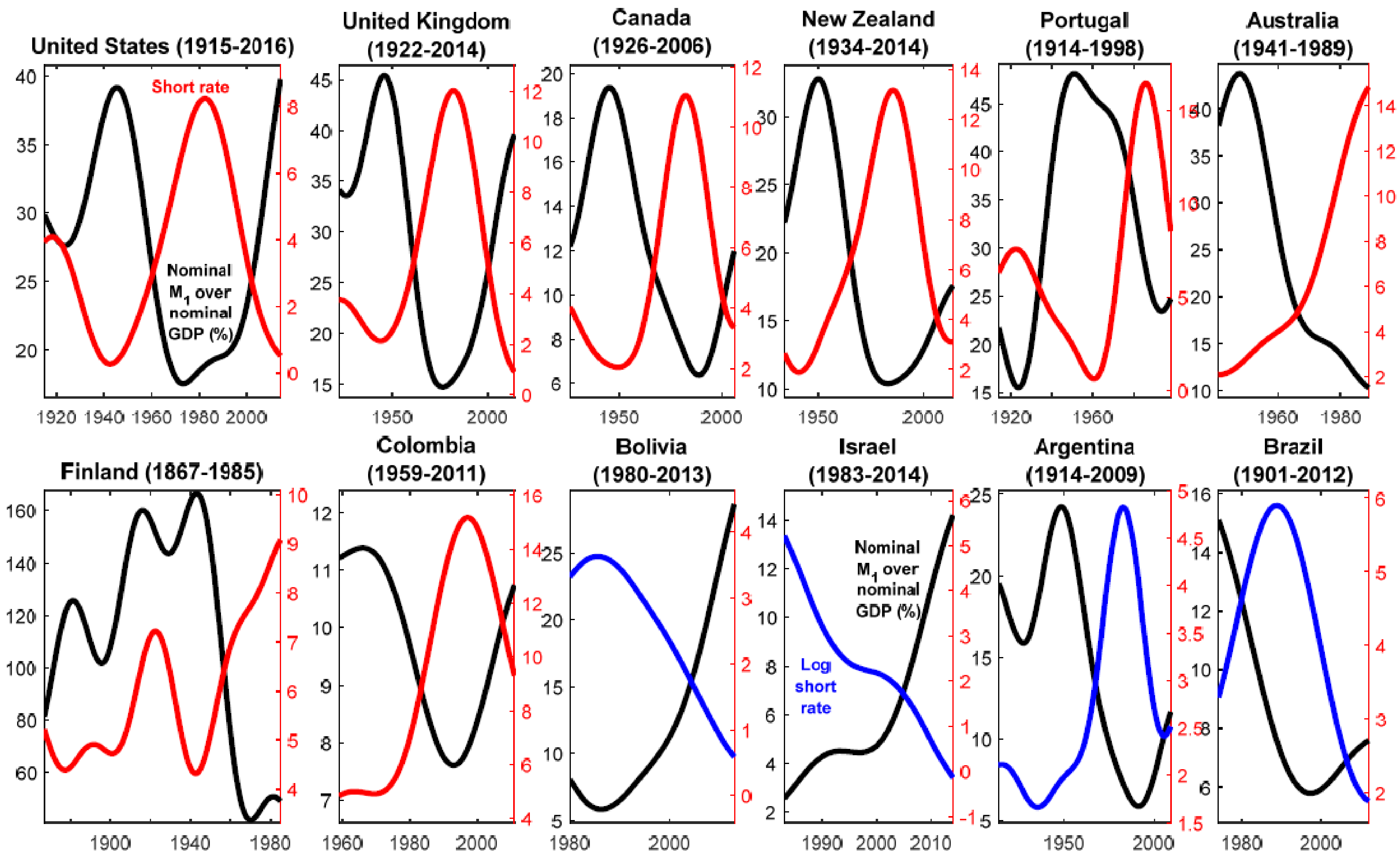

Figure 2a Low-frequency components of short rate and ratio between nominal M_1 and nominal GDP for selected countries

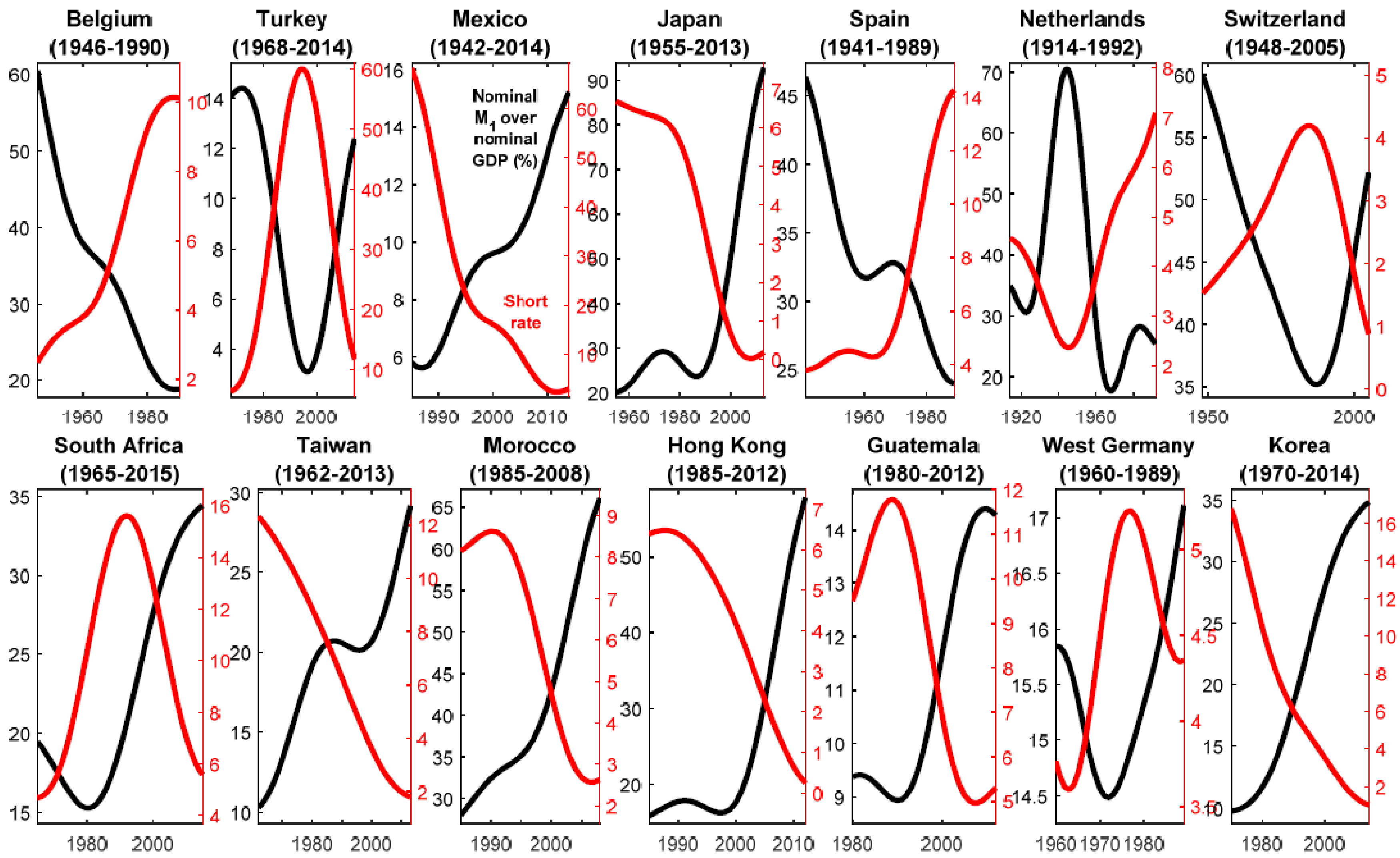

Figure 2b Low-frequency components of short rate and ratio between nominal M_1 and nominal GDP for selected countries

years; (*ii*) when we have data for non-consecutive sub-periods (this is the case, e.g., for France); (*iii*) when we have different series for the short rate which cannot be linked (this is the case, e.g., for Venezuela); and (*iv*) when we want to highlight drastic changes in the relationship between velocity and the short rate (this is the case, e.g., for the Netherlands and Portugal). Finally, for the United States we show with a different color the 'standard' M1 aggregate for the period since 1984, in order to highlight how failure to correct M1 as (e.g.) in Lucas and Nicolini (2015) leads to the apparent breakdown of the relationship between velocity and the short rate documented by several authors.[11] In our view, it is remarkable how well this simple theory performs in this first inspection for a large set of countries, in spite of a few apparent failures.

Figures 2*a* and 2*b* present evidence in the spirit of Lucas (1980), by plotting the low-frequency components of the same series shown in the scatterplots in Figures 1*a*-1*c*.[12] The components have been extracted *via* the filter proposed by Christiano and Fitzgerald (2003).[13] We find this evidence, consistently pointing towards a negative correlation between M1 velocity and the short rate at the very low frequencies, quite simply remarkable. Although the main empirical body of the paper will be based on cointegration tests, the evidence in Figures 2*a*-2*b* is, possibly, even *more convincing*, because it based on a simple technique such as linear filtering, which uniquely hinges on defining a specific frequency band of interest.

Despite the attractiveness of looking at simple plots, however, the previous analysis has several limitations. One would like to formally test if, as some of our simple technologies imply, the ratio between real money balances and output inherit a unit root when the short term interest rate exhibits a unit root. We also want to formally test if, indeed, the estimated elasticities are equal to 1/2, as the simple BT specification suggests, when using the log-log specification. In addition, we would also like to let the data indicate which of the three specifications appear to provide a better fit, and therefore learn something regarding the shape of the function $\theta(n_t, \nu_t)$. To the extent that the interest rate and velocity exhibit a unit root—which appears to be overwhelmingly the case—we can use cointegration techniques to test whether there is a statistical long-run relationship between the two series, and therefore between the interest rate and the ratio of money balances to GDP.

We now turn to a brief discussion of the main features of our approach, and of several methodological issues.

started with WWI. (For a fascinating discussion of such progressive unravelling, see e.g. Barro (1982).) We take August 1914 as the date marking the end of the Gold Standard mostly because we regard WWI as the single most important shock to the system.

[11] See, first and foremost, Friedman and Kuttner (1992).

[12] To be precise: We left out six countries for which evidence was weaker.

[13] Specifically, if the sample length, T, is greater than 50 years, we extract the components of the series with cycles slower than 30 years. If $40<T\leq 50$, $30<T\leq 40$, $20<T\leq 30$, we extract the components with cycles slower than 25, 20, and 15 years, respectively.

4 Main Features of Our Approach

In this paper we explore the long-run demand for M1 *via* cointegration methods. The key reason is that, as we show, in the overwhelming majority of cases the null hypothesis of a unit root cannot be rejected for either velocity, or the short rate, either in levels or in logarithms. At the same time, the debate over the stability of the money demand has long made the distinction between the short and the long run. This distinction is entirely absent in our model, but a large theoretical literature has developed to try to understand the large and sustained deviations of observed real money balances from their theoretical counterparts: The 'short-run' deviations of money demand.[14] The notion of cointegration boils down to the existence of a long-run relationship between series driven by permanent shocks: Those shocks are the source of identification of the relationship between the short rate and velocity. The existence of the cointegration relationship implies that, in the long-run, any permanent increase in the interest rate maps into a corresponding permanent increase in velocity, and therefore decrease in real money balances: The exact amount will be captured by the cointegration vector. Further, any deviation of the two series from their long-run relationship—i.e., the *cointegration residual*—is transitory, and bound to disappear in the long-run. The persistence of the residual is therefore a measure of how long-lived short-run deviations are.

We perform tests taking either *cointegration*, or *no cointegration*, as the null hypothesis (specifically, Shin's, and Johansen's). Although the overwhelming majority of the papers in the literature have been based on Johansen's procedure, there is no reason why no cointegration should be regarded as the 'natural null hypothesis'. Rather, it might be argued that, since we are here searching for a long-run money demand for *transaction* purposes, cointegration should be the natural null,[15] so that tests should just be based on Shin's (1994) procedure. As we discuss in the next sub-sections, however, Monte Carlo evidence clearly suggests that Johansen's procedure performs markedly better than Shin's, and it produces more informative results. Accordingly, in Section 6 we will mostly focus on the results from Johansen's tests.

We perform our analysis separately for the Gold Standard and for the subsequent period. As documented by Barsky (1987) and Benati (2008), the stochastic properties of inflation in the former period had been radically different from the latter, with inflation being most of the times statistically indistinguishable from white noise. By the Fisher equation, this implies that, unless the natural rate of interest had contained a sizeable permanent component (due, e.g., to permanent shifts in trend productivity growth), nominal interest rates should be expected to have been stationary, too, which

[14] See Grossman and Weiss (1983) or Rotemberg (1984) for early contributions or Alvarez and Lippi (2015) for a recent one.

[15] The reason is that basic economic logic suggests that, up to fluctuations in the opportunity cost of money, the nominal quantity of money demanded should be proportional to the nominal volume of transactions (proxied by nominal GDP).

would preclude them from being entered in any cointegrated system, or cointegrating regression.[16] The integration properties of nominal rates during the Gold Standard period ought therefore be separately checked, or otherwise we would run the risk of performing cointegration analysis based on a series which had been stationary for a significant portion of the sample period.

4.1 Integration properties of the data

A necessary condition for using cointegration methods is that all series feature a unit root. Online Appendix C reports results from our extensive investigation of the integration properties of the data (see in particular tables C.1, C.2, and C.3) based on Elliot *et al.*'s (1996) tests. In a nutshell, in the overwhelming majority of the cases the series under investigation are I(1), which justifies our use of cointegration techniques. In the few instances in which this is not the case, we eschew the relevant specifications (e.g., if we can reject the null of a unit root for the logarithm of the short rate, but not for the level, we eschew the log-log specification, and we uniquely focus on the Selden-Latané and semi-log).

4.2 Methodological issues pertaining to cointegration tests

4.2.1 Issues pertaining to bootstrapping

Everything in this paper is bootstrapped—specifically, both the p-values for the cointegration tests, and, more generally, all of the objects of interest, such as the coefficients on the short rate in the estimated money demand functions. In this section we briefly discuss details of the bootstrapping procedures we use, and how such procedures perform, in particular in terms of *comparative* performance. In our discussion we will extensively refer to Appendices D and E, which contain the Monte Carlo evidence motivating both our choices, and the way we will interpret the evidence.

Details of the bootstrapping procedures We bootstrap Johansen's tests *via* the procedure proposed by Cavaliere *et al.* (2012; henceforth, CRT), which is based on the notion of computing critical and p-values by bootstrapping the model which is *relevant under the null hypothesis*.[17] All of the technical details can be found in

[16] A key assumption underlying both Johansen's and Shin's tests is that all of the variables entering either the multivariate system (in the former case), or the single-equation cointegrating regression (in the latter one) are integrated of order one. See Hamilton (1994, very first sentence of p. 636) and Shin (1994, p. 92).

[17] This means that for tests of the null of no cointegration against the alternative of one or more cointegrating vectors the model which is being bootstrapped is a simple, non-cointegrated VAR in differences. For the maximum eigenvalue tests of h versus h+1 cointegrating vectors, on the other hand, the model which ought to be bootstrapped is the VECM estimated under the null of h cointegrating vectors.

CRT, which the reader is referred to. We select the VAR lag order as the maximum[18] between the lag orders chosen by the Schwartz and the Hannan-Quinn criteria[19] for the VAR in levels.

As for Shin's tests, to the very best of our knowledge nobody has yet provided anything comparable to what CRT did for Johansen's procedure. The bootstrap procedure we propose in Appendix E is based on the same general principle underlying CRT, i.e., bootstrapping the model which is relevant under the null. Within the present context, this implies that the process to be bootstrapped is the vector error-correction model (VECM) estimated under the null of one cointegration vector. Apart from this, and with the exception of two minor technical issues we discuss in Section E.2.1 of Appendix E, the procedure is very similar to the one proposed by CRT for Johansen's tests.

Monte Carlo evidence Tables E.1 and E.2 in the online appendix[20] report Monte Carlo evidence on the performance of the two bootstrapping procedures, which is discussed in detail in Sections E.3.1 and E.3.2 of Appendix E. We perform the simulations based on two types of data-generation processes (DGPs), featuring *no cointegration* and *cointegration*, respectively. For either DGP, we consider several alternative sample lengths, and alternative extents of persistence of the cointegration residual. Our main results can be summarized as follows.

As for Johansen's tests, if the true DGP features *no cointegration*, CRT's procedure performs very well irrespective of sample size, with empirical rejection frequencies (ERFs) very close to the nominal size. This is in line with the Monte Carlo evidence reported in CRT's Table I, p. 1731, and with the analogous evidence reported in Benati (2015). If, however, the true DGP features *cointegration*, the tests perform well *only* if the persistence of the cointegration residual is sufficiently low, and/or the sample size is sufficiently large: If the residual is persistent, and/or the sample is short, the tests fails to detect cointegration a non-negligible fraction of the times. This is in line with some of Engle and Granger's (1987) evidence, and it has a simple explanation: As the residual becomes more and more persistent, it gets closer and closer to a random walk (in which case there would be no cointegration), so that the procedure needs larger and larger samples to detect the truth (i.e.: that the residual is highly persistent, but ultimately stationary). As for Shin's tests, if the true DGP features *cointegration*, the more persistent the cointegration residual, the more the bootstrap procedure improves upon Shin's asymptotic critical values.

[18] We consider the maximum between the lag orders chosen by the SIC and HQ criteria because the risk associated with selecting a lag order smaller than the true one (model mis-specification) is more serious than the one resulting from choosing a lag order greater than the true one (over-fitting).

[19] On the other hand, we do not consider the Akaike Information Criterion since, as discussed (e.g.) by Luetkepohl (1991), for systems featuring I(1) series the AIC is an inconsistent lag selection criterion, in the sense of not choosing the correct lag order asymptotically.

[20] The online appendix can be found at: https://sites.google.com/site/lucabenatiswebpage/

If the DGP features *no cointegration*, however, even in large samples the procedure produces ERFs far from the ideal of 100 per cent.

This can be summarized as follows. If Johansen's tests detect cointegration, we should have a reasonable presumption that cointegration is there. If however they do not detect it, a possible explanation is that the sample is too short, and/or the cointegration residual is highly persistent. As for Shin's tests, lack of rejection of cointegration does not represent strong evidence that cointegration truly is there. Further, rejection of the null does not appear to be especially informative about the true nature of the DGP, as the ERFs are not markedly different conditional on the two possible states of the world. To put it differently, results from Shin's tests appear, overall, as less informative than those produced by Johansen's.

We now turn to the issue of how cointegration tests should be interpreted.

4.2.2 Interpreting the results from cointegration tests *via* Monte Carlo

Tables SELA.1, SL.1, LL.1, and LLCO.1 in the online appendix report Hansen (1999) 'grid bootstrap' median-unbiased (henceforth, MUB) estimates of the sum of the AR coefficients in AR(2) representations for the 'candidate cointegration residuals' in our dataset.[21] By 'candidate cointegration residual' (henceforth, CCR) we mean the linear combination of the I(1) variables in the system which will indeed be regarded as a cointegration residual *if* cointegration is detected.[22] Evidence points towards both a non-negligible extent of persistence of the CCRs, and a wide extent of heterogeneity across countries. Focusing on results based on the log-log specification for high-inflation countries, and the Selden-Latané one for all other countries, the MUB estimate based on Johansen's estimator of the cointegration vector—let's label it as $\hat{\rho}^J_{MUB}$—ranges from a minimum of 0.30 for Australia, to a maximum of 1.00 for Portugal (1966-1998). By classifying the $\hat{\rho}^J_{MUB}$'s, in an admittedly arbitrary fashion, as 'highly persistent' ($\hat{\rho}^J_{MUB} \geq 0.8$); 'moderately persistent' ($0.4 < \hat{\rho}^J_{MUB} < 0.8$); and 'not very persistent' ($\hat{\rho}^J_{MUB} \leq 0.4$), we end up with sixteen $\hat{\rho}^J_{MUB}$'s in the first group, fifteen in the second, and three in the third. Results based on Stock and Watson's estimator point, overall, towards an even greater extent of persistence.

Under these circumstances, statistical tests will often have a hard time in detecting cointegration even if it truly is there. This will be especially so in those cases in which $\hat{\rho}_{MUB}$ is high and the sample period is comparatively short. This implies that results from cointegration tests should *not* be taken strictly at face value, and they should rather be interpreted in the light of the Monte Carlo evidence in Tables E.1

[21] Results are based on 2,000 bootstrap replications for each possible value of the sum of the AR coefficients in the grid. Bootstrapping has been performed as in Diebold and Chen (1996). For reasons of robustness, we report results based on two alternative estimators of the cointegration vector, Johansen's, and Stock and Watson's (1993).

[22] The reason why we label it as 'candidate' is because, as the Monte Carlo evidence in the previous section has shown, if a residual is highly persistent, cointegration might well *not* be detected *even if it is there*, which would prevent the candidate from being identified as a true cointegration residual.

and E.2. In what follows we will therefore also report Monte Carlo-based ERFs of the tests computed *under the null of cointegration*. Specifically, we will estimate the VECM under the null of one cointegration vector; we will stochastically simulate it 2,000 times; and for each artificial sample we will perform the same bootstrapped cointegration tests we previously performed based on the actual data. This will allow us to gauge an idea of how likely it would be to detect cointegration if it were truly there in *all* of the samples we are working with.[23]

4.2.3 Testing for stability in cointegration relationships

We test for stability of cointegration relationships based on the three tests discussed by Hansen and Johansen (1999): Two Nyblom-type tests for stability in the cointegration vector and the vector of loading coefficients, respectively; and a fluctuation test, which is essentially a joint test for time-variation in the cointegration vector and the loadings. In either case, we bootstrap the test statistics *via* CRT's procedure, based on the VECM estimated conditional on one cointegration vector, and not featuring any break, or time-variation of any kind.

Table H.1 in the Appendix reports Monte Carlo evidence on the performance of the tests conditional on bivariate cointegrated DGPs, for alternative sample lengths, and alternative extents of persistence of the cointegration residual, which is modelled as an AR(1). The main results can be summarized as follows. The two Nyblom-type tests exhibit an overall reasonable performance, incorrectly rejecting the null of no time-variation, most of the time, at roughly the nominal size. Crucially, this is the case irrespective of the sample length, and of the persistence of the cointegration residual. The fluctuation test, on the other hand, exhibits a good performance only if the persistence of the cointegration residual is low. The higher the residual's persistence, however, the worse the performance, so that (e.g.) when the AR root of the residual is equal to 0.95, for a sample length $T = 50$, the test rejects at twice the nominal size. This is clearly problematic since, as previously discussed, residuals are typically moderately-to-highly persistent. In what follows we will therefore focus on the results from the two Nyblom-type tests, whereas we will eschew results from fluctuations tests (these results are however reported in tables H.2 and H.5).

We now turn to the results from cointegration tests, and tests for time-variation in cointegration relationships.

5 Searching for a Long-Run Money Demand

Table 1*a* reports results from Johansen's maximum eigenvalue tests of 0 *versus* 1 cointegration vectors for the United States, together with the Monte Carlo-based ERFs computed conditional on the null of one cointegration vector, whereas Table

[23] This is very much in the spirit of Lucas' (1988) interpretation of econometric results which, taken at face value, appeared to contradict the findings of Meltzer (1963).

1b reports the corresponding results for all other countries. Tables G.1 and G2 in the appendix report the corresponding results based on Shin's tests. The full set of results based on Selden-Latané, semi-log, log-log, and log-log without the Alvarez-Lippi correction to the short rate are reported in the online appendix, in Tables SELA.2, SL.2, LLCO.2, and LL.2, respectively, and are discussed in Appendix G.

Figures SELA.1-SELA.6, SL.1-SL.6, LLCO.1-LLCO.6, and LL.1-LL.6 in the online appendix report, based on either specification, in the top rows the candidate cointegration residuals produced by either Johansen's or Stock and Watson's (1993) estimators; and in the bottom rows the bootstrapped distributions of the corresponding estimates of the coefficient on (the log of) the short rate. For each bootstrapped distribution we also report the mean, the median, and the 5th and 95th percentiles. In all cases we report both candidate cointegration residuals, and estimates of the coefficients on the short rate, for *all* countries, rather than only for those for which statistical tests detect evidence of cointegration.

5.1 Evidence from cointegration tests

5.1.1 Testing the null of cointegration

Although this paper mostly focuses on the results produced by bivariate systems, we want to briefly discuss those produced by Shin's tests of the null of cointegration applied to unrestricted specifications featuring (the logarithm of) the short rate, and the logarithms of nominal GDP and M1. The reason for doing so is that they represent one 'extreme end' of the spectrum within the full set of results: As we discuss in Appendix G.1, based on unrestricted three-variables systems *it is almost impossible to reject the null of cointegration*.[24] Evidence is just slightly weaker for tests based on bivariate specifications for velocity and the short rate, in which unitary income elasticity has been imposed from the outset: We obtain just ten rejection of the null based on Selden-Latané, thirteen based on semi-log, and seven based on log-log.

For the reasons discussed in Section 4.1.1, however, these results should be downplayed: As we stressed there, lack of rejection of the null of cointegration based on Shin's tests does not represent strong evidence that cointegration truly is there. We therefore turn to discussing the results from Johansen's tests, which, as we pointed out, appear as uniformly more informative than Shin's.

5.1.2 Testing the null of no cointegration

Evidence from bivariate systems for velocity and the short rate Tables 1a and 1b contain this paper's most important body of evidence (at least, as far as

[24] Specifically, at the 10 per cent level we obtain *just four* rejections of the null based on the semi-log specification, whereas based on the log-log specification with the 1 per cent correction to the short rate we obtain *only one* rejection. (For the Selden-Latané specification it is not possible to consider unrestricted specifications.)

Table 1a United States: Bootstrapped p-values[a] for Johansen's maximum eigenvalue[b] tests for (log) M1 velocity and (the log of) a short-term rate, and Monte-Carlo-based empirical rejection frequencies of the tests under the null of cointegration

		I: Bootstrapped p-values			II: Empirical rejection frequencies		
Monetary aggregate	*Period*	*Selden-Latané*	*Semi-log*	*Log-log*	*Selden-Latané*	*Semi-log*	*Log-log*
		Without adjusting for the share of currency held abroad					
Standard M1	1915-2016	0.838	0.712	0.531	0.099	0.129	0.283
Standard M1 + MMDAs	1915-2016	0.044	0.081	0.155	–	–	0.662
Standard M1 + MMMFs	1915-2016	0.006	0.022	0.289	–	–	0.649
Standard M1 + MMDAs + MMMFs	1915-2016	0.002	0.007	0.617	–	–	0.444
Standard M1	1926-2016	0.913	0.822	0.694	0.134	0.140	0.140
Standard M1 + MMDAs	1926-2016	0.070	0.104	0.235	–	0.782	0.724
Standard M1 + MMMFs	1926-2016	0.003	0.021	0.405	–	–	0.619
Standard M1 + MMDAs + MMMFs	1926-2016	0.003	0.008	0.700	–	–	0.424
		Adjusting for the share of currency held abroad[c]					
Standard M1	1926-2016	0.960	0.780	0.418	0.153	0.138	0.163
Standard M1 + MMDAs	1926-2016	0.069	0.170	0.120	–	0.693	0.700
Standard M1 + MMMFs	1926-2016	0.003	0.030	0.383	–	–	0.538
Standard M1 + MMDAs + MMMFs	1926-2016	0.118	0.020	0.707	0.647	–	0.254

[a] Based on 10,000 bootstrap replications. [b] Null of 0 *versus* 1 cointegration vectors. [c] Adjustment performed as in Judson (2017); for details, see text.

statistical tests are concerned). In either table we highlighted in yellow all p-values for maximum eigenvalue tests smaller than 10 per cent; and all ERFs smaller than 50 per cent, corresponding to a less-than-even chance of detecting cointegration if this is truly in the data.

The United States Starting from the United States, which has been the focus of most previous investigations, the main results in Table 1a can be summarized as follows:

(i) In line with, e.g., Friedman and Kuttner (1992), based on the standard M1 aggregate the null of no cointegration is never rejected.[25]

(ii) A second consistent pattern is that no cointegration is also never rejected based on the log-log specification. It is important to stress that these results are based on appyling Alvarez and Lippi's 1 per cent correction to the short rate, which should drastically improve the fit at levels of the short rate close to zero. In spite of this, the log-log functional form still does not allow to detect cointegration.

(iii) Based on the Selden-Latané and semi-log specifications, on the other hand, evidence of cointegration is very strong across the board. Specifically, based on the M1 aggregate which has not been adjusted for the share of currency held abroad, cointegration is *always* detected, based on either specification, and either of the three 'expanded' M1 aggregates. Based on the adjusted aggregates, evidence is slightly weaker, and cointegration is detected in four cases out of six. It is to be noticed, however, that this is partly explained by the shorter sample period, as clearly shown by the set of results based on the unadjusted aggregates for the sample period 1926-2016 (i.e., the same sample as for the adjusted aggregates). E.g., focusing on the Lucas-Nicolini aggregate, the p-value produced by the Selden-Latané specification increases from 0.044 to 0.07 *uniquely* as a consequence of the shorter sample period. Further adjusting the aggregate as in Judson (2017) does not make any material difference, with the p-value being now equal to 0.069. For the semi-log, on the other hand, this is the case only to a minor extent, with the p-value increasing from 0.081 to 0.104 due to the shorter sample, and further increasing to 0.170 as a result of the M1 adjustment. So in the end the shorter sample explains only part of the deterioration of the results from cointegration tests based on adjusted aggregates.

Should we put more trust in the results based on the unadjusted or adjusted aggregates? In our own view, the answer is not obvious: Although Judson's clearly is a sensible approach, the adjustment is still based on an estimate of the amount of currency held by foreigners. Because of this, it is not entirely obvious that results based on adjusted aggregates should be preferred. A key advantage of unadjusted aggregates is that (if we trust the data-collection process) we know exactly what these aggregates are, and since, as mentioned in Section 3, currency is quite small

[25] It is to be noticed, however, that the ERFs are uniformly very low (ranging from 0.099 to 0.283), thus implying that if cointegration were in the data, there would be little chance of detecting it.

compared to deposits,[26] adjusting or not adjusting the aggregates should not make much of a difference. At any rate, we report all the results so that the reader can decide for herself.

A more important issue—as we illustrate by example in Section 7—is which specification (Selden-Latané, or semi-log) should be preferred. If we interpret the p-values produced by the two specifications for either aggregate as an informal 'test' of which of the two functional forms the data would seem to prefer, evidence points quite clearly to Selden-Latané, as in only a single case out of nine (the adjusted aggregate also including MMMFs) the p-value produced by the semi-log is smaller than the one produced by Selden-Latané. In all other cases, the opposite is true. As we show in Section 7.1, this has (at least) one important implication: If the true functional form is indeed Selden-Latané, estimating a log-log specification produces entirely spurious evidence of money demand non-linearities at low interest rates.

Other countries Turning to Table 1*b*, the following main findings should be highlighted:

(*i*) cointegration is detected based on *all* estimated specifications[27] for Argentina, Brazil, Guatemala, Israel, Norway, Portugal (1914-1965), South Africa, Switzerland (1948-2005), and the United Kingdom. It is further detected for two specifications out of three for Australia (1969-2015), Belgium, Canada (for either period), Chile (1940-1995), Korea, Mexico, and New Zealand. Finally, in three cases—Bolivia, Chile (1941-2012), and Switzerland under the Gold Standard—cointegration is detected based just on a single specification.

(*ii*) In most cases in which cointegration is not detected, the ERFs show that this is what we *should* indeed expect if cointegration truly were in the data. Consider, for example, Australia for the period 1941-1989. In spite of the very strong negative correlation between the low-frequency components of the two series in the top-right panel of Figure 2*a*, the p-values in Table 1*b* range between 0.642 and 0.973. Crucially, however, the ERFs show that if cointegration were in the data, the chance of detecting it would be *between 8 and 20 per cent*. The same argument holds for *all* three specifications for Germany and Japan under the Gold Standard; and for Colombia, Finland, the Netherlands, Portugal (1966-1998), Venezuela, and West Germany. This is a simple explanation for the results from cointegration tests, in spite of the fact that, in most of these cases, visual evidence points towards a very strong correlation between velocity and the short rate.

(*iii*) As for which specification the data seem to prefer, evidence is much less clear-cut than for the United States. By applying the same informal argument we used for

[26] E.g., for the broadest M1 aggregate also including MMMFs, currency has oscillated, since 1990, between 10.0 and 15.6 per cent of the overall aggregate. So even if, on average, roughly half has been in the hands of foreigners, this means that we are talking about 5 to 8 per cent of overall M1.

[27] We say 'estimated specifications' because, for the reasons discussed in Section 4.1, in a few cases we only estimate two functional forms, rather than three.

Table 1b Bootstrapped p-values[a] for Johansen's maximum eigenvalue[b] tests for (log) M1 velocity and (the log of) a short-term rate, and Monte Carlo-based empirical rejection frequencies of the tests under the null of cointegration

		I: Bootstrapped p-values			II: Empirical rejection frequencies		
Country	*Period*	*Selden-Latané*	*Semi-log*	*Log-log*	*Selden-Latané*	*Semi-log*	*Log-log*
Argentina	1914-2009	–	0.010	0.023	–	0.993	0.801
Australia	1941-1989	0.642	0.973	0.709	0.168	0.079	0.200
	1969-2015	0.063	0.099	0.405	0.720	0.677	0.438
Belgium	1946-1990	0.361	0.016	0.010	0.699	0.635	0.744
Bolivia	1980-2013	0.053	0.423	0.154	0.686	0.414	0.114
Brazil	1974-2012	0.008	0.042	0.093	0.625	0.995	0.658
	1934-2012	–	0.004	0.037	–	0.510	0.339
Canada	1926-2006	0.007	0.078	0.229	0.968	0.804	0.630
	1967-2012	0.007	0.117	0.003	0.965	0.740	0.942
Chile	1940-1995	0.133	0.065	0.033	0.111	0.156	0.864
	1941-2012	0.035	0.151	0.119	0.824	0.371	0.624
Colombia	1959-2011	0.717	0.692	0.872	0.169	0.144	0.143
Finland	1914-1985	0.622	0.659	0.839	0.231	0.218	0.209
Germany	1876-1913	0.503	0.534	0.532	0.152	0.141	0.146
Guatemala	1980-2012	0.049	0.043	0.052	0.536	0.529	0.454
Japan	1885-1913	0.333	0.365	0.331	0.159	0.135	0.144
	1955-2013	0.427	0.154	0.120	0.363	0.596	0.605
Korea	1970-2014	0.060	0.070	0.715	0.086	0.099	0.172
Israel	1983-2014	0.000	0.000	0.000	0.767	0.646	0.204
Italy	1949-1996	0.171	0.182	–	0.629	0.581	–
Mexico	1985-2014	0.007	0.002	0.205	0.537	0.313	0.190
Netherlands	1950-1992	0.349	0.286	0.401	0.463	0.427	0.324
New Zealand	1934-2014	0.093	0.109	0.044	0.690	0.686	0.822
Norway	1946-2013	0.031	0.021	0.015	0.749	0.756	0.792
Portugal	1914-1965	0.004	0.038	0.032	0.183	0.154	0.481
	1966-1998	0.511	0.722	0.125	0.100	0.064	0.316
South Africa	1967-2014	0.068	0.060	0.080	0.563	0.562	0.330
Spain	1941-1989	0.120	0.215	0.537	0.636	0.472	0.197
Switzerland	1851-1906	0.158	0.115	0.057	0.802	0.787	0.788
	1948-2005	0.000	0.000	0.001	0.923	0.891	0.775
Taiwan	1962-2013	–	0.742	0.794	–	0.167	0.141
Turkey	1968-2014	0.896	0.444	–	0.229	0.546	–
United Kingdom	1922-2014	0.011	0.021	0.077	0.975	0.976	0.659
Venezuela	1962-1999	0.776	0.844	0.888	0.087	0.079	0.062
West Germany	1960-1989	–	0.857	0.261	–	0.148	0.221

[a] Based on 10,000 bootstrap replications. [b] Null of 0 *versus* 1 cointegration vectors.

the United States, based on the comparative p-values produced by the three specifications, it could be possible to argue that Selden-Latané is preferred for Australia (1969-2015), Bolivia, Brazil (1974-2012), Canada (1926-2006), Chile (1941-2012), Finland, Germany under the Gold Standard, Korea, Portugal (1914-1965), Spain, and the United Kingdom. By the same token, the semi-log seems to be preferred for Argentina, Brazil (1934-2012), Colombia, Guatemala, Mexico, the Netherlands, South Africa, and Taiwan, whereas the log-log appear as preferred for Belgium, Canada (1967-2012), Chile (1940-1995), Japan for either period, New Zealand, Norway, Portugal (1966-1998), Switzerland under the Gold Standard, and West Germany.

So, although for a few countries the preference for one particular specification is quite clear (e.g., Selden-Latané for the United Kingdom), the data do not exhibit a consistent pattern across countries. In the light of the theoretical discussion in Section 2, a natural explanation is that the technology available to households in order to adjust their portfolios differs across countries.

Towards a unified framework? An alternative possible interpretation of the evidence in Tables $1a$-$1b$ is however that the true specification for money demand is the same for all countries, and that the lack of consistency in the results from cointegration tests simply reflects a combination of small-sample issues, and the 'luck of the draw' which is unavoidably associated with statistical testing. Under this interpretation, the next step would be trying to assess which, among the three functional forms, could be regarded as the most plausible representation of the data for the entire set of countries. Since the three models are not nested, however, such an assessment is not straightforward: E.g., even if we used panel methods, we would not be able to compare the three specifications based on the entire dataset.

One possible avenue would be to compare the point estimates of the parameters on the (log) short rate produced by either specification for two sets of low- and high-inflation countries. Intuitively, if a specific functional form provides a better overall fit for the entire set of countries, it should produce less variation in the point estimates across the two sets of low- and high-inflation countries. The first set comprises the United States, the United Kingdom, Australia, Canada and New Zealand: All of these countries experienced important variations on their nominal interest rates, but they are low-inflation. The second group is composed by Argentina, Brazil, Bolivia, Chile, and Israel, all high-inflation countries. For all countries in either set there is strong evidence of cointegration based on at least one of the three specifications. Unfortunately, even this approach does not produce clear-cut results. Consider, for example, the comparison between the two 'polar cases', Selden-Latané and log-log. Based on the results reported in Figures SELA.1-SELA.6 and LLCO.1-LLCO.6 in the online Appendix, and considering, for the sake of the argument, the median estimates of the coefficients on the (log) short rate produced by Johansen's estimator, the ranges of estimated coefficients for the low- and high-inflation countries produced by Selden-Latané are [-1.281; -0.446] and [-1.293; -0.009], respectively, whereas the corresponding ranges produced by log-log are [-0.96; -0.43] and [-0.67; -0.03], respectively. Based on these numbers, it is not at all clear which of the two specifications should be thought of as producing the more stable estimated coefficients across the two sets of low- and high-inflation countries. Results based on Stock and Watson's (1993) estimator of the cointegration vector are qualitatively the same. Finally, results for the other possible comparisons across functional forms are also equally inconclusive.

Evidence from unrestricted specifications Turning to unrestricted specifications, Tables SL.4, LL.4, and LLCO.4 report results from Johansen's tests based on systems featuring the (logarithm of) the short rate, and the logarithms of nominal GDP and M1. As we discuss more extensively in Appendix G.3, based on the semi-log specification we detect cointegration for most high-inflation countries (Argentina, Bolivia, Brazil, Chile, and Israel); for post-WWII Japan and Switzerland; for the Netherlands, Norway, Taiwan, and the United Kingdom. Based on the log-log specification with the 1 per correction to the short rate, cointegration is detected for post-WWII Japan and Switzerland; and for Argentina, Brazil, Canada, Korea, Israel, the Netherlands, Norway, and Portugal (1914-1965).

We next turn to the issue of stability of the cointegration relationship.

5.2 Testing for stability in cointegration relationships

Tables H.2 and H.3 in the Appendix report results from Hansen and Johansen's (1999) Nyblom-type tests for stability in either the cointegration vector, or the vector of loading coefficients. The key finding in the two tables is that evidence of breaks in either the cointegration vector, or the loading coefficients, is weak to non-existent. Specifically, for the United States, based on the Selden-Latané specification the null of no breaks in either feature is *never* rejected for either of the three 'expanded' M1 aggregates. Stability in the cointegration vector is also never rejected based on semi-log and log-log, whereas breaks in the loadings are detected based on the semi-log specification, and in one case out of six based on the log-log. Evidence for other countries is qualitatively the same. E.g., based on Selden-Latané, stability in the cointegration vector is rejected in three cases, whereas stability in the loadings is rejected in six cases. Results for the other two specifications are along the same lines.

5.3 The estimated coefficients on the short rate

We now turn our discussion to the bottom rows of Figures SELA.1-SELA.6, SL.1-SL.6, and LLCO.1-LLCO.6 in the online Appendix, showing the estimated coefficients on the (log) short rate; and to Tables SELA.3, SL.3, and LLCO.3 in the online appendix, reporting bootstrapped p-values for testing the null hypothesis that the coefficients be equal to a benchmark value. For the log-log specification, the natural benchmark is Baumol and Tobin's, i.e. -1/2. By the same token, based on previous evidence—see e.g. Stock and Watson (1993)—the natural benchmark for the semi-log is -0.1. As for Selden-Latané, since theory does not provide us with a numerical benchmark, we set it to -0.4, which is roughly equal to the median or modal estimates we obtain for the United States based on the Lucas-Nicolini aggregate (see Figure SELA.6).[28] As for the log-log, results are overall mixed, with the Baumol-Tobin null being rejected in 17 cases out of 32 based on Johansen's estimator of the cointegration vector, and in 21 cases based on Stock and Watson's. As for Selden-Latané, the null of -0.4 is rejected in 19 cases out of 33 based on Johansen's estimator, and in 25 cases based on Stock and Watson's. Finally, as for the semi-log we reject the null in 20 cases out of 34 based on Johansen, and in 27 based on Stock and Watson.

We now turn to two substantive issues: Whether there might be sizeable non-linearities in money demand at low interest rates, and the pitfalls originating from using the incorrect money demand specification.

6 Are There Sizeable Non-Linearities in the Behavior of Money Demand at Low Interest Rates?

A strand of literature—see, first and foremost Mulligan and Sala-i-Martin (2000)—has argued that, at low interest rates, money demand exhibits sizeable non-linearities due to the presence of fixed costs associated with the decision to participate, or not to participate, in financial markets. This implies that at sufficiently low interest rates money demand (and therefore money velocity) should be largely *unresponsive* to changes in interest rates, since most (or all) households will simply not participate in financial markets. The implication is that it should not be possible to reliably estimate money demand functions based on aggregate time series data, as only the use of micro data allows to meaningfully capture the non-linearity associated with the cost of participating in financial markets.

Figure 3 shows evidence on the possible presence of non-linearities for eight countries for which both of the sub-samples with the short rate above and below 5 per cent are sufficiently long.[29] The first row of either figure shows the short rate together velocity, which has been rescaled in such a way as to have the same sample mean, and the same sample standard deviation as the short rate. The bottom row shows the low-frequency components of the series shown in the corresponding panels in the top rows. The components have been extracted exactly as in Section 3.

The evidence in the figures provides *no support* to the notion that velocity—and therefore money demand—may be less responsive to interest rate changes at low interest rates. On the contrary, by no means the relationship between the two series appears to be different at high and low interest rates. The evidence based on the low-

[28] This is why Table SELA.3 does not report results for the United States based on the Lucas-Nicolini aggregate.

[29] 5 per cent was the threshold considered by Mulligan and Sala-i-Martin (2000).

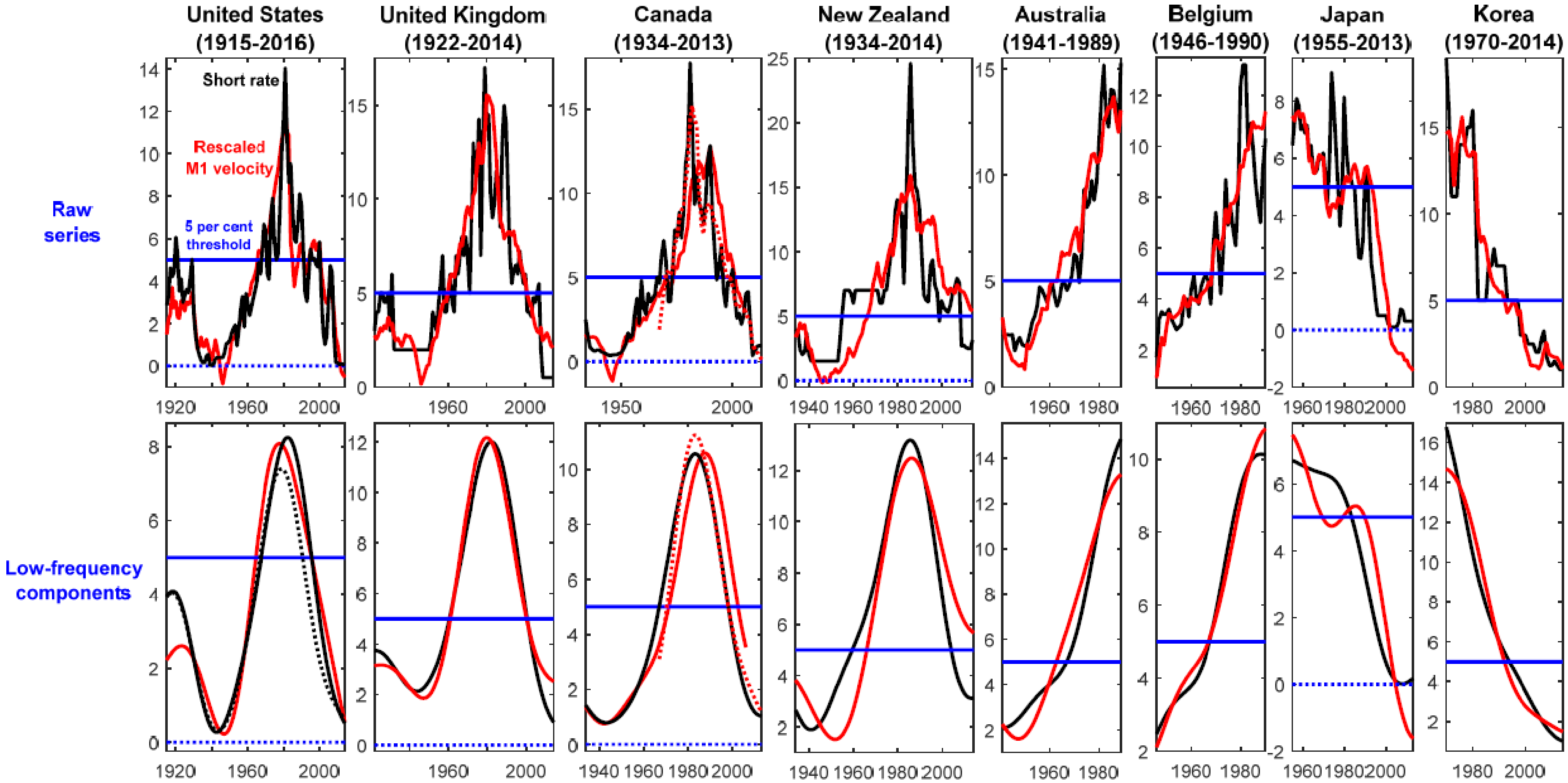

Figure 3 Informal evidence on the possible non-linearity of M_1 velocity at low interest rates

frequency components of the data is especially stark: Once stripping velocity and the short rate of higher-frequency fluctuations, the relationship between the series clearly appears as remarkably strong and stable at all levels of the interest rate. In particular, taking, just for the sake of the argument, 5 per cent as a 'reference threshold' for the short rate (see the previous footnote), the following should be noted:

(1) for the U.S., the U.K., and Canada, the two periods before the mid-1960s, and since the end of the 1990s—during either of which the short rate had, or has been, below 5 per cent—appear as remarkably similar to the period in-between, during which the short rate systematically exceeded 5 per cent. In no way do these data suggest that at low interest rates the relationship between velocity and the short rate is any different from what it is at higher rates.

(2) Qualitatively similar evidence holds for Australia and Belgium, for which the relationship between the series appears as the same either before or after the 1960s, and for Korea, for which the period since the 1990s appears as very similar to previous years.

(3) For Japan the relationship between the series appears to have broken down since the beginning of the XXI century. On the other hand, it is worth stressing that during the period between the mid-1990s and the beginning of the new century, when the short rate plummeted from about 5 per cent to about 0, velocity likewise collapsed with a lag of a few years. Further, it is to be noticed that since the start of the new millennium velocity has kept decreasing, whereas the interest rate has remained close to zero, a pattern which is the *opposite* of that implied by the presence of fixed costs associated with the decision to participate, or not to participate, in financial markets.

Figure 4 buttresses the visual evidence in Figure 3 with statistical evidence for five countries for which we could find sufficiently long post-WWII quarterly sample, in order to get reasonably precise estimates. In either case, we estimated Selden-Latané specifications for velocity, i.e.

$$V_t \equiv \frac{Y_t}{M_t} = C + \delta R_t \tag{8}$$

via Stock and Watson's (1993) estimator of the cointegration vector,[30] based on consecutive sample periods during which the short rate had consistently been above 4.5 per cent,[31] characterizing uncertainty by bootstrapping the VECM estimated conditional on one cointegration vector as in CRT. For either country we show the

[30] We set K, the number of leads and lags used for the dynamic OLS estimator, to one. On the other hand, we eschewed Johansen's procedure for the following reason. In order to be able extract the money demand from the estimated Selden-Latané specification, we need not only the slope coefficient δ in (8)—which Johansen's procedure does indeed produce—but also the intercept C. Johansen's procedure, however, produces an estimate of the intercept of the *VECM representation* of (8), rather than of (8) itself, and from this it is impossible to recover C.

[31] E.g., for the United States the sample period is 1972Q4-1991Q3 (details for other countries are available upon request). We take 4.5 per cent as the threshold uniquely for practical reasons, as it allows us to use materially longer samples for estimation.

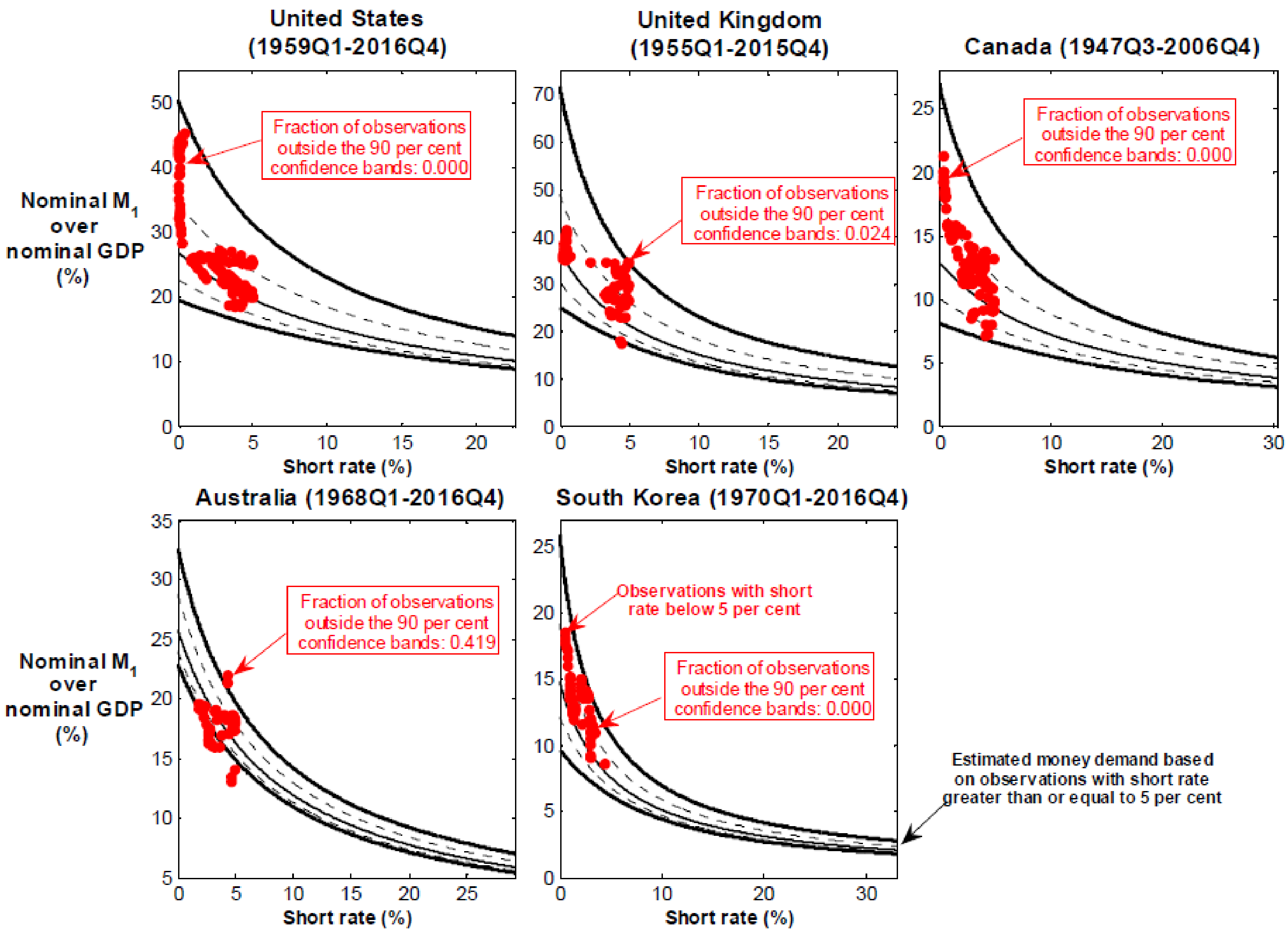

Figure 4 Statistical evidence on the possible presence of non-linearities in the demand for M_1 at low interest rates

estimated implied money demand curves, with 16-84 and 5-95 per cent confidence bands, and (in red) the observations corresponding to a short rate below 5 per cent. Under the null hypothesis that the model is the same at high and low interest rates, low-interest-rate observations should fall outside of the 5-95 per cent bands 10 per cent of the times. In fact, this *never* happens for the U.S., Canada, and South Korea, and it happens 2.4 per cent of the times for the U.K.. For Australia the fraction, at 41.9 per cent, is much higher, but it is to be noticed that, for the purposes of Mulligan and Sala-i-Martin's (2000) argument, the outliers are on the *wrong* side of the demand curve: Rather than being above the curve, most of them are below.

This evidence questions the notion that there might be sizeable non-linearities in money demand at low interest rates, and it rather suggests that the behaviour of M1 velocity—and therefore the demand for M1—is essentially the same at *all* interest rate levels (at least, for the range of interest rates experienced by the countries in our sample). In turn, this suggests that it should indeed be possible to reliably estimate the welfare costs of inflation associated with the mechanism first highlighted by Bailey (1956) based on aggregate time series data.

An important question then is how to rationalize the finding of a smaller elasticity at low interest rates. In the next section we provide a possible explanation.

7 Pitfalls of Using the Incorrect Functional Form

Is the previous discussion of alternative functional forms an ultimately sterile exercise, or do alternative specifications have materially different implications for issues of central importance? In this section we show, by example, that in several cases identifying the correct functional form does indeed have material implications.

7.1 Spurious non-linearity of money demand from estimating log-log specifications

Suppose that the data have been generated by a Selden-Latané specification, so that the relationship between the levels of velocity and the short rate is identical at all levels of the short rate. Since a given percentage change in the *level* of the short rate (say, 1 per cent) is associated with a larger change in its *logarithm* at low interest rates than it is at higher interest rates,[32] this automatically maps into lower estimated elasticities (in absolute value) at low interest rates than at higher interest rates. This implies that if the true specification is the Selden-Latané, estimating a log-log will automatically produce smaller elasticities (in absolute value) at lower than at higher interest rates. The same argument obviously holds if the true specification is the semi-log.

[32] E.g., ln(9)-ln(10)=-0.105, whereas ln(2)-ln(3)=-0.406.

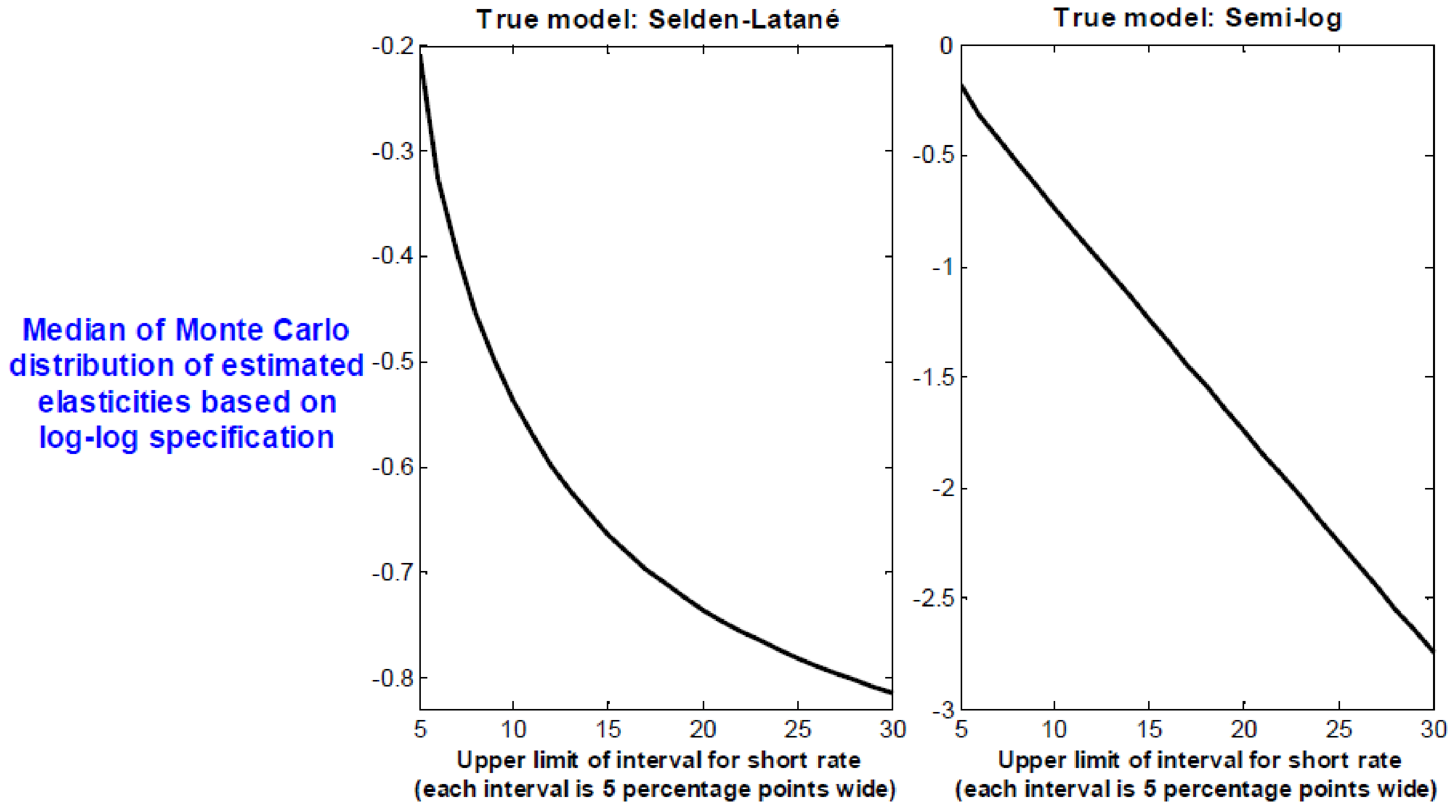

Figure 5 Monte Carlo evidence on the distortions originating from estimating a log-log specification when the true model is either Selden-Latané or semi-log

Figure 5 provides a simple illustration of this based on the following Monte Carlo experiment. We model the short rate, R_t, as a random walk with reflecting barriers with $N(0, 1)$ innovations.[33] The reflecting barriers are designed to keep the short rate within a specific bounded interval—thus allowing us to explore how changes in such interval are going to impact upon the estimated elasticity in log-log specifications—and are enforced *via* simple rejection, as in, e.g., Cogley and Sargent (2002). Conditional on R_t, we then generate artificial samples for velocity based on either the Selden-Latané specification (8), or the semi-log one

$$\ln V_t = B + \xi R_t \tag{9}$$

In (8) we set C to 2.5 (corresponding to a satiation level of 40 per cent of GDP) and δ to 0.4, which are very close to estimates for the United States based on Lucas and Nicolini's aggregate. By the same token, in (9) we set B to 1, and ξ to 0.1, which are, once again, near-numerically identical to the U.S. estimates.

We use artificial samples of length T=10,000, and we set the number of Monte Carlo replications to 1,000. We consider the following set of bounded intervals for the short rate: [0, 5], [1, 6], [2, 7], ..., [25, 30]. For each interval, and each replication, we estimate a log-log specification, and we store the elasticity. Figure 5 plots, for each lower bound of the corresponding bounded interval, the median of the Monte Carlo distribution of the estimated elasticities: In spite of the fact that the data-generation process is *the same* at *all levels* of the short rate, the lower the short rate, the smaller (in absolute value) *estimated* elasticities are. This implies that if the data have truly been generated by either a Selden-Latané or a semi-log specification, the popular practice of estimating log-log specifications *automatically* produces spurious evidence of non-linearity of money demand at low interest rates.

To be sure, this does not mean that evidence of non-linearity in the literature is spurious: What it means, however, is that, by estimating log-log specifications, a researcher would obtain these results even if the true model were of the Selden-Latané or semi-log type. This provides a first illustration of the importance of understanding what the true specification is. We now turn to a second example.

7.2 The welfare costs of inflation

Since Bailey (1956), the welfare costs of inflation have been a classic topic in monetary economics. As a second illustration of the pitfalls of using the incorrect functional form for money demand, we now show that, for several countries in our dataset, the three specifications analyzed herein produce materially different estimates of inflation's welfare costs. We estimate either of the three specifications for velocity based on Stock

[33] The standard deviation of reduced-form innovations to the U.S. short rate produced by the Selden-Latané specification is equal to 0.97.

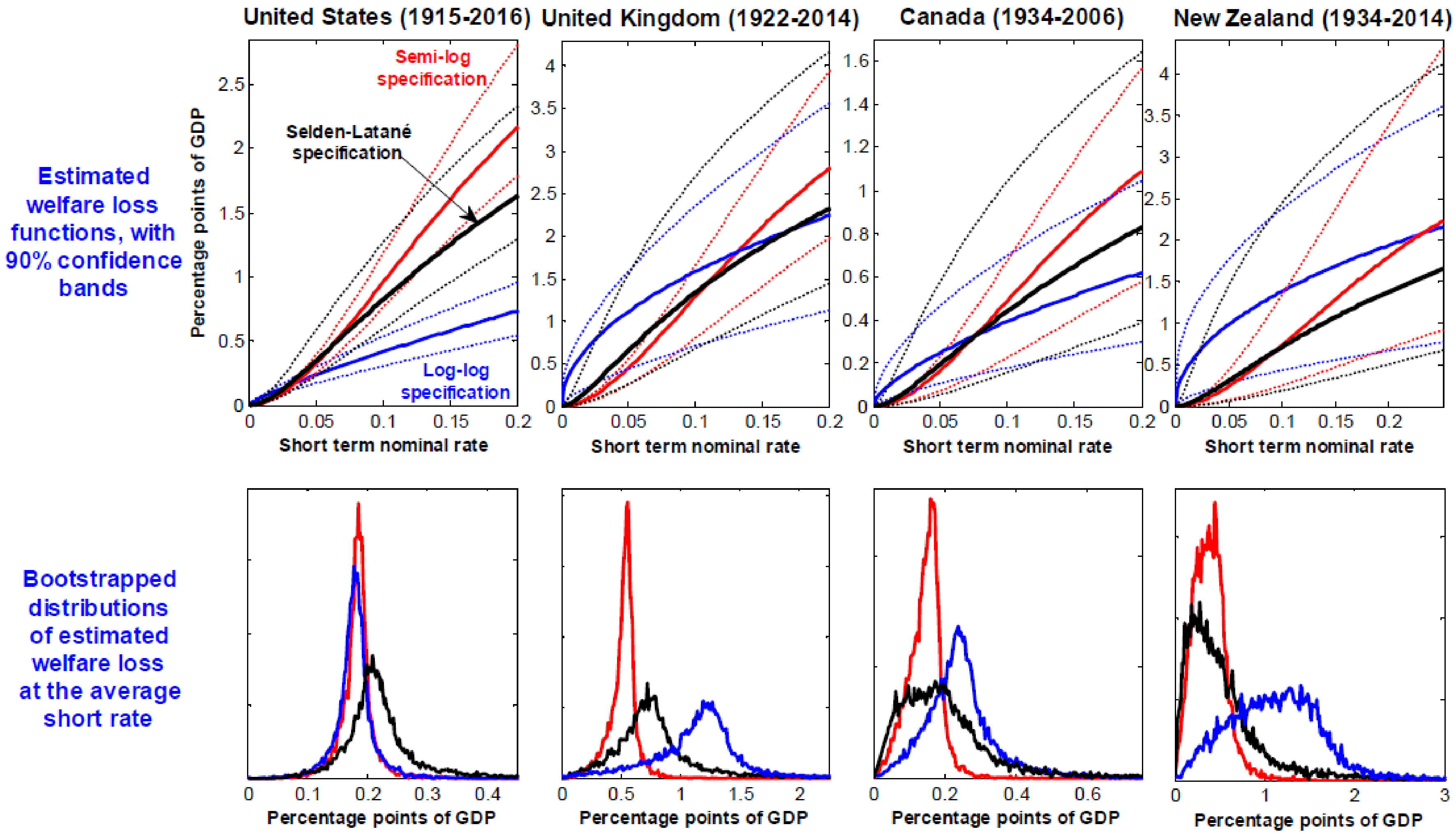

Figure 6 Implications of using alternative functional forms for the welfare costs of inflation

and Watson's (1993) estimator of the cointegration vector.[34] Based on the implied money demand functions, the welfare costs of inflation can then be immediately recovered along the lines of Bailey (1956), Friedman (1969) and Lucas (2000). We characterize the uncertainty surrounding the point estimates by estimating the VECM conditional on Stock and Watson's (1993) estimate of the cointegration vector,[35] and bootstraping it as in CRT (2012). Results are based on 10,000 bootstrap replications of the estimated VECM.

Lucas (2000, p. 251) provided the expressions for the welfare cost associated with a specific level of the interest rate R_t for the semi-log and log-log specifications. By the same token, it can be trivially shown that the welfare cost function associated with the Selden-Latané specification (8) is given by

$$WC_{SELA}(R_t) = \frac{\ln(C + \delta R_t) - \ln C}{\delta} - \frac{R_t}{C + \delta R_t} \tag{10}$$

Figure 6 reports, for four of the countries in our dataset, the estimated welfare loss functions, together with the implied welfare losses at the *average* short rate which has prevailed over the sample period (expressed in percentage points of GDP).[36] We will not comment upon the figures in detail because they speak for themselves. For the present purposes, what ought to be stressed is that the three specifications imply materially different estimates of the welfare costs. Focusing on the comparison between the popular log-log specification, and the Selden-Latané one, for the United Kingdom the welfare costs at the average short rate implied by the former are about *twice* those implied by the latter. Specifically, based on Selden-Latané and log-log, the median and 90 per cent confidence intervals are 0.75 [0.31; 1.72] and 1.43 [0.72; 2.18], respectively. The same holds for New Zealand, whereas the difference is less marked for Canada. For the United States[37] the three specifications tend to produce similar results, with median estimates equal to 0.21 per cent for the Selden-Latané specification, and 0.18 for the other two. Results for several other countries (not reported for reasons of space, but available upon request) are in line with those in Figure 6: Just to mention two cases, for Switzerland the median estimates based on

[34] We set K, the number of leads and lags used for the dynamic OLS estimator, to one. On the other hand, we eschew Johansen's estimator for the following reason. Focusing, e.g., on the Selden-Latané specification (8) (the same argument holds for the other specifications), in order to be able to compute the welfare costs of inflation we need not only the slope coefficient δ, which Johansen's procedure does indeed produce, but also the intercept C. Johansen's procedure, however, produces an estimate of the intercept of the VECM representation of (8), rather than of (8) itself, from which it is impossible to recover C.

[35] So, to be clear, we follow the two-step procedure discussed in Luetkepohl (1991, pp. 370-371). This approach is valid in the case in which there is just one cointegration vector, which is the case here.

[36] The differences across countries in the estimates reported in the second row therefore also reflect differences in average short rates, which for the U.S., U.K., Canada, and New Zealand were equal to 3.6, 5.6, 4.7, and 6.6 per cent, respectively.

[37] Estimates are based on Lucas and Nicolini's (2015) aggregate.

Selden-Latané and log-log are 0.22 and 0.53 per cent of GDP, respectively, whereas the corresponding figures for South Korea are 0.63 and 1.37 per cent.

These simple examples illustrate the importance of correctly identifying and using the right functional form for money demand. In particular, in several cases the Selden-Latané specification consistently delivers results which are materially different from those produced by the popular log-log form.

8 Conclusions

We use a simple model of a transaction demand for money to guide a thorough investigation of the stability of the long run relationship between the ratio of money to output and a short term nominal interest rate. Our data set comprises 32 countries for periods that range from 35 to 100 years. We report six main findings: (1) Evidence of cointegration between velocity and the short rate is widespread. For the U.S. we detect strong evidence based on several of the adjustments to the standard M1 aggregate originally proposed by Goldfeld and Sichel (1990). (2) Evidence of breaks or time-variation in cointegration relationships is weak to non-existent. (3) For several low-inflation countries the data prefer the specification in the levels of velocity and the short rate originally estimated by Selden (1956) and Latané (1960). This is the case in particular for the U.S., based on either of the adjustments to M1 proposed by Goldfeld and Sichel. (4) There is no evidence of non-linearities at low interest rates. (5) Using the correct money demand specification has important implications for the ability to correctly estimate the welfare costs of inflation. (6) If the data are generated by a Selden-Latané specification, estimation of a log-log specification spuriously causes estimated elasticities to be smaller at low interest rates, as found e.g. by Mulligan and Sala-i-Martin (2000).

References

Alvarez, F., and F. Lippi (2009): "Financial Innovation and the Transactions Demand for Cash," *Econometrica*, 77(2), 363–402.

Bailey, M. J. (1956): "The Welfare Cost of Inflationary Finance," *Journal of Political Economy*, 64(2), 93–110.

Barro, R. J. (1982): "United States Inflation and the Choice of Monetary Standard," *in R. E. Hall, ed., 'Inflation: Causes and Effects', University of Chicago Press.*

Barsky, R. (1987): "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19(1), 3–24.

Benati, L. (2008): “Investigating Inflation Persistence Across Monetary Regimes,” *Quarterly Journal of Economics*, 123(3), 1005–1060.

——— (2015): “The Long-Run Phillips Curve: A Structural VAR Investigation,” *Journal of Monetary Economics*, 76(November), 15–28.

——— (2017): “Money Velocity and the Natural Rate of Interest,” University of Bern, mimeo.

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2012): “Bootstrap Determination of the Cointegration Rank in Vector Autoregressive Models,” *Econometrica*, 80(4), 1721–1740.

Christiano, L. J., and T. J. Fitzgerald (2003): “The Bandpass Filter,” *International Economic Review*, 44(2), 435–465.

Cogley, T., and T. J. Sargent (2002): “Evolving Post-WWII U.S. Inflation Dynamics,” *in B. Bernanke and K. Rogoff, eds. (2002), NBER Macroeconomics Annuals 2001.*

Diebold, F. X., and C. Chen (1996): “Testing Structural Stability with Endogenous Breakpoint: A Size Comparison of Analytic and Bootstrap Procedures,” *Journal of Econometrics*, 70(1), 221–241.

Friedman, B. M., and K. N. Kuttner (1992): “Money, Income, Prices, and Interest Rates,” *American Economic Review*, 82(3), 472–492.

Friedman, M. (1969): “The Optimum Quantity of Money,” *in Milton Friedman, The Optimum Quantity of Money and Other Essays, 1Ű50. Chicago: Aldine Publishing Company.*

Goldfeld, S. M., and D. M. Sichel (1990): “The Demand for Money,” *in Friedman, B.M., and Hahn, F.H., eds., Handbook of Monetary Economics, Vol. I, Amsterdam, North Holland.*

Grossman, S., and L. Weiss (1983): “A Transactions-Based Model of the Monetary Transmission Mechanism,” *American Economic Review*, 73(5), 871–880.

Hansen, B. E. (1999): “The Grid Bootstrap and the Autoregressive Model,” *Review of Economics and Statistics*, 81(4), 594–607.

Hansen, H., and S. Johansen (1999): “Some Tests for Parameter Constancy in Cointegrated VAR Models,” *Econometrics Journal*, 2, 306–333.

Judson, R. (2017): “The Death of Cash? Not So Fast: Demand for U.S. Currency at Home and Abroad, 1990-2016,” Federal Reserve Board, Division of International Finance Working Paper, forthcoming.

Lucas Jr., R. E. (1980): "Two Illustrations of the Quantity Theory of Money," *American Economic Review*, 70(5), 1005–1014.

——— (1988): "Money Demand in the United States: A Quantitative Review," *Carnegie-Rochester Conference Series on Public Policy*, 29, 137–168.

——— (2000): "Inflation and Welfare," *Econometrica*, 68(2), 247–274.

Lucas Jr., R. E., and J.-P. Nicolini (2015): "On the Stability of Money Demand," *Journal of Monetary Economics*, 73, 48–65.

Luetkepohl, H. (1991): *Introduction to Multiple Time Series Analysis, 2nd edition.* Springer-Verlag.

Meltzer, A. H. (1963): "The Demand for Money: The Evidence from the Time Series," *Journal of Political Economy*, 71(3), 219–246.

Mulligan, C. B., and X. SalaiMartin (2000): "Extensive Margins and the Demand for Money at Low Interest Rates," *Journal of Political Economy*, 108(5), 961–991.

Rotemberg, J. J. (1984): "A Monetary Equilibrium Model with Transactions Costs," *Journal of Political Economy*, 92(1), 40–58.

Selden, R. T. (1956): "Monetary Velocity in the United States," *in M. Friedman, ed., Studies in the Quantity Theory of Money, University of Chicago Press*, pp. 405–454.

Stock, J. H., and M. W. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61(4), 783–820.