UNIVERSITÄT
BERN

Faculty of Business, Economics
and Social Sciences

Department of Economics

Long-Run Money Demand Redux

Luca Benati

18-04

January 2018

DISCUSSION PAPERS

Schanzeneckstrasse 1
Postfach
CH-3001 Bern, Switzerland
http://www.vwi.unibe.ch

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. (2) Evidence of breaks or time-variation in cointegration relationships is weak to nonexistent. (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 nonlinearities 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 Giuseppe Cavaliere, Jesus Fernandez-Villaverde, Helmut Luetkepohl, Albert Marcet, Edward Nelson, Pedro Teles, as well as seminar participants at SED 2017, Minneapolis FED, Oxford, Bonn, Bern, Banca d'Italia, Norges Bank, and ECB for useful discussions or comments. Special thanks to Joao Ayres, Nicole Gorton, Eva Werner, and Junji Yano for invaluable help with the data, and to Ruth Judson for help in adjusting U.S. currency for the fraction held by foreigners. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

[†]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

This paper describes and analyzes a new dataset containing annual measurements of money supplies, both real and nominal output (GDP), and thus price levels, and short term nominal interest rates for 32 countries ranging from 1851 to 2016. The framework we use for organizing these data is a money demand function that relates the money that the public and private sectors of the economy choose to hold to the rate of production of goods and the short-term interest rate

$$M_t = P_t Y_t \phi(r_t) \quad (1)$$

where ϕ is a decreasing function of r_t. Think of an individual, or a business, or a government agency choosing how much cash to hold if it expects a spending flow per unit of time of $P_t Y_t$, where r_t is the opportunity cost of holding low- or no-interest bank deposits instead of equally risky assets with higher returns: a unit-elastic income effect, and a price effect. For the individual, then, (1) is a demand function, a description of a decision problem, a choice, that every agent in the economy must make. For our purposes, we need to think of (1) as an equilibrium condition for the economy as a whole.

The formula (1) contains some strong implications. One is that, if r_t is stationary, M_t and $P_t Y_t$ should grow at common rate, for any function ϕ. If, on the other hand, r_t has a unit root—possibly, because inflation is driven in part by permanent shocks—M_t and $P_t Y_t$ should grow at a common rate once controlling for the impact of permanent shocks to r_t. A second implication is that it should be possible to use cross-country, and even within-country time-series to trace out a stable function ϕ. This is the agenda carried out in this paper.

In recent years, many economists and central bankers have come to doubt the usefulness of measures of the money supply (like M1) in the conduct of monetary policy. What was thought to be a central 'pillar' of the monetary policy strategy of the European Central Bank (ECB) has come to be seen as too unreliable to be of any use at all. These concerns were not without empirical basis: conventional time-series models of money demand could be unstable, especially at the high frequencies. Seeking 'long run' relationships seemed therefore the only feasible course of action. Our central idea, in the present work, is to use cointegration methods in order to precisely characterize what we mean by 'long run' relationships, and to apply them in a uniform way to a very wide variety of countries. We take pains to ensure that terms such as 'short rate' and 'M1' are measures of the same thing (or almost!) in different countries, and over time within countries. Then, we simply let the results produced by these methods speak for themselves. These findings need discussion, country by country, and we will provide it. But we will argue that the basic features of the demand function for money are in general very solid, maybe the most solid finding in the field of macroeconomics.

In his paper 'Long Run Evidence on Money Growth and Inflation', Luca Benati (2009) used a band-pass filter in order to illustrate how the low-frequency components

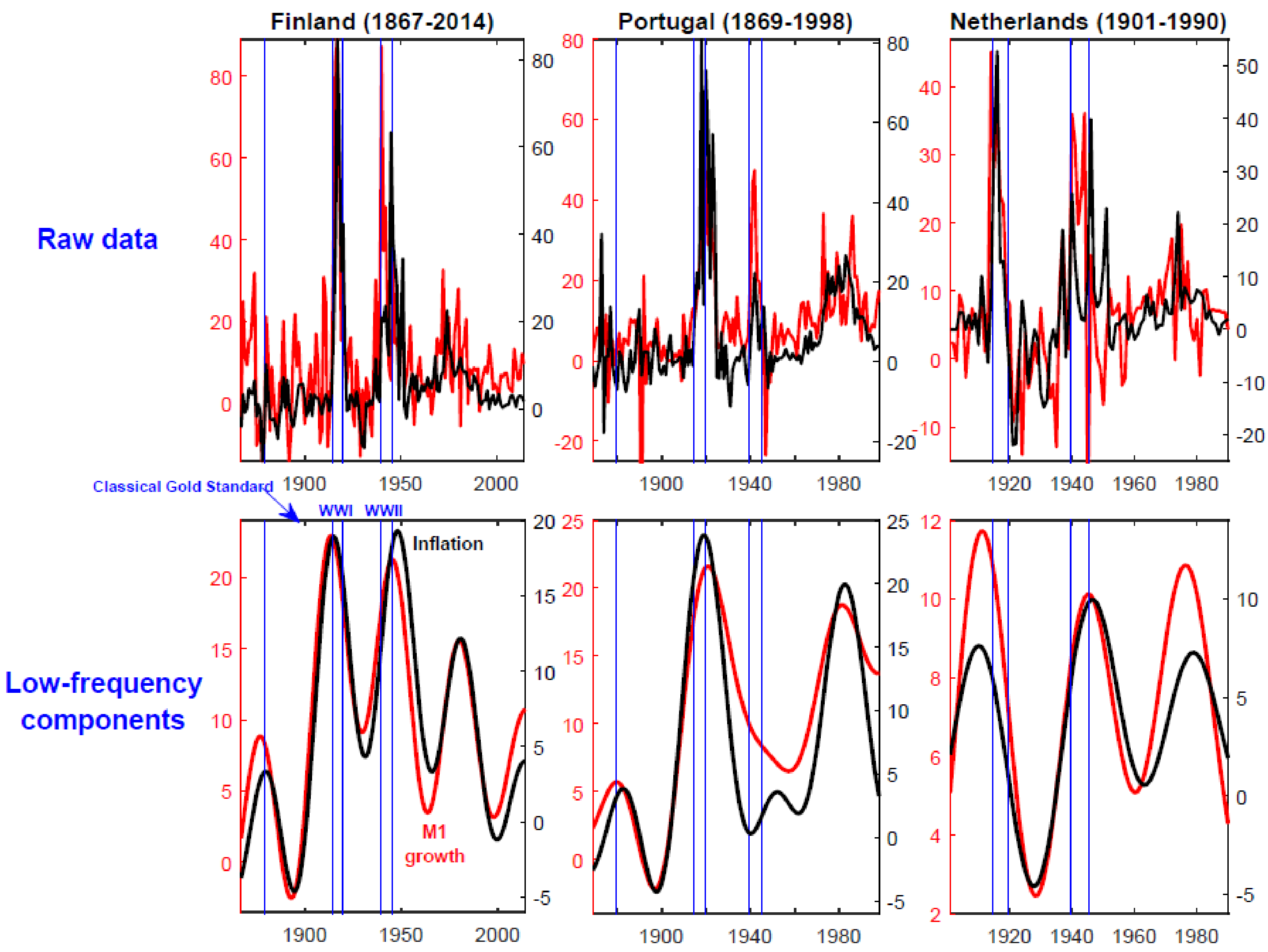

Figure 1 M1 growth and inflation since the Gold Standard era
(low-frequency components: cycles slower than 30 years)

of money growth and inflation exhibit, in most cases, a nearly one-for-one relationship. As an example, the top row of Figure 1 shows data on M1 growth and inflation for three European countries since the Gold Standard era, whereas the bottom row shows the components of the two series with cycles slower than 30 years.[1] That paper contains evidence for many other countries and definitions of money growth. There are differences, of course, but the long-run relationship between the two series is, in most cases, very clear.

The current paper has many features not contained in Benati (2009), but there are two that are central.

First, in that paper money growth and inflation were treated simply as given time series. In this paper, by contrast, we borrow from a vast post-World War II literature and take the interest rate as a given series—chosen by monetary policy—and assume that individual agents divide their work effort between producing goods, and economizing their holding of low-interest-bearing cash.[2] We derive an equation like (1), where the familiar Baumol-Tobin model is an example, but a few others are possibilities. We address these elements of consumers' decision problems in detail in Section 2. Then in Section 3 we plot the implied predictions of a particular case of the model against the data for all countries in our dataset, and we show low-frequency evidence, extracted *via* the band-pass filter. We find this informal visual evidence quite remarkable.

Second, as mentioned above, in this paper we make central use of cointegration methods, which replace the distinction between high and low frequencies used by Benati (2009). Section 4 describes the methodology, whereas Section 5 discusses our main findings. Evidence of cointegration between real money demand and a short-term interest rate is widespread, whereas evidence of breaks or time-variation in cointegration relationships is weak to nonexistent. 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* expect to obtain if cointegration were truly present in the data, because of the short sample length, the high persistence of the cointegration residual, or both.

Having provided what we believe is very robust evidence of a long-run money demand relationship, we use our dataset to revisit two issues that have been widely discussed in the literature. First, in Section 6, we study the behavior of real money demand at very low interest rates. It has been shown that model economies with heterogenous agents that face fixed costs of participating in the asset markets can lead to nonlinearities for low values of the nominal interest rates, which has implications regarding the proper way to estimate the welfare cost of inflation.[3] We find no evidence of nonlinearities in our dataset. In Section 7, we revisit the computation of the welfare

[1] The components have been extracted *via* the filter proposed by Christiano and Fitzgerald (2003).

[2] That literature was led by Milton Friedman, Anna J. Schwartz, Karl Brunner, Allan Meltzer, William J. Baumol, and James Tobin.

[3] See, e.g., Mulligan and Sala-i-Martin (2000).

cost of inflation in the tradition of Martin Neal Bailey (1956), Milton Friedman (1969), Robert E. Lucas, Jr. (2000) and Peter Ireland (2009). 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 the 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 exactly replicates 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 $\alpha P(s^t)x(s^t) \leq M(s^t)n(s^t)$, where α is a constant of proportionality.[4]

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 of this cost function. Specifically, we describe the total cost of making transactions, measured in units of time, by a nonnegative, 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 randomness is

[4] The model omits the use of cash by firms to pay employees and suppliers of intermediate goods and to clear asset exchanges. The parameter α implies that we are implicitly treating all these payments as proportional to final goods payments.

essential to motivate the econometric analysis at 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}. If we 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

$$m(s^t) + b(s^t) + E\left[q(s^t, s_{t+1})\pi(s^t, s_{t+1})\widetilde{P}^Q(s^t, s_{t+1})\right] \leq \omega(s^t), \tag{2}$$

where lowercase 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 state 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.[5] 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 online Appendix B.1.[6]

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 monetary aggregate and its return in detail later on, in our discussion of the empirical analysis.

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}, is given by

$$\begin{aligned} \omega(s^t, s_{t+1}) \leq \; & \frac{m(s^t)R^m(s^t) + b(s^t)R(s^t)}{\pi(s^t, s_{t+1})} + q(s^t, s_{t+1}) \\ & + \frac{[1 - \theta(n(s^t), \nu(s^t))]\, z(s^t) - x(s^t)}{\pi(s^t, s_{t+1})} + \tau(s^t, s_{t+1}), \end{aligned} \tag{3a}$$

[5] 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.

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

where $\tau(s^t, s_{t+1})$ is the real value of the monetary transfer the government makes to the representative agent. Finally, the cash-in-advance constraint can be written in real terms as

$$\alpha x(s^t) \leq m(s^t) n(s^t). \tag{4}$$

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 $x\left(s^t\right)$, the number of bank trips $n\left(s^t\right)$, and the assets $m\left(s^t\right)$, $b\left(s^t\right)$, 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 (2), (3a), and (4).

The Bellman equation describing the decision problem is

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

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

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

$$r^* \equiv (R - R^m) = n^2 \frac{\alpha\theta_n(n)}{1-\theta(n)}, \tag{5}$$

which gives an extended squared root formula for the equilibrium value of n.[7]

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

$$\frac{m\left(r_t, \nu_t\right)}{x\left(r_t^*, \nu_t\right)} = \frac{\alpha}{n\left(r_t^*, \nu_t\right)},$$

which does not depend on z_t.

We now discuss several empirical implications of this solution that do not depend on the particular functional form assumed for the function $\theta(n)$. First, the theory implies an income elasticity equal to one. This is the specification we will study for much of the paper. In online Appendix G, we allow for a more general specification that 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 $m\left(r_t^*, \nu_t\right)$. In particular, up to a linear approximation the stochastic properties of

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

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, m_t/x_t should be too, whereas if r_t^* has a unit root, m_t/x_t should have a unit root as well. 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. But before discussing specific cases, notice that as long as $\theta_n(n) > 0$ for all n, then $n \to 0$ as $r^* \to 0$. This means that there is no satiation point for real balances as its opportunity cost goes to zero. This property of the BT specification therefore extends to any cost function that is strictly increasing with the number of portfolio adjustments. As we show below, however, the data strongly prefer specifications with finite real money balances when $r \to 0$. This is where our assumption of $R^m < 0$ becomes key in the three examples below, since it implies that $r^* > 0$ even when $R \to 1$.[8]

The exponential case Consider first the function $\theta(n_t, v_t) = \gamma\nu_t n_t^{\sigma}$. In this case, equation (5) 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 barely above 4%, so the denominator in the expression above would 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

$$\ln \sigma\gamma + \ln \nu_t + (\sigma + 1) \ln n_t = \ln r_t^*, \tag{6}$$

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.

[8] A very interesting alternative that fits the micro data very well is proposed in Alvarez and Lippi (2009). In that model, agents get free portfolio adjustments randomly, which could potentially be interpreted as a sections of the function $\theta(n)$ that is flat in the aggregate.

The Selden-Latané specification A less well-known 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.

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

$$n_t b - \nu_t \simeq r_t^*, \tag{7}$$

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 U.S.—the data seem to quite clearly prefer it over the traditional log-log specification discussed above and the semi-log specification that we discuss next.

The semi-log specification 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 on 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 is the one before. The main difference between this function and the two studied above is that it asymptotes a constant (the term in parentheses on the right-hand side) as the number of trips grows arbitrarily large.[9] The solution is given by

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

[9] This implies that in this case, the welfare cost of inflation will be bounded above.

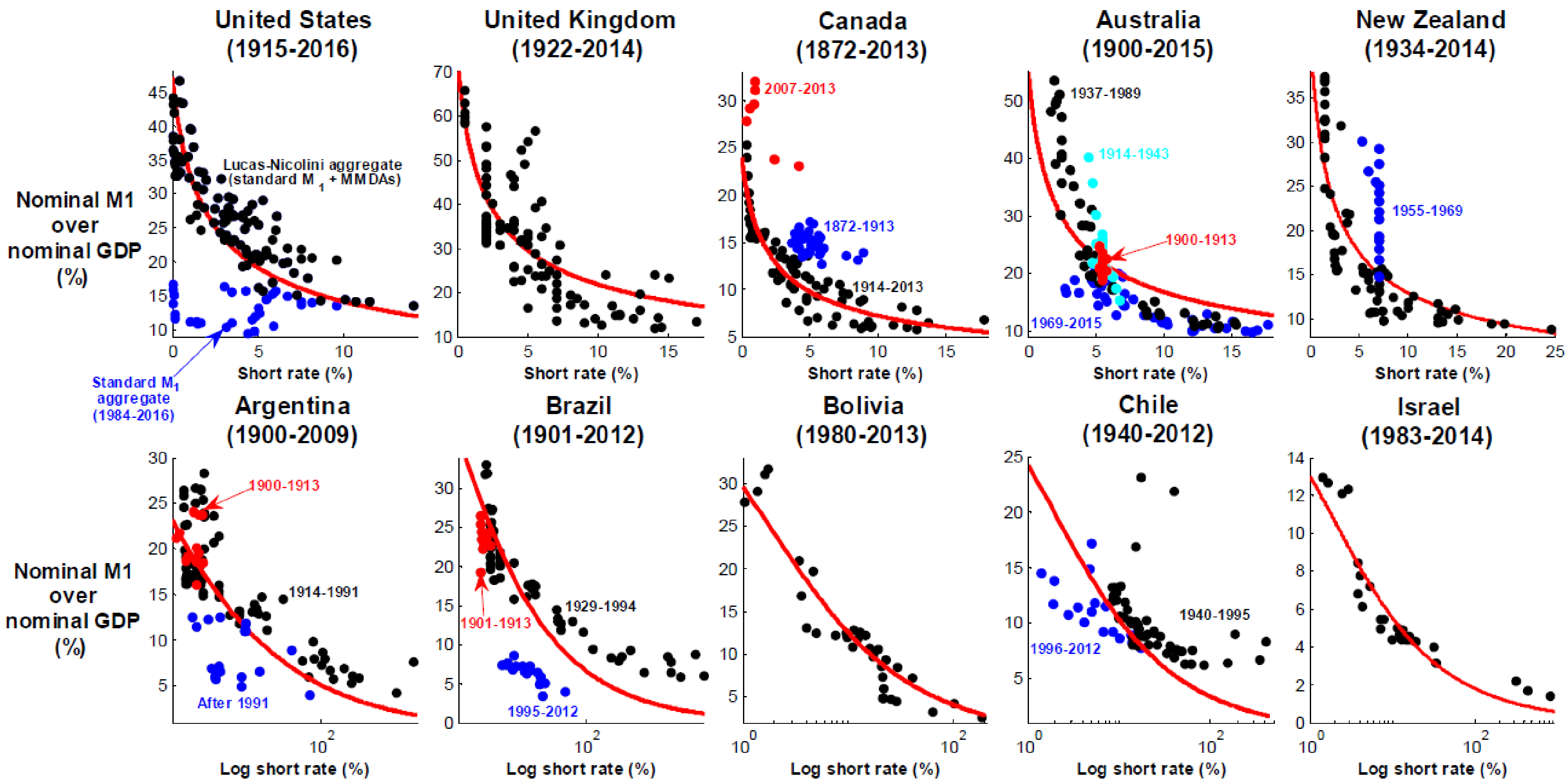

Figure 2a The raw data: short rate, ratio between nominal M1 and nominal GDP, and calibrated Baumol-Tobin specification

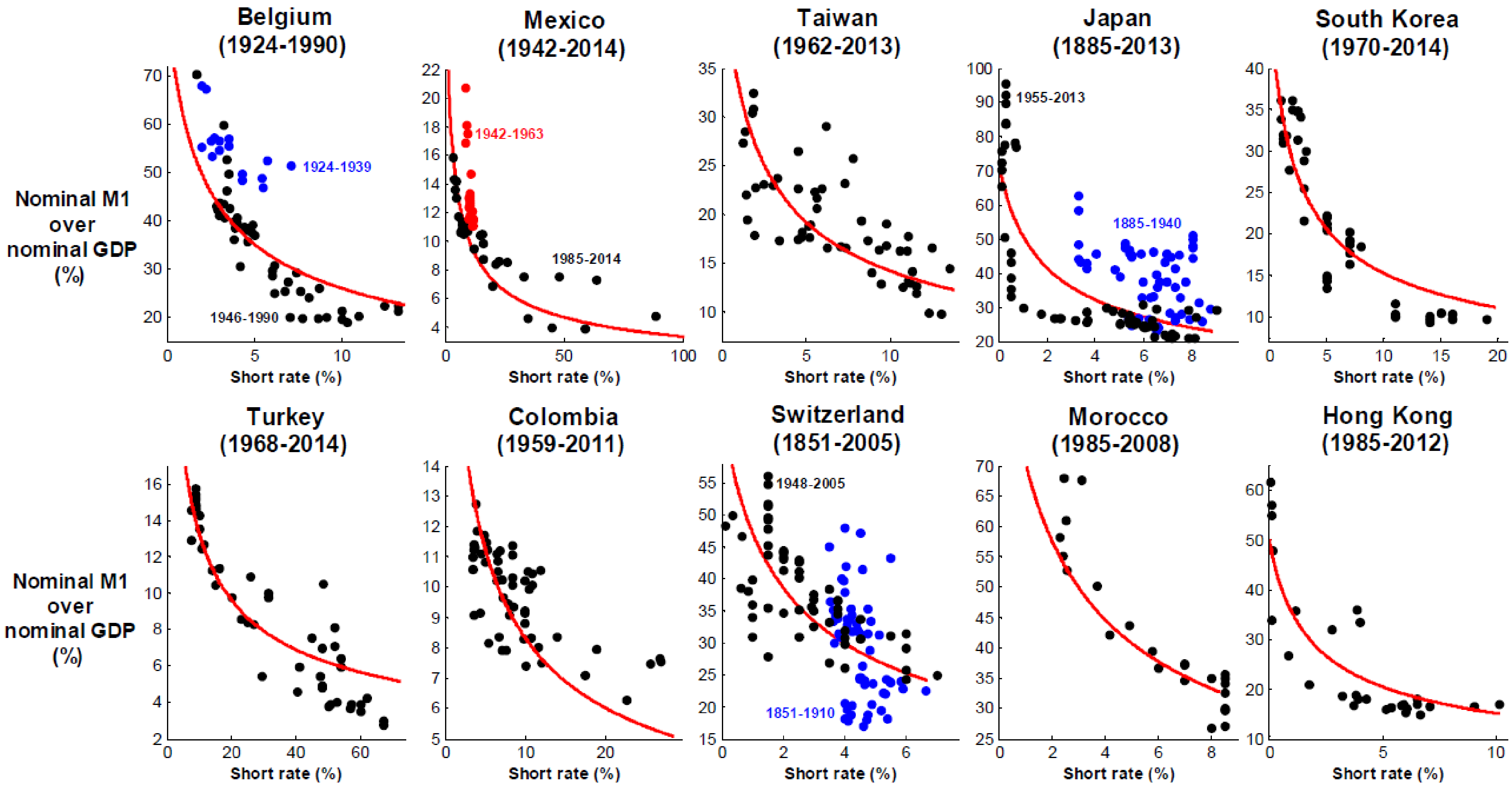

Figure 2b The raw data: short rate, ratio between nominal M1 and nominal GDP, and calibrated Baumol-Tobin specification

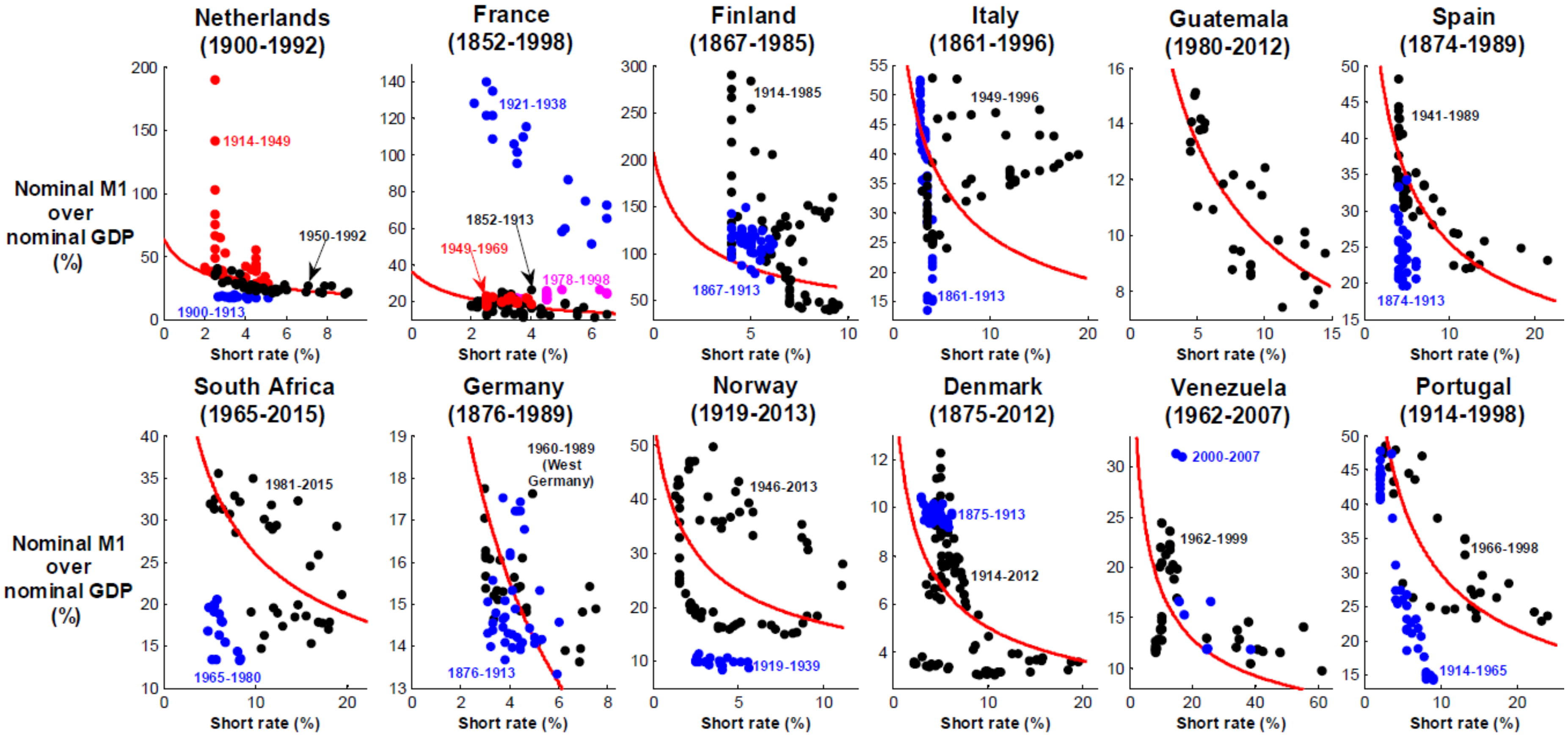

Figure 2c The raw data: short rate, ratio between nominal M1 and nominal GDP, and calibrated Baumol-Tobin specification

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. Note, however, that the condition for the cost function to be incresing implies that ε cannot be too low. Thus, the approximation in this case is less precise that in the previoustwo cases, since we cannot obtainan apression that is arbitrarely close to the semi-log specification.[10]

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 is associated with 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.[11] 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) 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 $R^m = 0.99$. As discussed above, if we assume that $R^m = 1$, the log-log curve goes to infinity as $R \to 1$. As can be seen in the evidence presented in this section, this does not seem to be the case for countries that did experience several periods of almost zero interest rates, such as the U.S. and Japan.

Online Appendix A describes the data and the data sources in detail. All of the series are standard, with the single exception of the U.S., where we consider three of the alternative adjustments to the Federal Reserve's standard M1 aggregate. These adjustments were originally suggested by Goldfeld and Sichel (1990, pp. 314-315)

[10] A detailed analysis of the behavior of this cost function for low values of ε is available upon request.

[11] For instance, deposits did pay interest in the United States after Regulation Q was modified in the early 1980s. Also, some deposits included in M1 did pay interest in very high-inflation countries such as Argentina and Brazil.

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 (*i*) money market deposit accounts (MMDAs), as in Lucas and Nicolini (2015);[12] (*ii*) Money Market Mutual Funds (MMMFs); or (*iii*) both MMDAs and MMMFs. Finally, for reasons of robustness, for any 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) to take into account the sizeable expansion in the fraction of U.S. currency held by foreigners since the early 1990s.[13] So, for the U.S., we essentially consider six alternative M1 aggregates. As we discuss below, regardless of whether we adjust for the fraction of U.S. currency held by foreigners, the results do not change materially, since the currency component of M1 is ultimately quite small compared to the deposits component.

Figures 2*a* to 2*c* are scatterplots of the short rate and the ratio between nominal M1 and nominal GDP (i.e., the inverse of M1 velocity), together with the theoretical curve that corresponds to an approximation of equation (6), namely, the BT case, so

$$\frac{M_t^j}{Y_t^j} = \frac{\alpha^j}{(r_t^j + 1)^{1/2}}, \tag{8}$$

where Y_t^j is nominal income at time t in country j and we let the constant of proportionality α^j to be country-specific. Therefore, one way to see our descriptive exercise is as using one free parameter per country in order to allow for a country-specific intercept, whereas the slope will be given by the BT assumption of a linear technology, so that the elasticity is calibrated to 0.5. 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,[14] 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 2*a* contains countries that belonged to the Commonwealth at some point. The second row contains countries that experienced very high inflation rates. The interest rate axis (i.e., the horizontal axis) is in a logarithmic scale, because of the magnitudes reached by inflation and interest rates in these countries. The second row of Figure 2*a* contains two countries, Argentina and Brazil, for which we highlight the most recent period (since 1991 and 1995, respectively). These are the two countries in our sample that experienced recurrent periods of very high inflation that lasted over a decade. The blue squares correspond to the periods that followed the successful stabilization years: 1991 for Argentina and 1995 for Brazil. These points are highlighted because

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

[13] The way the adjustment is performed is described in detail in online Appendix A.

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

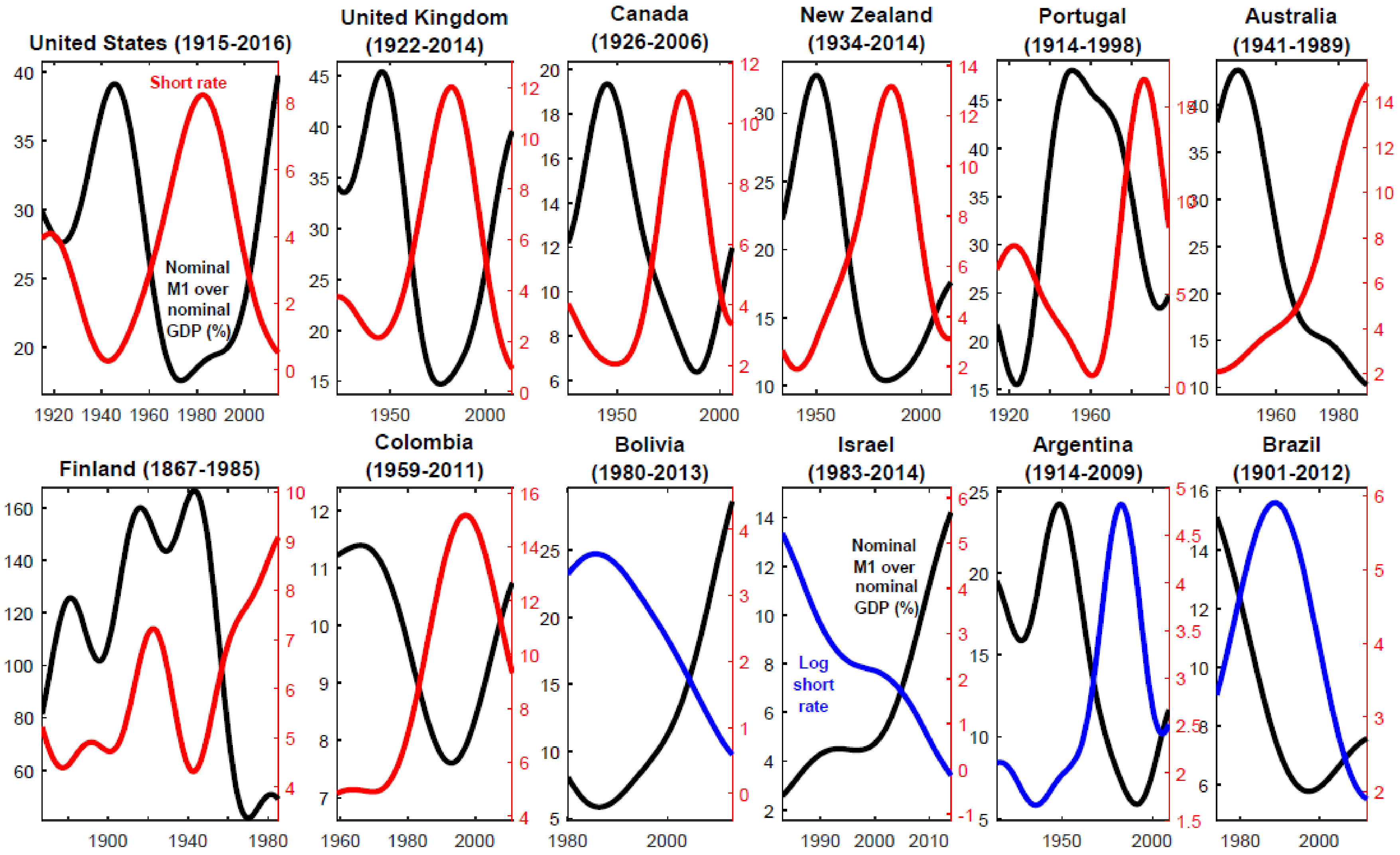

Figure 3a Low-frequency components of short rate and ratio between nominal M1 and nominal GDP for selected countries

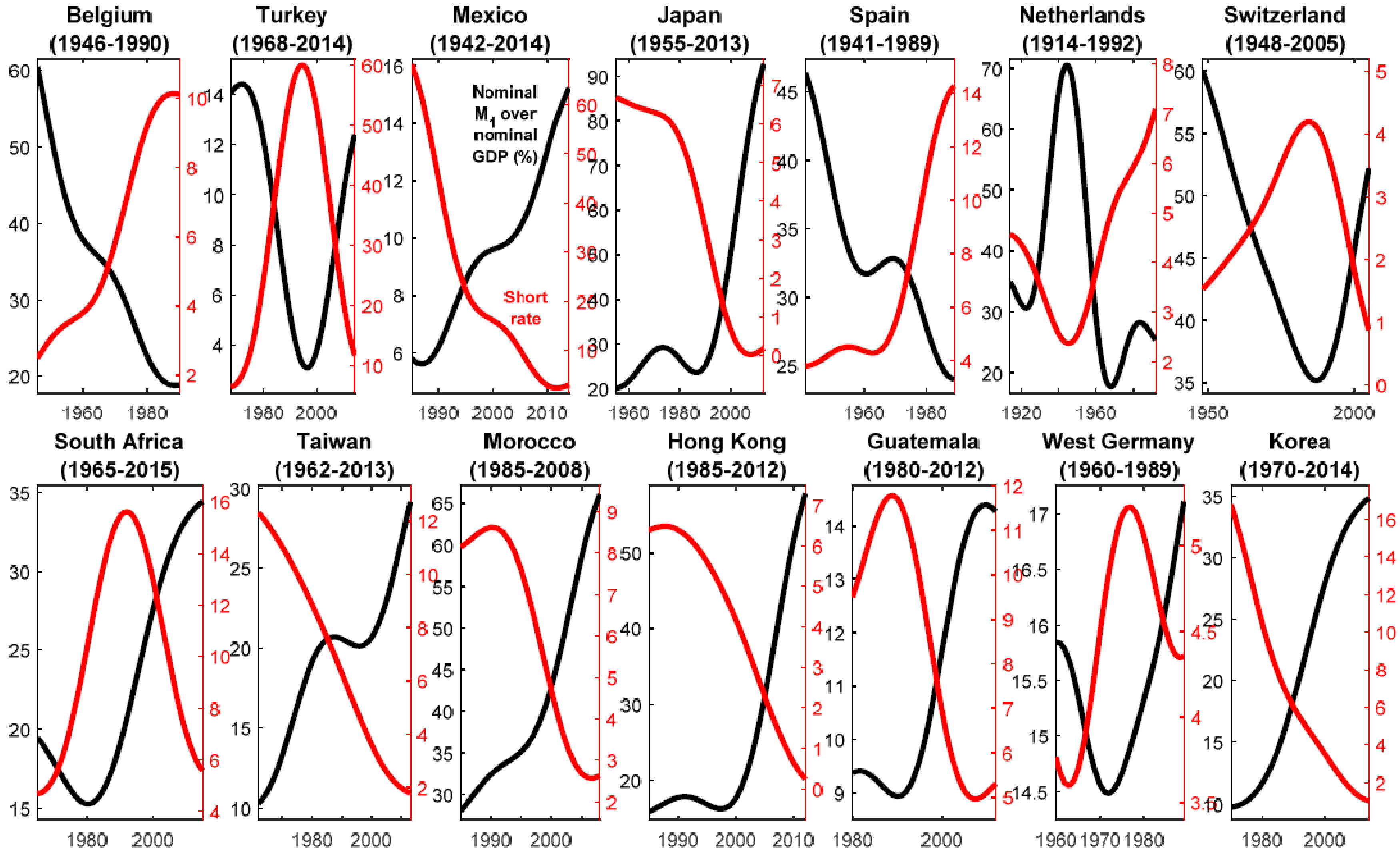

Figure 3b Low-frequency components of short rate and ratio between nominal M1 and nominal GDP for selected countries

in both cases, the points following a successful stabilization lie below the theoretical curve that matches the previous period.

Figure 2*b* reports countries for which the theoretical curve is still visually a good approximation to the data. The first row of Figure 2*c* shows countries for which the fit gets worse,[15] but there still seems to be some relationship between the theory and the data, whereas the second row of Figure 2*c* 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*) to indicate data for the Gold Standard, up until 1913,[16] which are always shown with a color different from that used in subsequent years; (*ii*) to indicate data for nonconsecutive subperiods (as in the case for France); (*iii*) to indicate different series for the short rate that cannot be linked (as in the case for Venezuela); and (*iv*) to highlight drastic changes in the relationship between velocity and the short rate (as in the case for the Netherlands and Portugal). Finally, for the U.S., the 'standard' M1 aggregate for the period since 1984 is indicated with a different color to emphasize how a failure to correct M1 (as in Lucas and Nicolini (2015), e.g.) leads to the apparent breakdown of the relationship between velocity and the short rate documented by several authors. In our view, it is remarkable how well this simple theory performs in this first inspection for a large set of countries, despite a few apparent failures.

Figures 3*a* and 3*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 2*a*-2*c*.[17] The components have been extracted *via* the filter proposed by Christiano and Fitzgerald (2003).[18] We find this evidence, which consistently points toward 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 is based on cointegration tests, the evidence in Figures 3*a*-3*b* is, possibly, even *more convincing*, because it is 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. We would like to formally test whether, as some of our simple technologies imply, the ratio between real money balances and output inherits

[15] For the Netherlands, the two World Wars and their aftermaths had been characterized by an anomalous behavior 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.

[16] Although we consider the Gold Standard to have ended in August 1914, with the outbreak of World War I, marking its *exact* end is all but impossible, as Richard Nixon's closing of the 'gold window' in August 1971 was the culmination of a decades-long unraveling process that had started with World War I. We use August 1914 as the end date primarily because we regard World War I as the single most important shock to the system.

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

[18] 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.

a unit root when the short-term interest rate exhibits a unit root. We also want to formally test whether the estimated elasticities are indeed equal to 1/2, as the simple BT specification suggests, when using the log-log specification. 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 overwhelmingly appears to be 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 several methodological issues.

4 Main Features of Our Approach

In this paper we explore the long-run demand for M1 *via* cointegration methods. The key reason for this approach 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, in either levels or logarithms. At the same time, the debate over the stability of the money demand has long made the distinction between the short run 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.[19] 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—that is, 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 searching here for a long-run money demand for *transaction* purposes, cointegration should be the natural null,[20] so that

[19] See Grossman and Weiss (1983) for an early contribution, or Alvarez and Lippi (2014) for a recent one.

[20] Basic economics 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

tests should be based only on Shin's (1994) procedure. As we discuss in the following subsections, 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 primarily 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 statistically indistinguishable from white noise most of the times. By the Fisher equation, this implies that, unless the natural rate of interest had contained a sizeable permanent component (because of permanent shifts in trend productivity growth, e.g.), nominal interest rates should be expected to have been stationary, too, which would preclude them from being entered in any cointegrated system, or cointegrating regression.[21] The integration properties of nominal rates during the Gold Standard period therefore ought to be separately checked, or otherwise we would run the risk of performing cointegration analysis based on a series that 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 specifications.)

4.2 Methodological issues pertaining to cointegration tests

4.2.1 Issues pertaining to bootstrapping

Everything in this paper is bootstrapped. In this section, we briefly discuss details of the bootstrapping procedures we use, and how such procedures perform, particularly in terms of *comparative* performance. In our discussion, we extensively refer to online Appendices D and E, which contain the Monte Carlo evidence motivating both our choices, and the way in which we will interpret the evidence.

(proxied by nominal GDP).

[21] 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, first sentence on p. 636) and Shin (1994, p. 92).

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 that is *relevant under the null hypothesis*.[22] All of the technical details can be found in CRT, to which the reader is referred. We select the VAR lag order as the maximum[23] between the lag orders chosen by the Schwartz and Hannan-Quinn criteria[24] for the VAR in levels.

As for Shin's tests, to the best of our knowledge no one has yet provided anything comparable to what CRT did for Johansen's procedure. The bootstrap procedure we propose in online Appendix E is based on the same general principle underlying CRT, that is, bootstrapping the model that is relevant under the null hypothesis. 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 online 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 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, the sample size is sufficiently large, or both: if the residual is persistent, the sample

[22] This means that for tests of the null of no cointegration against the alternative of one or more cointegrating vectors the model that 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 that ought to be bootstrapped is the VECM estimated under the null of h cointegrating vectors.

[23] We consider the maximum between the lag orders chosen by the SIC and HQ criteria because the risk associated with selecting a lag order that is smaller than the true one (model misspecification) is more serious than the one resulting from choosing a lag order that is greater than the true one (overfitting).

[24] On the other hand, we do not consider the Akaike Information Criterion since, as discussed for example 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.

is short, or both, the tests fail to detect cointegration a nonnegligible fraction of the time. 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 (that is, 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. If the DGP features *no cointegration*, however, even in large samples the procedure produces ERFs that are far from the ideal of 100%.

Thse results 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, the cointegration residual is highly persistent, or both. 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, to be less informative than those produced by Johansen's tests.

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.[25] 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.[26] Evidence points toward both a nonnegligible 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 specification 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 fash-

[25] 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).

[26] We label it as candidate cointegration residual 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 present*, which would prevent the candidate from being identified as a true cointegration residual.

ion, 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, toward 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 present, especially when $\hat{\rho}_{MUB}$ is high and the sample period is comparatively short. This implies that results from cointegration tests should *not* be taken strictly at face value, but rather should be interpreted in the light of the Monte Carlo evidence in Tables E.1 and E.2 in the online Appendix. In what follows, we therefore also report Monte Carlo-based ERFs of the tests computed *under the null of cointegration*. Specifically, we estimate the VECM under the null of one cointegration vector; we stochastically simulate it 2,000 times; and for each artificial sample we 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 present in *all* of the samples we are working with.[27]

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 online 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 modeled 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 therefore focus on the results from the two Nyblom-type tests, but we eschew instead results

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

from fluctuations tests (these results, however, are reported in tables H.2 and H.5 in the online Appendix).

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 1a reports results from Johansen's maximum eigenvalue tests of 0 *versus* 1 cointegration vectors for the U.S., together with the Monte Carlo-based ERFs computed conditional on the null of one cointegration vector. Table 1b reports the corresponding results for all other countries. Tables G.1 and G2 in the online 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 specifications 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.

The top rows of 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 any specification, the candidate cointegration residuals produced by either Johansen's or Stock and Watson's (1993) estimators; the bottom rows report 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 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 online Appendix G.1, based on unrestricted three-variable systems *it is almost impossible to reject the null of cointegration.*[28] Evidence is just slightly weaker for tests based on bivariate specifications for velocity and the short rate, in which unitary income

[28] Specifically, at the 10 per cent level we obtain *just four* rejections of the null of cointegration 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.)

elasticity has been imposed from the outset: we obtain just 10 rejections of the null based on Selden-Latané, 13 based on semi-log, and 7 based on log-log.

For the reasons discussed in Section 4.2.1, 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 present. We therefore turn to discussing the results from Johansen's tests, which, as we pointed out, appear to be 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 1*a* and 1*b* contain this paper's most important body of evidence (at least, as far as statistical tests are concerned). In both tables we highlight in yellow all p-values for maximum eigenvalue tests smaller than 10%; and all ERFs smaller than 50%, corresponding to a less-than-even chance of detecting cointegration if this is truly present in the data.

The U.S. Starting from the U.S., which has been the focus of most previous investigations, the main results in Table 1*a* 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.[29]

(*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 applying Alvarez and Lippi's (2009) 1% correction to the short rate, which should drastically improve the fit at levels of the short rate that are close to zero. In spite of this, the log-log functional form still does not allow cointegration to be detected.

(*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 any specification, and any of the three 'expanded' M1 aggregates. Based on the adjusted aggregates, evidence is slightly weaker, and cointegration is detected in four out of six cases. Note, 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 (2015) 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 now being equal to 0.069. For the semi-log specification, on the other

[29]Note, 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.

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; Monte-Carlo-based empirical rejection frequencies of the tests under the null of cointegration; and fractions of Monte Carlo replications for which the bootstrapped p-values are smaller for Selden-Latané than for semi-log, conditional on either model being the true data-generation process

		I: Bootstrapped p-values			II: Empirical rejection frequencies			III: Fractions of replications for which p-values are smaller for Selden-Latané than for semi-log if true model is:	
Monetary aggregate	*Period*	*Selden-Latané*	*Semi-log*	*Log-log*	*Selden-Latané*	*Semi-log*	*Log-log*	*Selden-Latané*	*semi-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	0.670	0.299
Standard M1 + MMMFs	1915-2016	0.006	0.022	0.289	–	–	0.649	0.574	0.346
Standard M1 + MMDAs + MMMFs	1915-2016	0.002	0.007	0.617	–	–	0.444	0.618	0.222
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	0.665	0.344
Standard M1 + MMMFs	1926-2016	0.003	0.030	0.383	–	–	0.538	0.597	0.422
Standard M1 + MMDAs + MMMFs	1926-2016	0.118	0.020	0.707	0.647	–	0.254	0.600	0.265

[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.

hand, this is the case only to a minor extent, with the p-value increasing from 0.081 to 0.104 as a result of 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 compared to deposits,[30] adjusting or not adjusting the aggregates should not make much of a difference. At any rate, we report all the results so that readers can decide for themselves.

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 as an informal 'test' of which functional form the data would seem to prefer, evidence points to Selden-Latané, as in only a single case out of nine (based on the adjusted aggregate that also includes MMMFs) the p-value produced by the semi-log specification is smaller than the one produced by Selden-Latané. In all other cases, the opposite is true. The last two columns of Table 1a show additional supporting evidence, by reporting results from the following Monte Carlo experiment. We estimate either Selden-Latané or semi-log specifications imposing one cointegration vector. Then, we simulate either DGP 2,000 times, and we perform cointegration tests based on either specification, bootstrapping the tests as we did based on the true data. The last two columns report the fractions of replications for which the p-value based on Selden-Latané is smaller than the one based on semi-log, conditional on either model being the true DGP. Evidence shows that obtaining smaller p-values based on Selden-Latané rather than based on semi-log—as we did based on the true data—is significantly more likely if the true model is Selden-Latané rather than if the true specification is semi-log. Although this evidence is not overwhelming, we regard it as quite clearly pointing toward Selden-Latané as being the preferred specification for the U.S..

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

(i) Cointegration is detected based on *all* estimated specifications[31] for Argentina,

[30] For example, for the broadest M1 aggregate that also includes 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 of the currency has been in the hands of foreigners, this amounts to about 5 to 8 per cent of overall M1.

[31] We say 'estimated specifications' because, for the reasons discussed in Section 4.1, in a few cases

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.

Brazil, Guatemala, Israel, Norway, Portugal (1914-1965), South Africa, Switzerland (1948-2005), and the U.K.. 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 present in the data. Consider, e.g., 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 3*a*, the *p*-values in Table 1*b* range between 0.642 and 0.973. Crucially, however, the ERFs show that if cointegration were present in the data, the chance of detecting it would be *between 8 and 20%*. 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, even though, in most of these cases, visual evidence points toward 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 U.S.. By applying the same informal argument we used for the U.S., based on the *p*-values produced by the three specifications, it might be argued that the Selden-Latané specification 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 U.K.. By the same token, the semi-log specification seems to be preferred for Argentina, Brazil (1934-2012), Colombia, Guatemala, Mexico, the Netherlands, South Africa, and Taiwan, whereas the log-log specification appears to be 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 U.K.), 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.

Toward a unified framework? An alternative possible interpretation of the evidence in Tables 1*a*-1*b* is 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 to try to assess which, of the three functional forms, could

we only estimate two functional forms, rather than three.

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. For instance, 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 any 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 U.S., the U.K., Australia, Canada and New Zealand: all of these countries experienced important variations on their nominal interest rates, but they are low-inflation countries. 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, e.g., 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 in the online Appendix 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 online 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; and for the Netherlands, Norway, Taiwan, and the U.K.. 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 online 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 nonexistent. Specifically, for the U.S., 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 specifications, 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. For instance, based on the Selden-Latané specification, 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 specification is -0.1. As for the Selden-Latané specification, 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 U.S. based on the Lucas-Nicolini (2015) aggregate (see Figure SELA.6).[32] As for the log-log specification, results overall are 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 the Selden-Latané specification, 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 specification 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.

[32] This is why Table SELA.3 does not report results for the United States based on the Lucas-Nicolini (2015) aggregate.

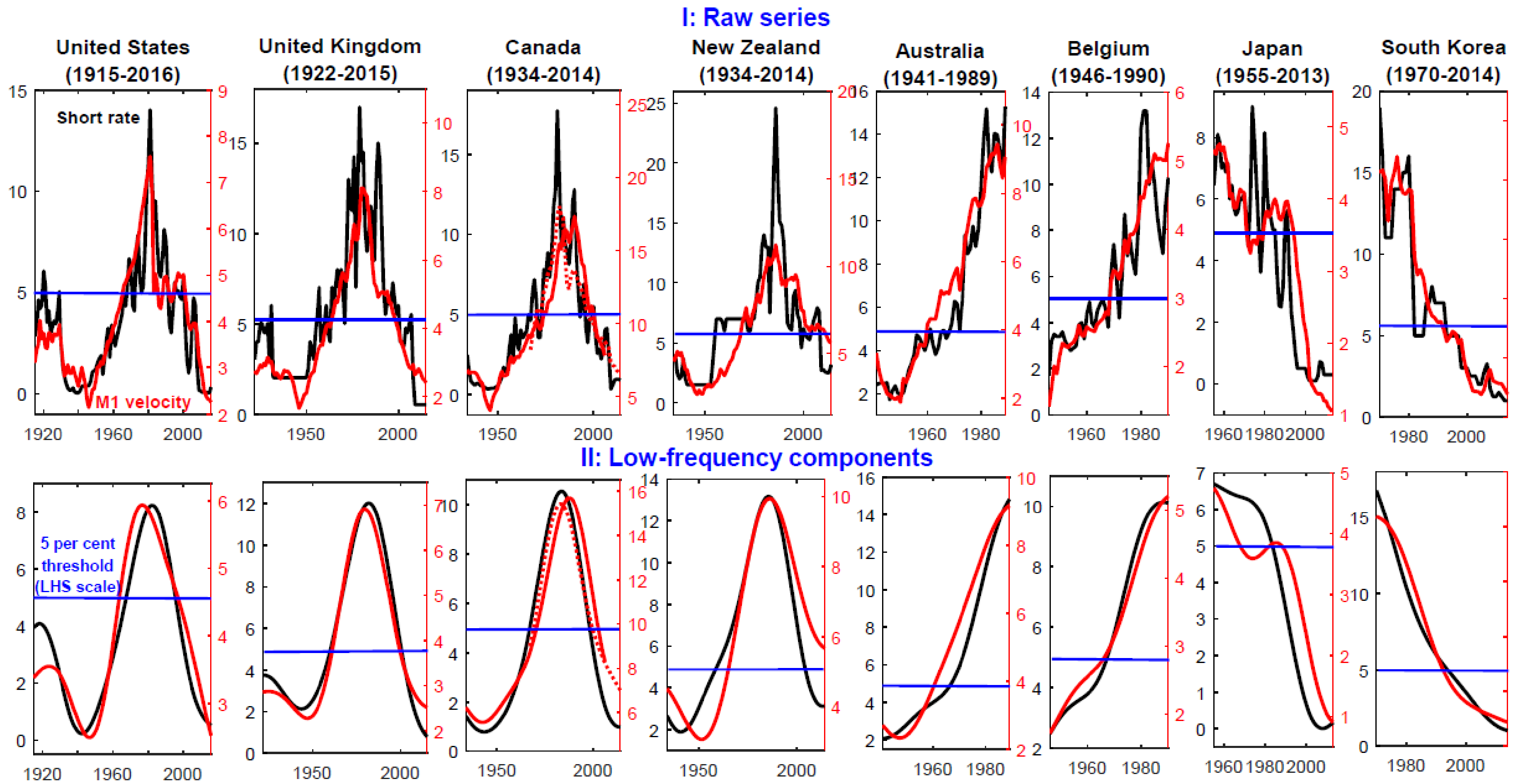

Figure 4 Informal evidence on the possible non-linearity of M1 velocity at low interest rates

6 Does the Behavior of Money Demand at Low Interest Rates Exhibit Sizeable Nonlinearities?

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 nonlinearities 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 us to meaningfully capture the nonlinearity associated with the cost of participating in financial markets.

Figure 4 shows evidence of the possible presence of nonlinearities for eight countries for which either of the subsamples with the short rate above and below 5% is sufficiently long.[33] The first row shows the short rate together with velocity (the raw series), and the bottom row shows the low-frequency components of the two series (the components have been extracted exactly as in Section 3).

The evidence in the figure 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 does the relationship between the two series appear to be different at high and low interest rates. The evidence based on the low-frequency components of the data is especially stark: once we strip higher-frequency fluctuations from velocity and the short rate, the relationship between the series clearly appears remarkably strong and stable at all levels of the interest rate. In particular, taking, for the sake of argument, 5% as a 'reference threshold' for the short rate (see footnote 33), the following should be noted:

(1) for the U.S., 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%—appear as remarkably similar to the period in-between, during which the short rate systematically exceeded 5%. 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: the relationship between the series appears the same both before and after the 1960s. For Korea, 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% to about 0, velocity likewise collapsed with

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

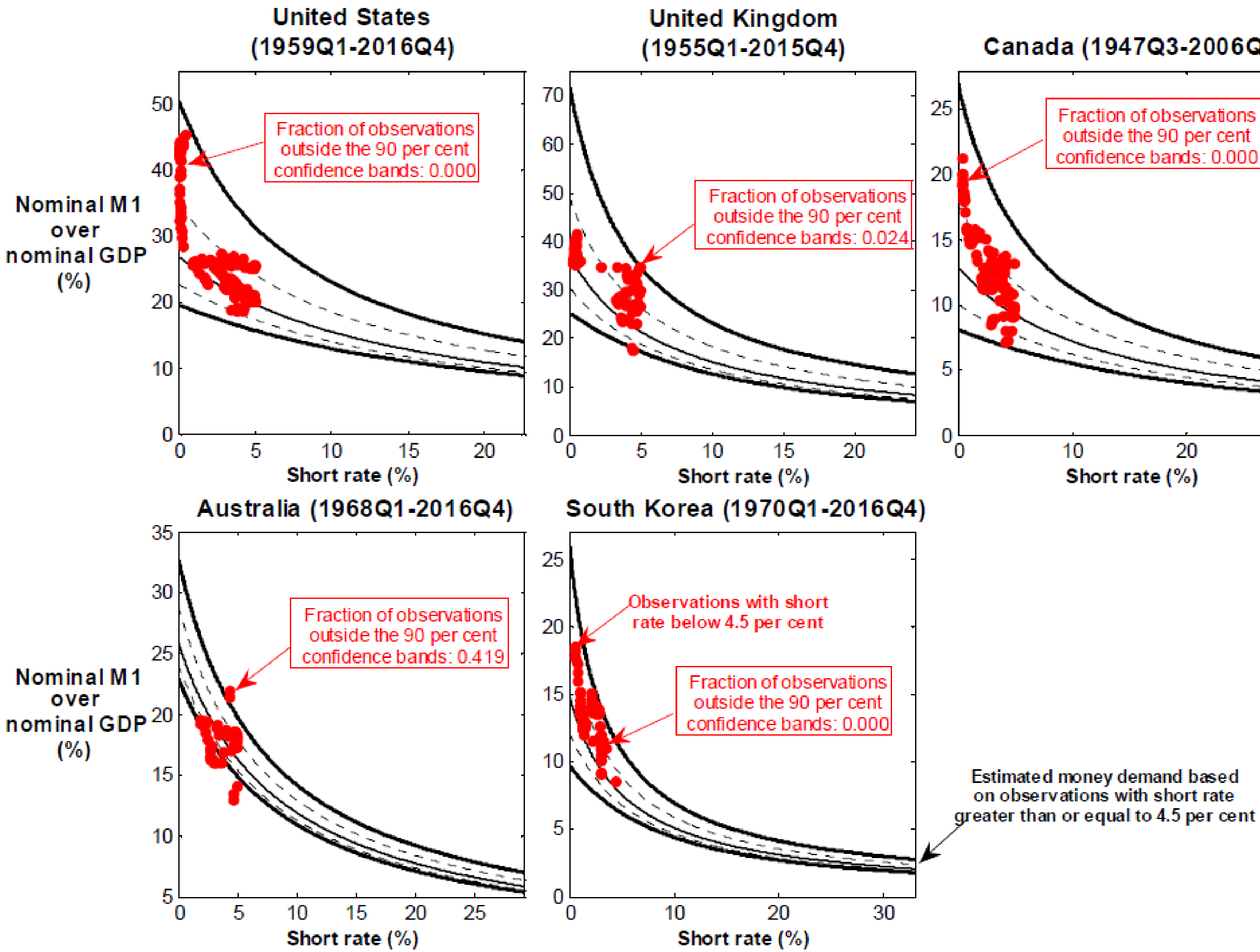

Figure 5 Statistical evidence on the possible presence of non-linearities in the demand for M1 at low interest rates

a lag of a few years. Further, note that since the start of the new millennium, velocity has kept decreasing, whereas the interest rate has remained close to zero, a pattern *opposite* to that implied by the presence of fixed costs associated with the decision to participate, or not participate, in financial markets.

Figure 5 buttresses the visual evidence in Figure 4 with statistical evidence for five countries for which we could find sufficiently long post-WWII quarterly samples, in order to get reasonably precise estimates. In either case, we estimated Selden-Latané specifications for velocity, that is,

$$V_t \equiv \frac{Y_t}{M_t} = C + \delta R_t, \tag{9}$$

based on consecutive sample periods during which the short rate had consistently been above 4.5%,[34] characterizing uncertainty by bootstrapping the VECM estimated conditional on one cointegration vector as in CRT. For each country we show the estimated implied money demand curves, with 16-84 and 5-95% confidence bands, and (in red) the observations corresponding to a short rate below 4.5%. Under the null hypothesis that the model is the same at both high and low interest rates, low interest rate observations should fall outside of the 5-95% bands 10% of the times. In fact, this *never* happens for the U.S., Canada, and South Korea, and it happens 2.4% of the time for the U.K.. For Australia, the fraction, at 41.9%, is much higher, but note that, for the purposes of Mulligan and Sala-i-Martin's (2000) argument, the outliers are on the *wrong* side of the demand curve: most of them are below, rather than above the curve.

This evidence questions the notion that there might be sizeable nonlinearities in money demand at low interest rates, and rather suggests that the behavior 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 identifying the correct functional form does indeed have material implications in several cases.

[34] For example, 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.

7.1 Spurious nonlinearity 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%) is associated with a larger change in its *logarithm* at low interest rates than it is at higher interest rates,[35] 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 specification, estimating a log-log specification will automatically produce smaller elasticities (in absolute value) at lower rather than higher interest rates. The same argument obviously holds if the true specification is the semi-log.

This can be illustrated as follows. Specification (9) for velocity implies the following expression for M1 balances as a fraction of nominal GDP:

$$\frac{M_t}{Y_t} = \frac{1}{V_t} = \frac{1}{C + \delta R_t}. \tag{10}$$

If, however, the estimated equation for money demand is of the log-log type, that is,

$$\ln\left(\frac{M_t}{Y_t}\right) = \alpha + \beta \ln R_t + \epsilon_t, \tag{11}$$

the theoretical value of the estimated elasticity turns out to be equal to

$$\frac{d\ln\left(\frac{M_t}{Y_t}\right)}{d\ln R_t} = -\frac{\delta R_t}{C + \delta R_t}. \tag{12}$$

This function is plotted on the left-hand side of Figure 6. We set C to 2.5 (corresponding to a satiation level of M1 balances of 40% of GDP) and δ to 0.4, which are very close to estimates for the U.S. based on Lucas and Nicolini's (2015) aggregate. The estimated elasticity tends to 0 for $R_t \rightarrow 0$ (in fact, at the Zero Lower Bound, it is exactly equal to 0), whereas it tends to -1 for $R_t \rightarrow \infty$.

By the same token, if the true specification is of the semi-log type, that is,

$$\ln\left(\frac{M_t}{Y_t}\right) = B - \gamma R_t, \tag{13}$$

whereas the estimated equation for money demand is still given by (11), we have

$$\frac{d\ln\left(\frac{M_t}{Y_t}\right)}{d\ln R_t} = -\gamma R_t. \tag{14}$$

[35] For example, ln(9)-ln(10)=-0.105, whereas ln(2)-ln(3)=-0.406.

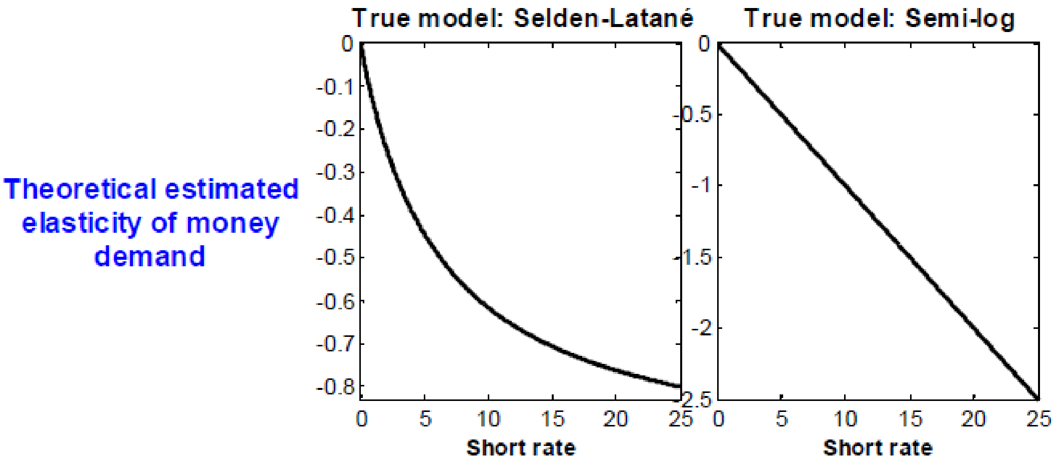

Figure 6 Evidence on the distortions originating from estimating a log-log specification when the true model is either Selden-Latané or semi-log

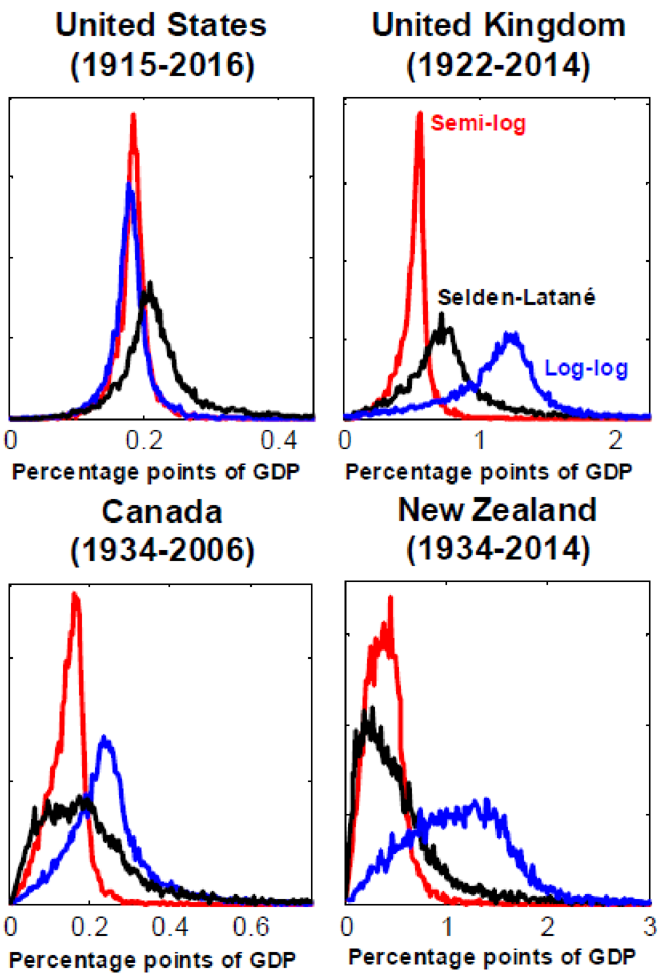

Figure 7 Implications of using alternative functional forms for the welfare costs of inflation

This function is plotted on the right-hand side of Figure 6. We set B to 1, and γ to 0.1, which are, once again, nearly identical, numerically, to the U.S. estimates. The estimated elasticity tends to 0 for $R_t \rightarrow 0$, is exactly equal to 0 at the Zero Lower Bound, and it tends to -∞ for $R_t \rightarrow \infty$.

The bottom line is that in either case, estimating a log-log specification produces entirely spurious evidence of a lower elasticity at interest rates approaching zero. To be sure, this does not mean that evidence of nonlinearity in the literature is spurious: what it does mean, 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. Two more recent contributions are Lucas (2000) and Ireland (2009), which stress the relevance of the functional form used in the empirical analysis in obtaining estimates of the welfare costs. As a second illustration of the pitfalls of using the incorrect functional form for money demand, in this section we revisit this issue bringing more countries, a new functional form (the Selden-Latané) that is preferred by the data at low inflation rates; and including the most recent years, which provide additional evidence of the behavior of real money demand at very low nominal interest rates. We estimate each of the three specifications for velocity as before.[36] 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).

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 shown that the welfare cost function associated with the Selden-Latané specification (9) is given by

$$WC_{SELA}(R_t) = \frac{\ln(C + \delta R_t) - \ln C}{\delta} - \frac{R_t}{C + \delta R_t} \tag{15}$$

Figure 7 reports, for four countries, the estimated welfare losses at the *average* short rate that has prevailed over the sample period (expressed in percentage points of GDP).[37] We will not comment on 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

[36] We characterize the uncertainty surrounding the point estimates by bootstrapping the estimated VECM as in CRT (2012). Results are based on 10,000 bootstrap replications.

[37] The differences across countries in the estimated welfare losses 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.

comparison between the popular log-log specification, and the Selden-Latané one, for the U.K. the welfare costs implied by the former are about *twice* those implied by the latter, with the median and the 90% confidence intervals being 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 U.S.[38] the three specifications tend to produce similar results, with median estimates equal to 0.21% 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 on request) are in line with those in Figure 7.[39] These simple examples illustrate the importance of correctly identifying and using the right functional form for money demand.

8 Conclusions

We use a model of a transaction demand for money to guide an investigation of the stability of the long-run relationship between M1 velocity and a short term nominal interest rate. Our dataset comprises 32 countries for periods that range from 35 to 100 years. Evidence of cointegration between velocity and the short rate is widespread, whereas evidence of breaks or time-variation in cointegration relationships is weak to non-existent. For the U.S. we detect strong evidence based on three of the adjustments to the standard M1 aggregate originally proposed by Goldfeld and Sichel (1990). For several low-inflation countries (in particular, the U.S.) the data prefer the specification in the levels of velocity and the short rate originally estimated by Selden (1956) and Latané (1960). We detect no evidence of non-linearities at low interest rates, but we show that if the data are generated by a Selden-Latané or a semi-log specification, estimating a log-log specification spuriously causes estimated elasticities to be smaller at low interest rates. Using the correct functional form has important implications for the ability to correctly estimate the welfare costs of inflation.

References

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

——— (2014): "Persistent Liquidity Effects and the Long Run Money Demand," *American Economic Journal: Macroeconomics*, 6(2), 71–107.

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

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

[39] To mention just two cases, for Switzerland the median estimates based on the Selden-Latané and log-log specifications are 0.22 and 0.53% of GDP, respectively, whereas the corresponding figures for South Korea are 0.63 and 1.37%.

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.

——— (2009): "Long-Run Evidence on Money Growth and Inflation," *European Central Bank Working Paper N. 1027, March 2009.*

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

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.

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.

ELLIOT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–836.

ENGLE, R. F., AND C. W. GRANGER (1987): "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–276.

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.

HAMILTON, J. (1994): *Time Series Analysis.* Princeton, NJ, Princeton University Press.

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.

Ireland, P. (2009): "On the Welfare Cost of Inflation and the Recent Behavior of Money Demand," *American Economic Review*, 99(3), 1040–1052.

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.

Latané, H. A. (1960): "Income Velocity and Interest Rates: A Pragmatic Approach," *Review of Economics and Statistics*, 42(4), 445–449.

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.

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.

Shin, Y. (1994): "A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration," *Econometric Theory*, 10(1), 91–115.

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