The Quantity Theory of Money: An Empirical and Quantitative Reassessment

Xi Wang‡

Washington University in St. Louis

First Draft: May, 2016
Last Date of Update: October 23, 2017
Click for Updated Version

Abstract

Almost forty years ago, Robert Lucas showed (Lucas (1980)) that a simple formulation of the Quantity Theory of Money (QTM from now onward) did surprisingly well at capturing the long-run co-movements of consumer price inflation and a certain measure of money supply (M1). The goal of this article is to re-evaluate that claim using, to the extent possible, the same statistical and economic criteria but a much larger data set covering both a longer period and many more countries.

We find Lucas’ result to be extremely fragile. It appears that the period 1955-1980 is the only period during which QTM fits data well in most of our sample countries. It starts to break down when we go beyond this period. Furthermore, the recent breaking down of QTM is global, even though different countries have different breaking dates. Fragility is robust.

To explain this breaking down for U.S during Pre-crisis Period (1980-2007), we show $M_2$ is a more robust monetary index by investigating the historical performance of $M_1$. Under the view of endogenous money. Namely, broad money($M_2$) is generated from loan issuing. We decompose the structure of loans for U.S. We found that real estate is the major collateral asset for Household and Firms. We thus propose money is after composition goods, a bundle of land and final goods. To confirm our theory, we investigate a historical nominal price index of U.S and find that (long-run) growth of nominal house price co-moves with(leads) growth of broad money more robustly. Furthermore, the timing of recent financial innovation matches with breaking data. We thus propose a channel through which financial innovation can affect our estimation of QTM.

Introduction

The Quantity Theory of Money (QTM) has been at the heart of Monetary Economics since its birth. The QTM states that the general price level should, over the long-run, co-move with the quantity of money available in the economy. Hence general inflation should co-move with the growth rate of money, and such movement should be one-to-one. This means
that the QTM is both a theory of money (it says what "money really is") and a theory of how markets for monetary exchanges function. In fact, the QTM begins with a well-known accounting identity

\[ M \cdot V = P \cdot Y \]

and turns into a theory of how the price level \( P \) is determined as a function of the available quantity of \( M \) by making assumptions about

- the way in which \( Y \) (GNP or any other pile of goods traded in monetary exchanges) is determined;

- the way in which \( V \) (velocity of money) moves over time and is or is not affected by \( Y, M \) and other economic variables such as interest rates and what not;

The QTM is a few centuries old and, from a technical vantage point, it has assumed a variety of forms. After Friedman’s classical spelling out of its modern version (Friedman (1956)), there is no doubt that Lucas has become - both on theoretical and empirical grounds - the point of reference for the contemporary research on this subject. In fact, Lucas (Lucas (1980)) was able to show that, in US data covering the years 1955-1977, M1 growth and CPI inflation moved together when short-run movements had been reasonably filtered out. Following in the steps of Lucas, many other researchers have also contributed to strengthen the view that ...the central prediction of the quantity theory are that, in the long run, money growth... should affect the inflation rate on a one-for-one basis... the application of the quantity theory of money is not limited to currency reforms and magical thought experiments. It applies, with remarkable success, to co-movements in money and prices generated in complicated, real-world circumstances...1

1 There are bunches of paper confirming this observation with different measure and from different angles. From a cross-country standpoint, McCandless Jr and Weber (1995) uses 30 years (1960-1990) averages of annual inflation and growth rates of \( M_2 \) across 110 countries to show that they line up almost perfectly along a 45-degree line. Lucas (1996) views this as a great success of QTM and Monetary Economics. From now on, I would like to refer methods analyzing the properties of data in the time domain as the temporal approach, for example, the method in McCandless Jr and Weber (1995), simply adopting sample average of raw data. More recently, Benati et al. (2016) applies co-integration tests to long spanned dataset and propose the existence of long-run money demand. Co-integration test is also a method explore (dynamic) properties of data in time domain. On the other hand, papers like Lucas (1980), Christiano and Fitzgerald (2003) and Sargent and Surico (2011) adopt frequency domain method.
In the present paper, I examine recent data from the US and from a group of advanced economies over the half-century 1955-2016 to evaluate if this statement still stands. The executive summary of my findings is that Lucas’ original formulation works quite well for M1(2) until the early-middle 1980s(early-1990s) but begins to break down after that. By the middle 1990s the one-to-one relationship between M1 (or M2 for that matter) and inflation, which was so stable for about three decades, is all but disappeared. I investigate some, somewhat ”natural,” measures of money supply other than M1 (including ”NewM1” as defined in Lucas and Nicolini (2015)) and find that, when the last twenty years or so are taken into consideration, none of them is capable of replicating what M1 used to do. Figures-1 and 1 report the (raw) time series we are studying, in levels and growth rates respectively; measures of money supply are in the upper panel while prices indices are in the lower panel in both figures.

A number of technical and theoretical issues are involved in the empirical study of the QTM: (i) the definition of what is the ”money” used in transactions; (ii) the definition of what is being transacted, which needs not necessarily be GNP; (iii) a convincing way of measuring the ”long-run movements” of the various variables. These are difficult problems, and I will next describe how I approached them, starting from the last, which is in some sense the easiest.

Figure 1: Does Price of Final Goods Follow any Money index? Source: H.6 Money Stock Measure table provided by the Federal Reserve Board, 1959-2016, Annual data
In his pioneering work Lucas (1980) used an elementary band-pass filter to extract the long run\(^2\) signal from M1, CPI and the T-bill rate. This procedure is not necessarily immediate because at least M1 is a non-stationary time series. To make the band-pass filter work and avoid a spurious regression Lucas (1980) used the annual growth rates of \(M_1\) (hence of the CPI) in his statistical analysis. The stationarized sequences are used to estimate “long-run” signal.

After estimating the long run trends of inflation and money growth, Lucas (1980) plots (1) (QTM) long-run(filtered) growth rate of money against inflation rate; (2) (Fisher Effect) long-run(filtered) growth of money against T-bill rate. Lucas (1980) finds that these points line up almost perfectly along a 45-degree line. Hence inflation rate co-moves with money growth on a one-to-one basis. Furthermore, nominal interest rate is sum of real interest rate and expected inflation rate. Lucas (1980) argues that interest rate is determined by the status of the real economy, which is in long run irrelevant with money stock. The long-run variation of nominal interest thus reflects variations of expected inflation. The relationship between money and the nominal interest rate is thus determined by the effect of money onto (expected) inflation. As a result, QTM reveals itself with direct and indirect evidence from

\(^2\)Long-run signal, in Spectrum Analysis, refers to the signals with low frequencies, for example, frequencies around \(\omega = 0\). But it is practically impossible to identify the spectrum on \(\omega = 0\). For more details, one can refer to Sargent (1987), Hamilton (1994) and Lütkepohl (2007).
its influence over inflation. In this article, we have no intent to defend or attack whether money neutrality, we instead focus on the validity of QTM. Thus, all the empirical findings below are all adjusted for real income.

Recent evidence of U.S presents a counterexample to QTM, no matter real income adjusted or not. Inflation reacts to money growth, at best, in a drawing way, certainly not on a one-for-one basis anymore. In upper(down) panel of Figure-1 and -2, we plot out the level(growth) of these normalized index during 1979-2007. In the upper panel, we normalize the level of $M_1(0/2)$ stock and price all to be 100 at 1979. From the figures, one can tell that price of consumption(Grey lines) did not keep pace with any monetary index(Black lines), during 1979 to 2007, no matter what kind of price index we are using to represent the price level. In Figure-1 and 2, we used CPI index, PCE Chain price index and GDP deflator. One can tell that PCE Chain price has almost the same trend with GDP deflator and neither of them comoves with any money index. In sum, all price index grows at an almost constant rate ignoring fluctuations from money part at the raw data level. We thus investigate whether QTM holds for U.S in a long history.

We are not the first to investigate whether QTM was stable across time for U.S. Benjamin M. Friedman(1988) did a preliminary test\(^3\) to check the relationship between price and money, and found the collapse of the one-to-one relationship between money and prices in the 1980s. There are also several recent papers take “long run” seriously, documenting breaking down of QTM(after some date). For example: “...For most of the last 25 years, the quantity theory of money has been sleeping...(Sargent and Surico(2012), Page 110). “...in the later period, the relationship(between money and inflation) turns negative in frequencies 20 years or higher.....(Christiano and Fitzgerald(2003)).

Our empirical findings confirm this recent breaking down of QTM in U.S. We find that the period 1953-1977 which is under the investigation of Lucas (1980), is a special period, beyond which QTM is hardly a tight law governing the one-for-one relationship between money and inflation. For example, during the period from 1945-1954, (long run) inflation ran off the track of (long run) growth rate of the money stock, no matter we are using M1 or M2. However, U.S is not alone.

\(^3\)Benjamin M. Friedman(1988) does not take “long-run” title into consideration.
This fragility of QTM is robust across countries. In our robust check, we found QTM is not a global(universal) law: we have countries where QTM never holds and still holds. Furthermore, for more countries, QTM tends to hold for a while, then collapse. Though different countries have different breaking dates and degrees, collapsing QTM is qualitative robust. For example, QTM never holds in Germany and France. And QTM used to exist in Australia, but not after 2000. QTM used to exist in Italy, but not after 1998. Countries deviate away from QTM in their own ways. Before having a unifying theory to explain the breaking down, we believe one should investigate countries’ own monetary history.

Put evidence on one side; researchers have proposed several explanations for collapsing QTM. For example, Sargent and Surico (2011) adopts a general equilibrium framework and proposes disinflation policy as a candidate. Benati (2009) believes monetary velocity shock plays a role. Teles et al. (2016) adopts the temporal approach to test whether disinflation policy weakens the relationship between money growth and inflation. Following the same approach, McCallum and Nelson (2010) finds QTM deteriorates after disinflation policies. Nevertheless, all supportive evidence provided by the temporal approach is misleading to some degree: they never checked whether QTM exists in the investigated countries or not. For example, they include France in their sample. But it is hard to conclude that QTM used to exist in France. From cross countries long-dated data, we thus believe it is more useful to investigate in a case-by-case way.

Theoretically, disinflation policy must have a role in the revealing process of QTM, since the central bank is supposed to have the power of controlling the money supply. However, to make disinflation policy to have any effect on QTM or estimated QTM, one has to adopt a framework with uncertainty. If our economy is under a circumstance without any shock, and the central bank has an incentive to change(lower) inflation rate. The Central bank must have channels to make its inflation-rate choice. In another word, central bank needs a policy tool to implement its policy target. The channel there should be QTM or stable money demand function. In other words, QTM lays out a menu, and central bank picks its favorite from this menu. In the end, no matter how central bank dislikes inflation or what policy target central bank chooses, all the realized points should line up according to this menu: QTM. The one-to-one relationship will still hold in a deterministic world.
Logically disinflation policy should, therefore, have no effects on the relationship between money growth and inflation without the help of (relative) randomness (of money supply and demand). With the help of randomness, the problem becomes a simultaneous estimation process, classical wisdom applies. Relative volatility of supply and demand play a critical role, the effect of the slope of the supply curve, at best, is second-order, see Wang (2015).

Without the help of randomness, collapsing of QTM should be explained from the angle of money demand. To make QTM be A THEORY out of an identity(or to have a stable money demand), people place restrictions on money velocity, for example, Lucas and Nicolini (2015) links money velocity with interest rate$^4$. As long as velocity is not stable, QTM will be gone. Hence, a micro foundation to this changing velocity is necessary to explain collapsing QTM.

To micro-found changing(decreasing) velocity of QTM, we went back to money generation process. However, we have already established that Transaction purpose should not serve as a micro-foundation for the Money demand, i.e., $M_1$ does not robustly co-moves with inflation. Instead, $M_2$ does. Though $M_2$ version of QTM breaks, it sheds lights on the direction of us to investigate: $M_2$ is the major liability of depositary financial institution. On the asset side, loan plays a major role. In other words, instead of investigating deposit demand, we look at credit generation or Money generation process. Beside deposit, loan issuing also generates money. In another word, money demand is not only represented by deposit but also by the loans. As taught in Econ 101 class, a large portion of the money is generated by the financial system. We call the process of money generation “money multiplier. And money-multiplier heavily depends on the process of loan issuing. For example, Commercial and Industry loan can be used by firms to purchase intermediary goods and input. Consumer credit can be used to buy final consumption goods. These types of loan can boost prices of final goods. Instead of tracking the demand of deposit, we track the composition of loans. In U.S, major kinds of loans are associated with real estate. For

---

$^4$This cannot explain recent breaking of QTM either: points after 1984 all lie under the fitting curve. However, if one would like to link velocity with interest rate. Stationarity of interest rate implies a stable long-run velocity. Hence, a long run money regression still applies. Furthermore, the channel in Lucas and Nicolini (2015) cannot explain breaking QTM: since interest rate starts to become stationary again after 1990. It is hard to conclude that there is a trend in the process of the T-bill rate during 1990-2007. However, QTM still breaks after 1990.
example, nearly 70% of all commercial and industrial loans in the United States are secured by collateral assets (Berger and Udell (1990)). And real estate is an important tangible asset for small and large firms (from Z.1 Tables and Liu et al. (2013)). In another word, Real estate, as an important collateral, generates significant loans. Meanwhile, in U.S the fraction of real estate loan out of total loans stays high since the beginning of available data. In sum, Real estate not only serves as collateral generates loans, but it also generates loans directly by being a transaction target. This fact inspires us to propose that (long-run) demand of money and real estate are intertwined.

To confirm our judgment, we investigate the historical nominal price of house during pre-crisis period(S&P CoreLogic Case-Shiller Home Price Indices, available from Robert Shiller’s website,; Longest dataset we can find) for U.S. And we found that in U.S (filtered) growth rate of nominal house price co-moves nearly perfectly with (filtered) growth rate of broad money during 1955-2007, not only during Lucas period(1955-1980). Put in other words, index of house price moves more robust with the growth rate of money than CPI. Hence, the demand for money or money generation should be closely linked to housing purchase. Then here comes another question: can we explain the valid QTM during 1955-1980(90)? Can we have a general theory to nest the classical QTM? Both answers turn out to be yes.

During 1955-1990s, (filtered) price growth of house tracks CPI(PCE price) inflation well, or one can think reversely: CPI(PCE) price index tracks the price of the house well. However, after a certain date, (filtered) inflation fails to track the price of the house. As notes by Davis and Heathcote (2007), cost of land plays a more and more important role in determining the price of house. One can think house as a bundle, comprising consumption goods(reproducible structure) and non-reproducible plots of land. If consumption goods contribute a constant fraction, there will be a constant fraction of money used to purchase consumption goods. However, if the fraction going into final goods decreases, in other words, the fraction of land increases, we should expect a larger fraction of money going after land, leaving the price of consumption goods growing more slowly than money growth. Put in another words, the real price of land increases, thus creates larger money demand. With this endogenous money generating process, we then explore the effect of a potential
explanation of breaking QTM: financial innovation. Since financial innovation also starts around mid-1980s. We are thus to propose a broader money generating process: borrowing collateral by land also generates money. Under this endogenous money generation process, we then explore the implication of financial innovation.

Concerning to empirical method, we adopt the original method of Lucas (1980) in the main body of this article. As a robustness check, we also used bandpass filter like Christiano and Fitzgerald (2003) and (window) spectrum at like Benati (2009). Though the idea behind Lucas (1980) is to estimate long-run signal out of original data, it is worth to repeat that the estimation of the spectrum at $\omega = 0$ is notorious of unstable property. Furthermore, we do not have a criterion or would like to take a stand on how long should be named to be “Long run, 30 years or over or 40 years or above? For example, if we would like to explore dataset of U.S is from 1955-2015 (post-Lucas period), 30 years or above will also have the similar disadvantage of a spectrum estimation at $\omega = 0$. And laying out the gain estimator would be a misleading part, high value of this estimator tends to have a high variance. Hence, when we have high gain estimator, we tend to accept that QTM is still alive. However, an estimator significant higher than one should be viewed as a rejection of QTM. And it is also straight forward to extend our analysis to adopt a Fourier Analysis approach, for example one can take $\{\sum_{j<5}[\alpha_j\cos(\omega_j(t-1)) + \delta_j\sin(\omega_j(t-1))]|_{t=1}^{T}\}$ to be the estimated long-run signal. Then one must have a criterion on which frequencies should be chosen. Fourier Analysis is available from the author upon request; we closely follow Lucas (1980) in our main text.

In the related literature, various empirical methods were used. For example, Benati (2009) uses nonparametric spectrum estimation and found that QTM is not stable. Rolnick and Weber (1997) adopts the temporal approach and finds that correlation between inflation and money growth was weaker during the standard Gold Period. Sargent and Surico (2011) uses time-varying VAR to establish the recent failure of QTM. However, Sargent and Surico (2011) adopts filter of Lucas (1980) with 16 quarters window length. Benati et al. (2016) explore several co-integration tests(temporal approach) to prove the existence of long-run money demand$^5$.

$^5$Because of cross-correlation, cointegration specification is supposed to be a more appropriate one than
To summarize, in the first part of this article, we check whether QTM ever holds across time in U.S. To robustly check my results for U.S, I did a cross-countries analysis. Put in other words, we ask the following question: Does Bob get lucky when he explores the data of U.S during 1955-1977? Does QTM hold for developed countries during 1955-1977? As a robust check, we also checked the performance of QTM across countries during the years after 1980. We find that for U.S, QTM breaks down. Our results are robust even if we abandon the post-crisis period, namely QTM breaks way before 2007. And later we will focus on this pre-crisis period. In other words, 1953-1977 is a special period. In the robust check, I investigated data of 13 countries during a long historical period, mainly from 1870-1880s or early 1890s to 2016. For each country, we check whether QTM exits 1955-1980 and post-1980. During 1950s-1980, for most of the countries in our sample, the growth rate of money provides a good explanation of inflation. Furthermore, the latest breaking down of QTM happens at a different time for different countries (Monetary measure). For example, the breaking date for U.S is 1991 if we focus on M2, while it is 1984 if we focus on M1. And this breaking date also depends on window length of our filter, for example, Sargent and Surico (2011) used a window length of 5 years and identified 1984 is breaking date.

Logically, if quantity theory of money holds for most of the countries during 1953-1977 but not recently, it is meaningful to ask another question: Why does QTM break down recently? And an interesting policy question would be whether and how the central bank can control inflation. Why is there no co-movement of inflation after expansion of monetary aggregates? However, as shown in data, each country has its pattern, we believe it is beneficial to investigate country by country. We thus investigate U.S. case in this article and leave the rest for future research.

In the second half, we explore a possible explanation of the recent breaking down of QTM for U.S during the pre-crisis period. We explore the historical nominal house price of U.S (Robert Shiller’s price index). We found that for U.S, in long run, the growth rate of house price co-moves with the growth of M2. Since money is endogenously generated,

\[ \text{difference specification, which is taken by Lucas (1980) and us} \]

\[ \text{6It turn out that Quantity Theory of Money is not a global pattern before the 1950s, but one may doubt} \]

\[ \text{the quality of data during the pre-WWII period, we thus focus the results after 1955. Results of pre-1955} \]

\[ \text{are available from the author upon request.} \]
a higher growth rate of house price generates higher demand of loans, which will generate money endogenously. Before the 1980s, the relative price of land and final goods stays stable. Hence the fraction of money related to final goods transaction is stable. However, since the mid-1980s, following a sequence of financial innovations, the land becomes better and better collateral, the relative price of land becomes higher. With complementary assumption, this will cause the fraction of money related to final goods decreases. Or the growth of money is not generated by final good inflation but asset inflation. Hence the relationship between the growth of money and inflation breaks. Namely, narrowly defined QTM breaks.

To model financial innovation, we use a simple occasionally binding model with collateral borrowing and Bayesian Learning following Boz and Mendoza (2014), and our new money demand function. Under this framework, we can generate a pattern with growing money stock but with significantly lower inflation. To solve this model, we adopt the method of policy iteration with the endogenous grid point. Then we calibrate our model to U.S economy and check the implied relationship between money growth and inflation.

**Empirical Method**

When one stares at the identity $MV = PY$, where $M$ is the stock of money, $V$ is money velocity, $P$ is aggregate price level, and $Y$ is real income. Without any restriction, it is just an identity without any further empirical implications. QTM claims a constant money
velocity in the long-run. Then it implies that \(\Delta M = \Delta P + \Delta Y\) or \(\Delta P = \Delta M - \Delta Y\), with money neutrality assumption, \(\Delta Y\) can be viewed as a error term, uncorrelated with \(\Delta M\). Hence, when one run a regression \(\Delta P = \alpha + \beta \Delta M + \epsilon\), QTM is a hypothesis that \(\beta = 1\). Without money neutrality assumption, regression can be run as \(\Delta P/Y = \alpha + \beta \Delta M + \epsilon\). Now \(\epsilon\) is interpreted as measurement error. QTM is still a hypothesis stating \(\beta = 1\).

To extract low frequency signal from raw data, we follow the process of Lucas (1980). First, we take fourth order difference of (quarterly) the raw data to make them into stationary processes:

\[
g_m(t) = \log(M_t) - \log(M_{t-4})
\]
\[
\pi(t) = \log(CPI_t) - \log(CPI_{t-4})
\]
\[
g_y(t) = \log(y_t) - \log(y_{t-4})
\]

Where \(CPI_t\) and \(y_t\) stand for consumption Price index and real GDP at time \(t\). And \(M_t\) stands for Monetary aggregate, which can be currency plus reserves \(M_0, M_1, M_2\) even \(M3/4\) for some countries. Since the definitions of Monetary index are different across countries. We will try to label monetary stock as narrow or broad money in robust check section, but all empirical process are the same.

After stationarizing the data, Lucas (1980) then uses exponential decreasing sequence as filter coefficient to extract long-run information. To be more precise, Let filter coefficient \(\omega_{\rho,L}(n)\) with bandwidth \(L\), for \(n = 0, 1, 2, \ldots, T\), defined as

\[
\Psi_n(\rho, L) = \begin{cases} 
\rho^{|n|} & |n| \leq L \\
0 & |n| > L
\end{cases}
\]

I will denote \(x(t)\) and \(x^f(t)\) for \(t = 1, 2, 3, \ldots, T\) as raw and filtered data respectively. \(x^f(t) = \sum_{s=1}^{T} \Psi_s(\rho, L) x(t - s) = \Psi(L)x(t)\). We explicit write \(\Psi_n\) as function of \(\rho, L\) to emphasize that filter \(\Psi(L)\) depends on discounting rate \(\rho\) and bandwidth \(L\). One can try different bandwidth and discounting rate of this filter, to extract long run signal out of raw data\(^7\). And it is worth to note that here we use zero padding technique\(^8\), which is common in signal

\(^7\)when the gain function concentrate on low frequency. Variation of filtered date comes from low frequency variance of original data series; See Sargent(1987) or Stoica and Moses(2005)

\(^8\)To be more precise, we append the original data with infinite 0. Say \(\{Y_t\}_{t=1}^{T}\) is our observation, then we apply analysis to sequence \(\{Y_t\}_{t=-\infty}^{\infty}\), where \(Y_t = 0\) if \(t > T\) or \(T < 1\)
analysis. After fixed bandwidth and $\rho$ to be sixteen years and 0.95, we extract out long-run growth rate of Monetary stock, Inflation and real income. Then we would like to check how long-run inflation correlates with long run growth of money.

The idea of Long-run signal can be understood as following: when one applies a Fourier Transformation to our data, we can decompose the signal, represented by the data, into a sequences of sub-signals with different frequencies (periods) and their corresponding weights $^9$. Formally, for any $\{y_t\}_{t=1}^T$, there is $\{\alpha_j\}$, $\{\delta_j\}$ such that $y_t = \bar{y} + \sum_j [\alpha_j \cos(\omega_j(t-1)) + \delta_j \sin(\omega_j(t-1))]$, where $\omega_j = \frac{2j\pi}{T}, j = 1, 2, ..., \frac{T-1}{2} \left( \frac{T}{2} \right)^{1011}$. Here, with the help of formula, we can define long-run signal to be those signals with low frequencies $\omega_j^{12}$, for example $\{\alpha_1 \cos(\omega_1(t-1)) + \delta_1 \sin(\omega_1(t-1))\}_{t=1}^T$ is the long-run signal associated with frequency $\omega_1$ or period $\frac{2\pi}{\omega_1}$.

To understand why filter can extract a particular group of frequencies out. It is helpful to step back and view our data on frequency domain. Suppose our demeaned data $\{x_t\}_t$ is generated by a stationary and ergodic data generating process $^{14}$, we denotes auto correlation sequence as $\{\gamma_k\}_{k=1}^\infty$, where $\gamma_k = E[x_t x_{t-k}]$. According to inverse Fisher Reitz theorem, there is well-defined function $g(\omega) = \sum \gamma_k e^{-ik\omega}$, we denote $s(\omega) = \frac{1}{2\pi} g(\omega)$. $s(\omega)$ is then named to be spectrum density of $\{x_t\}_t$. Under a mild condition, an equivalent definition can be used as $s(\omega) = \frac{1}{2\pi} \lim_{N \to \infty} E\left[ \frac{1}{N} | \sum_{t=1}^N x_t e^{-it\omega} |^2 \right]^{15}$

Intuitively, we can view any random variable as a sequence with energy. A constant se-

---

$^9$Riesz Fischer theorem.

$^{10}$This perfect fitting comes from orthogonality of sequence $\{\sin(\omega_j(t-1))\}$, $\{\cos(\omega_k(t-1))\}$

$^{11}$The formula of $\{\alpha_j, \delta_j\}_{j=1}^{T-1}(\frac{T}{2})$ are from Residue Theorem.

$^{12}$Whiteman (1984) interprets the filter of Lucas (1980) is one to identify the signal with frequency 0. But as we have stated in previous footnotes, it is mostly practical impossible to accomplish. With the formula here, we can put it in a more explicit way: the longest cycle we can identify from a dataset with length $T$ is $T$. It is impossible to torture a dataset to identify a cycle with an infinite period. Furthermore, the gain of spectrum of filter in Lucas (1980) is not a precise delta function on frequency $\omega = 0$. In the language of spectrum analysis or electronical engineering, It has sidelobe leakage, see Chapter 2, Stoica and Moses (2005)

$^{13}$The so called “long-run risk” finance literature also refers to the lasting effects of a shock to the (small) persistent component of consumption growth. But LR model implicitly assume a stochastic volatility element, since persistent consumption or dividend process alone cannot generate a high enough equity premium. But concerning to “long-run”, both emphasize the persistent component.

$^{14}$We implicitly assume it is a one dimensional data process, it is straightforward to extend this intuition into multidimensional case.

$^{15}$This mild condition is absolute value of $\{\gamma_k\}$ decays sufficiently enough, so $\lim_{N \to \infty} \frac{1}{N} \sum_{k=-N}^N |k| |\gamma_k| = 0$. The equivalence can be proved as following:
quence is one with zero energy, it does not move at all. Similar to physics, a more energetic sequence is more active, in another words, more volatile. To measure volatility, we use variance as an index. Inverse Fourier transformation of \( s(\omega) \) or from the second representation of \( s(\omega) \), one can easily tell that \( \text{Var}(x) = \gamma_0 = \int_{-\pi}^{\pi} s(\omega) e^{ik\omega}|_{k=0} d\omega = \int_{-\pi}^{\pi} s(\omega) d\omega \). \( s(\omega) \) thus represents the energy contributed by sub-signal with frequency \( \omega \).

When we apply a filter, say \( H(L) \), to the raw data. The spectrum density of filtered data can be represented by \( H(e^{-i\omega})s(\omega)H(e^{i\omega}) \) or \( |H(e^{-i\omega})|^2s(\omega) \), where \( s(\omega) \) is the spectrum density of raw data. The ideal filter should mimic function shape of delta function: it is zero elsewhere except for one point. As Figure-3, we plot out the \( |H(e^{-i\omega})|^2 \) of Lucas filter with different \( \rho \) and window length. Since, gain values of these six specifications go to zero after \( \frac{\pi}{9} \), we plot out the value from \([0, \frac{\pi}{9}]\). One can tell that, though with a different shape, values \( |H(e^{-i\omega})|^2 \) of all six specifications peak at zero and then decay, like a delta function, but not perfect delta function. Thus Lucas filter is a filtering extracting signals with low frequencies. Furthermore, it is worth to notice that as \( \rho \) or window length \( L \) increases, extracting quality improves.

During the period under investigation of Lucas (1980), the growth rate of real GDP can be viewed as a stable series. However, during a long historical period, we are to explore, the growth rate of income cannot be taken as given. Furthermore, we do not tend to defend or attack money neutrality. To take this time-varying growth rate of real GDP into account, we adjust our money growth by the growth rate of real GDP, especially when we analyze

\[
\lim_{N \to \infty} E\left[\frac{1}{N} \sum_{t=1}^{N} x_t e^{-i\omega t}\right]^2 = \lim_{N \to \infty} E\left[\frac{1}{N} \sum_{t,s} x_t x_s^* e^{-i(t-s)\omega}\right] \\
= \lim_{N \to \infty} \frac{1}{N} E\left[ \sum_{k=-N+1}^{N-1} (N-|k|) \gamma_k e^{-ik\omega}\right] \\
= \sum \gamma_k e^{-ik\omega} + \lim_{N \to \infty} \frac{1}{N} \sum_{k=-N}^{N} |k| \gamma_k e^{-ik\omega} \\
= g(\omega)
\]

The last equation follows the fact that mode of last term goes to zero as previous assumption states.

---

16 As a more general statement, we can represent the data generating process as \( y_t = \mu + \int_0^\pi \alpha(\omega) \cos(\omega t) + [\delta(\omega) \sin(\omega t)] \), so signal is a compound by series of sub-signals.

17 Since \( s(\omega) \) is an even function and periodic, with period \( 2\pi \). Normally, we plot out the value of spectrum from \([0, \pi]\).
European countries in robust check part. Adjusting real income is important for QTM revealing. For example, without adjusting real income, QTM does not show up during 1955-1980 in Switzerland. We thus include and focus on the cases with income adjustment.

Following Lucas (1980), we plot filtered growth rate of money or excess (adjusted) growth rate of money against (filtered) inflation rate. Excess growth rate of money is defined to be growth rate of money minus growth rate of real GDP.

Lucas (1980) implicitly adopt an eyeball metric to illustrate the nearly perfect linear (one-to-one) relationship between (filtered) growth rate of money and inflation. This method, interpreted by Sargent and Surico (2011), can be boiled down to a test whether the coefficient before money growth equals to one if we run a regression of monetary growth onto inflation:

$$\pi(t) = \alpha + \beta g_m(t) + \epsilon$$

$$H_0: \beta = 1$$

Before we move on into the estimation results, it is worthy to note that $\beta = 1$ does not necessarily mean quantity theorem holds. One can easily scatter several points along the 45-degree lines to get unite slope. But it is hard to conclude that these two are closely related. Quantity theorem of money does not only requires more than $\beta = 1$, but it also states that money should be the dominating driving factor of inflation. Put it in another
Figure 5: Filtered Growth rate of Money stock against Inflation rate, 1955-2006, U.S

way, $R^2$ of this regression should be close to 1. And this is the reason why papers like Benati (2009) includes coherency. In spectrum analysis, coherency can be understood similar to $R^2$. However, coherency estimation without any window adjustment will be one by definition. It is reasonable to doubt this measurement even after adjustment. We thus prefer to implement a simple regression method in this section. Next, we turn to discuss the empirical results for U.S, data description and result for other countries is reserved in Appendix.

Empirical Results

In this section, we present data and extend Lucas’s scatter plots of filtered money growth and inflation as well as the cointegration test of long-run money demand. We find that recent decades experiences collapsing of QTM no matter what kind of empirical methods are adopted.

Data

To avoid data quality issue, we focus on data after WWII, but our results are robust when one would like to extend data. We use quarterly U.S Real GDP(M1/2) data from FRED from 1947Q1(1959Q1). M2/1 data before 1959Q1 are from Appendix B of Balke and Gordon
Figure 6: Ratio of (New) $M_1$ over GDP against interest rate (3 month) with Lucas (1980) Exercise, Baumol-Tobin Specification, 1915-2006, Annual data

(1986). We plot out the growth rate of raw data in Figure-4; the gray area in the Figure represents periods identified as the recession by NBER. We can confirm the era of Great moderation, a reduction in the volatility of business cycle fluctuations starting in the mid-1980s\(^{18}\). And this is one of the reasons we insist on adjusting for real income. Not only we can observe a decreasing volatility, but we can also tell that inflation trends down and never came back, no matter how money stock grows or nominal interest rate adjusts.

Scatter Graphs: Frequency Approach

We plot filtered money growth(income-unadjusted case in the upper panel and adjusted one in the down panel) against inflation in Figure-5. As we can tell from the figures, the relationship between growth rate of M1/2 and inflation, no matter controlling the growth of income or not, is nearly perfect during Pre-1980 period. To facilitate illustration, we label different periods with different colors. When we compare the graphics in upper panel with their counterparts, one can tell that for U.S under the period we are investigating, controlling the growth rate of income improves the fitting of QTM. However, adjusting income cannot save QTM from recent breaking down, the relationship between money and price flattens out after some points. After a certain point, the growth of money is associated

\(^{18}\)This pattern is well documented by Stock and Watson (2003), Bernanke (2004), and Clark (2009)
with inflation in a non-one-to-one way, if they are still related.

Growth rates of M1 and M2 fit inflation perfectly during Lucas Period. For upper panel in Figure-5, without controlling real income, $\beta$ and $R^2$ are 1.1(1.3), 0.86(.81), for M1(2). $\beta$ and $R^2$ will become 0.82(1.0075) and 0.82(0.96) for M1(2) when we control for real income. As one will see in robust check section, Lucas period(1955-1980) is a unique interval during which inflation co-moves with monetary indexes closely(or reversely) in most countries. Since we used filtered data for QTM regression, it is hard to use standard deviation to determine whether we can reject hypothesis $\beta = 1$ or $R^2 = 1$. We adopt an ad hoc criterion here\textsuperscript{19}, we used $\beta = 0.7, R^2 = 0.8$ as a benchmark\textsuperscript{20}. However, for $M_1(2)$, $\beta$ is significantly less than 1 during post-1984(1991). For the points in Figure-5, during post-1984(1991) era for $M_1(M_2)$, $\beta = 0.33489(-0.1), R^2 = 0.0889(0.0064)$ if not adjust for real income, $\beta = 0.002(-0.2), R^2 = 0(0.25)$ for adjusted case. It is also worth to note that with different bandwidth, it is possible to find different breaking dates. For example, when we use 16 years as bandwidth, QTM of $M_2$ broke around 1991(1993), while Sargent and Surico (2011) finds QTM broke around mid-1980(1984) by using four years(8 quarters before and eight quarters after) as bandwidth. However, as we showed before, shorter window length means poorer extracting quality.

**Evidence from Cointegration test: Temporal Approach**

We set out a general cointegration specification below, then interpret Benati et al. (2016) through this setting:

\[ p_t = m_t + v_t \]

\[ m_t = m_{t-1} + \xi_t \]

where $p_t = \ln(P_t)$, $v_t = \ln(V_t)$, $m_t = \ln(M_t) - \ln(y_t)$, $P_t$, $y_t$, $M_t$ and $V_t$ stand for Price level, real GDP, level of monetary stock and money velocity. There are two issue here may cause our original (difference) specification not appropriate: $E[\xi_t v_t], E[\xi_t v_{t+s}] \neq 0$. We explore this point in another paper Wang(2016). Here we merely replicate an exercise in Benati

\textsuperscript{19}These ad hoc criterion are justified from our Mont Carlo experiments.

\textsuperscript{20}Or be more precise $\beta$ should not stay too large either, we set $\beta < 1.4$, so $\frac{1}{\beta} < 0.7$
et al. (2016) and Lucas and Nicolini (2015), and extend it with $M_2$ data. Both paper try to use shoe cost to micro-found money velocity: money velocity is interpreted as frequency people go to the bank, interest rate serves as an opportunity cost of holding cash($M_1$). Benati et al. (2016) explores several co-integration tests to identify a long-run equilibrium relationship between real money balance and price under different specification. Benati et al. (2016) ignore the second equation in our specifications, focuses on the first equation and replaces $v_t$ as a function of interest rate. Functions to represent $v_t$ generally vary under different specifications.

We replicate Baumol-Tobin specification in Benati et al. (2016). We plot money balance over GDP against interest rate with fitted “long-run money demand” in Figure-6, upper panel plots the case of $M_1$, down panel plots $M_2$. As one can tell from Figure-6, the long-run demand for money performs well to fit such long historical data(1914-2015), points scatter tightly along the fitted curves. However, if we color different era with different colors, for example, in the upper panel we color the points post-1984 out, in the down panel we color the points post-1991 out, these points lie under the fitted money demand, as demonstrated in Figure-6. An eye ball metric can tell that the variation of data after 1984 is too little to justify the validity of this long-run demand function. The same breaks down of “long-run” demand apply to $M_2$ after 1991 as well.

This running off track is confirmed by a cointegration test. Intuition behinds cointegration is to identify a persistent relationship, under which a combination of $I(1)$ processes become stationary, or the residual will be stationary. In Figure-7, we plot out residuals of our estimated cointegration relationship\textsuperscript{21}. One can easily tell that after some date, residuals become negative uniformly: the points cannot be justified by historical (reasonable) track. The long-run demand for money fits long historical data well. But recent decades seems to be a totally different era.

Cross countries Robust Check

We summarize the empirical results here. Data and figures are reserved in Appendix. In robust check part, we investigate Australia, Canada, Denmark, France, Germany, Italy, \textsuperscript{21}We adopt Phillips and Hensen’s fully modified OLS estimates.
Netherlands, Norway, Spain, Sweden, Switzerland, United Kingdom.

QTM holds, at least for a while, for most of the countries under our investigation. However, QTM is not a universal law. For example, QTM never exists in Netherlands, France and Germany. Among those used to have QTM, breaking down of QTM happens recently in U.S, Sweden, Denmark, Spain, Switzerland, Australia and Canada at different dates, which do not coincide with their starting dates of disinflation policy. If we take into account the fact that QTM never exists in Netherlands, France and Germany. We can conclude that the recent breaking down of QTM is robust.

**Proposed Channel: Endogenous generated money and financial innovation**

As previous evidence shows, the recent breaking down of QTM is robust. To nail down our problem, we focus on the case of U.S: Quantity theory of money works for a while in U.S(1955 to mid-1980s(early-1990s)), then it breaks down. As stated previously, we choose to focus on the demand part of Money and try to find a micro-foundation of changing money velocity.\(^{22}\)

\(^{22}\)In Wang (2016), I explore the case in which central bank can affect money velocity
The demand for money or the incentive of holding money balance come from transaction purpose. And this is the reason why Lucas (1980) and Lucas and Nicolini (2015) adopt $M_1$ and similar indexes. However, during 1932-1954, (filtered) growth rate of $M_1$ has no explanation power to inflation ($\beta = -0.17, R^2 = 0.07$), while $M_2$ are valid after controlling income growth ($\beta = .91, R^2 = 0.76$)\(^{23}\). Furthermore, one should not contribute the uselessness of $M_1$ to Great recession period since $M_1$ cannot explain inflation variation during several years previous to 1929. However, it is hard to believe during 1932-1950, $M_2$ plays any role in implementing transactions. Though $M_2$ version of QTM breaks, it sheds lights on the direction we should investigate: $M_2$ is the major liability of depositary financial institution. On the asset side, loan plays a major role. In other words, instead of investigating deposit demand, we look at credit generation or Money creating process. $M_2$ stands for a major part of the liability of depositary system, namely banks. On the other side of the bank, there are kinds of assets, the most important of which are loans. Furthermore, money can be endogenously generated by loan issuing. When a loan got issued, the bank has facilitated a transaction, no matter how receivers of fund keep it in their checking or saving accounts. Experience of U.S during 1930-1955 justify this understanding, $M_2$ can explain inflation well. We choose to focus on loan issuing instead of deposit demanding, emphasizing that money is generated endogenously.

To decompose the loan and extract the most important components, we explore the data of flow of fund. For individuals, as Figure-8 demonstrates, the largest liability is from

\(^{23}\)This may serve as a reason why Sargent and Surico (2011) adopts $M_2$ index.
mortgage loan since the first available data point. For corporations, collateral loans are also important borrowing. Berger and Udell (1990) reports that nearly 70% of all commercial and industrial loans in the United States are secured by collateral assets. An important collateral asset for both small firms and large corporations is real estate. According to the S.5 of Z.1 tables provided by the Federal Reserve Board, during 1952Q1-2016Q4, averagely real estate represents 63% of the tangible assets held by non-financial corporate firms on their balance sheets. For the period from 1952Q1 to 2016Q4, tangible assets (the sum of real estate, equipment, and intellectual property products) average about 58% of total corporate assets. Furthermore, total corporate assets include financial asset; not all financial assets can serve as collateral, for example, firms’ foreign investment and kinds of deposit accounts. Each of them contributes 20% and 25% of financial asset. For non-farm noncorporate U.S. firms, averagely real estate accounts for 90% of tangible assets (which is in turn about 87% of total assets). And we can tell from Figure-?? commercial and industrial loan and mortgage loan plays a major role in loan issuing. Hence, it is now reasonable to propose that real estate purchase is a major component in money generating process.\textsuperscript{24}

To confirm this finding, we explore long-run data of house price during the pre-crisis period (S&P CoreLogic Case-Shiller Home Price Indices). In Figure-11, we plot raw data of house price index in the upper panel and filtered growth rate of house price with inflation and growth of $M_2$ in down panel. Again we adopt the filter of Lucas (1980). As one can tell from the figure, the filtered growth rate of this price index co-moves with money more closely and robustly, if one follows Lucas (1980) to plot growth rate of $M_2$ against house price, and run a QTM regression the $\beta(R^2)$ will be 0.89(0.93). In the time domain, the nominal price of real estate tracks money throughout the whole pre-crisis sample, including Lucas Period.

Then what happens during mid-1980s-early 1990s? Financial innovation. Under endogenous money generating framework, the structure of the financial market is supposed to have ignorable effects on the monetary phenomenon, including QTM. Furthermore, the well-known financial innovation process starts from the 1980s. Since the early 1980s, collateral-\textsuperscript{24} We even ignore the expenditure affected by house purchase, for example, home-related durables and home improvements sectors. See Benmelech et al. (2017).
eralized debt obligation (CDO) was introduced. Then residential mortgage-backed securities (MBS) and collateralized mortgage obligations (CMO) follow. Then credit default swaps follow. Meanwhile, a sequence of acts got passed to facilitate financial innovation, for example, Gramm-Leach-Bliley Act (1999), Commodity Futures Modernization Act (2000)\textsuperscript{25}. As a result, aggregate leverage ratio and the ratio of market value of residual land over GDP increase since the mid-1980s, as shown in Figure-10. In other words, the land becomes better and better collateral during the new era. Under the framework of endogenous money creating, the land generates more and more money, or more money and credit are generated because of land. QTM thus reveals itself in a different way.

To be more precise, as one will see in a model with financial innovation, for example, the model in next section, financial innovation promotes the real price of the asset, which will, on the other hand, generate more money endogenously. This fact will thus have a tremendous implication for QTM: more money is after the price of land. Furthermore, if people treat land and final goods are complementary (e.g., in-house production, one need to combine final goods with land), the higher relative price of land means more money will be generated by land but not final goods. Or put in another way, less money is chasing or generated from final goods. Narrowly defined QTM thus breaks down.

\textsuperscript{25}GLB Act allows bank holding companies to own other financial companies. CFM Act moves regulation out of OTC market.
Model outline

In this section, we would like to embed a (broader) money demand(generating) function(process) into a framework with financial innovation to explain collapsing QTM for U.S during Pre-crisis period. To model financial innovation, first we would like to outlay our economy in a frictional environment, then financial innovation improves agents’ situation. The friction we choose is borrowing limit, which is a fundamental feature of our financial system. We thus adopt a framework with endogenous borrowing limit. A critical feature of this limited borrowing framework is that borrowers are subject to an endogenous borrowing limit, which is itself endogenously determined by the status of the economy. This feedback loop has already been explored in financial friction literature, such like Kiyotaki and Moore (1997), Jermann and Quadrini (2012), Bianchi and Mendoza (2010) and Bianchi et al. (2012). A binding borrowing constraint will be reflected on economic status, which will further affect borrowing limit. In other words, this feedback loop is two-sided: economy status and borrowing limit are determined simultaneously.

For example, under the collateral constraint we are adopting, the degree of consumption smoothing is positively linked to the market value of collateral since people can borrow more when the collateral value is high. Furthermore, as people can have a more stable consumption path, price of collateral will also increase and become higher than its fundamental value. Vice versa, if the price of collateral decreases, borrowing limit will shrink accordingly. People then will face a “sudden stop” and reduced consumption. Lower consumption will further dampen the price of collateral.

We model financial friction as a process alleviating this borrowing friction; people are thus able to keep a higher leverage ratio if they are willing to. To make this innovation endogenous, we model financial innovation as a learning process, following Boz and Mendoza (2014). The main mechanism is that haircut of collateral is random, following a Markovian regime changing process(2 states Markovian process). However, agents in our economy have no perfect information about this regime-changing probability. They are Bayesian learners. Conditional on new observation, the agent will update her subjective belief, optimal planning will be then carried on accordingly. To make our analysis computation efficient, we follow
We will first lay out the basic setup, then briefly go through the numerical method to solve it. A calibration section then follows.

**Basic Setting: Real Economy**

Consider a economy in infinite discrete time $t = 1, 2, 3, \ldots$, there is a unit measure of agents. Agents face a stochastic endowment flow: she will receive $y_t$ (perishable) every period, where \( \{y_t\} \) follows a Markovian process. Agents act atomistically in a competitive market and value consumption \( \{c_t\}_{t=0}^{\infty} \) according to a standard time-separable expected utility function as below. The normal assumption on time discounting rate and shape of utility function applies.

$$E\left[ \sum_{T=0}^{\infty} \beta^T U(c_t) \right]$$

There is a risk-neutral bank, and it trades one-period non-state-contingent discount real bonds $b_t$ with the economic agents. We assume the bank is willing to hold any collateral asset directly, e.g., because of asymmetric information. To simplify our setting, we further fix the price of one period bond as $1/R$, where $R$ is the real interest rate. As explained
in Sargent and Ljungqvist(2004), we need restriction $R\beta < 1$ to ensure the existence of a well-defined long-run distribution of borrowing. Furthermore, this real interest rate $R$ is exogenous to our economy. For domestic agents, one-period real risk-free bond and land are two kinds of trade-able assets. Later we will allow a role for cash. Furthermore, land can serve as collateral. The period budget constraint is thus:

$$c_t + b_{t+1}/R + l_{t+1}q_t \leq b_t + l_t(d_t + q_t) + L_t$$

where $l_t$, $q_t$ is the land holding and price of land in period $t$, $d_t$ is the dividend flow from land and $L_t$ is labor income. We divide the total production $z_t f(l_t)$ into dividend(rent) from land, $d_t$ and $L_t$. $z_t$ follows a Markovian process ($z_t \in Z_t = \{z_1, z_2, \ldots, z_N\}$). To emphasize the effect of financial innovation, we assume that people has perfect statistical information about $\{z_t\}_t$ and land is not reproducible and available stock is normalized to 1. Hence it is straightforward to state that $d_t = z_t f'(1)$ and $L_t = z_t[f(1) - f'(1)]$.

Under this incomplete market framework, we add in an endogenous collateral borrowing constraint:

$$b_{t+1} \geq -\psi_t - \phi_t q_t l_{t+1}$$

And as we stated previously, stochastic process of $\psi$ and $\phi$ follows a “true” Markovian process. For simplicity, we adopt a binary support set. Namely, $(\phi_t, \psi_t)$ can take two value $(\phi_h, \psi_h)$ and $(\phi_l, \psi_l)$, where $1 > \phi_h > \phi_l$ and $\psi_h > \psi_l$. The learning process will be elaborated in later in this section. It is straightforward to extend our set into a setup with multiple states.

Furthermore, it is worth pointing out that $\phi_t$ is enough to represent the status of the financial structure if one only concerns about collateral borrowing. Moreover, we would like to give a role to the evolving credit payment system, as pointed out in Wang (2016). Figure-12 is borrowed from Wang (2016), plotting the time path of ratios between consumer credit and $M_1$(wealth). Furthermore, credit card balance is the major component in consumer credit.
credit. From the figure, one can tell that credit card keeps crowding out $M_1$ or money to implement a transaction. This is another reason we prefer to honor the role of loans and do not want to merely focus on cash or $M_1$. Here we use $\psi_t$ as a shortcut\textsuperscript{27} to represent this “non-collateral” borrowing.

Let $\mu_t$ denote the Lagrange multiplier of the collateral constraint, the Euler Equations for $b_{t+1}$ and $l_{t+1}$ will be

$$U'(c_t) = E^{B_t} [\beta RU'(c_{t+1})] + \mu_t$$

$$q_t = \frac{E^{B_t} [\beta U'(c_{t+1})(q_{t+1} + d_{t+1})]}{U'(c_t) - \phi_t \mu_t}$$

Where $E^{B_t} [...]$ represents the expectation according to belief at period $t$, $B_t$. We mark $B(.)$ with subindex $t$ to emphasize belief is evolving.

Equilibrium condition of this model consists of Euler equations, budget constraint, collateral borrowing, complementary slackness condition, and unit normalized stock of land:

\textsuperscript{27}This is a shortcut for limited commitment setup to justify the existence of unsecured credit, such as in Azariadis et al. (2016) and Wang (2016)
Collateral constraint brings a feedback loop into our setup. To see how feedbacks work, it is worth emphasizing that one can express policy functions of consumption (by complementary slackness condition) as the minimum of two branches, one denotes unconstrained case and another for constrained.

Through this formula, it is easier for one to understand the feedback between financial structure and consumption: binding agents choose a lower level of consumption than unconstrained case (so min function). Furthermore, reduced consumption will dampen asset price, which is a critical component in collateral borrowing, since in equilibrium asset holding is normalized to be 1. To make this observation more explicit, we take a difference between the two Euler Equations of risky and risk-free rate:

\[ 1 - \mu_t = E^{B_t}[\beta U'(c_{t+1})(R^q_{t+1} - R)] \]
\[ E^{B_t}[R^q_{t+1} - R] = -\frac{\text{Cov}(U'(c_{t+1}), R^q_{t+1}) - (1 - \phi_t)\mu_t}{E^{B_t}[U'(c_{t+1})]} \]

Except \((1 - \phi_t)\mu_t\) term, expected excess return follows a classical formulation:

Payments from a good-hedging security positively correlate with pricing kernel: pays better when people value consumption more. It offers insurance to some degree; people will thus require lower (future) excess return or price of a good hedging security is high. On the other hand, a binding borrowing constraint directly increases excess return, because of positive \(\mu\), e.g., Equation-0. In another word, the price of the asset will be dampened: according to Campbell and Shiller decomposition, higher expected future excess return means lower current asset price. To make this statement clearer under our setup, one can write \(E^{B_t}[R^q_{t+1}] \equiv \frac{E^{B_t}[(q_{t+1} + d_{t+1})]}{q_t}\) recursively (forward) to express price of asset as a

---

29 We present Campbell and Shiller as following, which one will tell that it is a little different with our
discounted value of a dividend sequence as following. And the discounting rate is general different with risk-free rate unless investors are risk-neutral:

\[ q_t = \frac{E^{B_t}[d_{t+1} + q_{t+1}]}{E^{B_t}[R_{t+1}^q]} \]

\[ = E^{B_t}[(d_{t+1} + q_{t+1}) \frac{1}{E^{B_t}[R_{t+1}^q]}] \]

\[ = E^{B_t}[(d_{t+1}) \frac{1}{E^{B_t}[R_{t+1}^q]}] + E^{B_t}[(d_{t+2} + q_{t+2}) \frac{1}{E^{B_t+1}[R_{t+2}^q]}] \]

\[ = E^{B_t}[(d_{t+1}) \frac{1}{E^{B_t}[R_{t+1}^q]}] + E^{B_t}[(d_{t+2} + q_{t+2}) \frac{1}{E^{B_t}[R_{t+1}^q]E^{B_t}[R_{t+2}^q]}] = ... \]

\[ = E^{B_t} \left[ \sum_{k=0}^{\infty} \frac{1}{\prod_{i=0}^{k} E^{B_t}[R_{t+1+i}^q]} \right] d_{t+1+k} \] (**)

From the formula (**), it is clear that higher future expected the excess return, which may be caused by binding borrowing constraint, positive \( \mu \), can dampen the price of the asset. Furthermore, endogenous borrowing constraint makes land an even worse hedging against consumption through an indirect channel: when constraint binds, higher realized return means higher land price, which further implies higher borrowing limit, better consumption. Hence covariance between consumption and realized return would be higher than the cases without borrowing limit. Reduced asset price will further trigger tighter binding of borrowing constraint again. Mathematically, \(|\text{Cov}(U'(c_{t+1}), R_{t+1}^q)|\) will be larger than frictionless case.

Furthermore, formulation (***) give us another self-full filling channel, especially under a framework with endogenous belief. When people expect borrowing constraint to be bind-

\[ R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} \]

\[ r_{t+1} = \ln(R_{t+1}) = p_{t+1} - p_t + \ln(1 + \frac{D_{t+1}}{P_{t+1}}) = p_{t+1} - p_t + \ln(1 + \exp(d_{t+1} - p_{t+1})) \]

\[ = p_{t+1} - p_t + \ln(1 + \exp(d - p)) + \frac{\exp[d - p]}{1 + \exp(d - p)}(d_{t+1} - p_{t+1} - (d - p)) \]

\[ = p_{t+1} - p_t + k + (1 - \rho)(d_{t+1} - p_{t+1}) \]

\[ \Rightarrow \]

\[ p_t = \rho p_{t+1} + k + (1 - \rho)d_{t+1} - r_{t+1}, \quad \rho = \frac{1}{1 + \exp(d - p)} \]

One can tell from the last equation (Campbell-Shiller decomposition) that \( p_{t+1} \) is decreasing in future return \( r_{t+1} \). Moreover, one can apply the last equation forward recursively to represent \( p_t \) as a discounted value of future dividends.
ing in the future, this expectation will also reduce asset price today, since it will increase $E^{B_t}[R_{t+k}]$ for some $k$. To focus on the purpose of this article, we will leave this sun-spot equilibrium exploration to future research.

To make financial innovation as a smoothly progressive process\(^{30}\), we model it as an updating optimistic belief through Bayesian learning channel and adopt equilibrium computation algorithm in Cogley and Sargent (2008). Bayesian learners update their belief to take new data points into account according to Bayesian law\(^{31}\). In our setup, condition on $(\phi_{t-1}, \psi_{t-1}) = (\phi_h, \psi_h)$, after the realization of $(\phi_t, \psi_t)$, agents update transition probability of period and form a new belief $B_t$, which can be summarized in $Pr((\phi_{t+1}, \psi_{t+1}) = (\phi_h, \psi_h) | (\phi_t, \psi_t) = (\phi_h, \psi_h))$.

Since belief itself evolves, a complete computation is time-consuming. Anticipated Utility approach in Cogley and Sargent (2008)\(^{32}\) adopts a simplifying computation strategy(assumption) that after observing every new sample, agents update their belief and choose consumption, investment and borrowing decision according to this updated belief. In another word, they ignore the possibility that their belief will change in next period. Moreover, to focus on financial sector part, we assume $(\phi_t, \psi_t)$ can be observed directly without any noise. Under a framework with observation error, it is straightforward to embed a (forward) Kalman filter process in this setting to “filtering” noise. Our numerical computation consists of two steps: (1) After setting the prior belief; we simulate a sequence of observation. Agents’ posterior belief $\{B_{t=1}\}$ will be generated according to Bayesian Law. (2) For any period $t$, given belief $B_t$, agents will make their optimal decision according to this belief. Put it in another way; we solve the model as if belief $B_t$ will last forever. Policy functions in any period-$t$ are time specific, only valid in the particular period. In next subsection, we elaborate how people’s belief is updated.

\(^{30}\)Under a smoothly progressive process, real shocks still have a place. Hence, our quantitative results takes care of shocks from real economy.

\(^{31}\)Generally, Bayesian learner has a prior belief first, say $q(\theta)$, where $\theta$ is the parameter learner feel interested in. For illustration purpose, we assume that $q(\theta), \theta \in \Theta$ is a probability density function, $\Theta$ is the support set of parameters. Bayesian learning still apply with any well defined random variable, not matter its probability density function exists or not. Posterior belief about $\theta$ will be updated according to $Pr(\theta | y_t) = \frac{Pr(y_t | \theta) q(\theta)}{\int_{\Theta} Pr(y_t | \theta) q(\theta) d\theta}$. Since the denominator is complicated in most cases, but it is a constant. In most analysis, researchers write $Pr(\theta | y_t) \propto Pr(y_t | \theta) q(\theta)$ and use MCMC algorithm to sample this posterior distribution. For more details on MCMC, see Chib (2003)

\(^{32}\)Cogley and Sargent (2008) gave the credit to Kreps(1998).
Bayesian Learning

Since agents have no perfect information about the transition probability of financial structure parameters $F_t = (\phi_t, \psi_t)$. We further simply assume that there are only two states of $F_t$, hence learning process degenerates to learn the transition probability $p = Pr(F_t = H|F_t = H)$ and $q = Pr(F_t = L|F_t = L)$. We can write the transition probability matrix of $F_t$ (a function of time) as:

$$
\Pi = \begin{bmatrix}
p & 1 - p \\
1 - q & q
\end{bmatrix}
$$

Furthermore, we set prior of $p, q$ are independent beta distribution. Posterior belief on $p, q$ will be beta distribution too: $p_t \sim \beta(n^{HH}_t, n^{HL}_t)$ and $q_t \sim \beta(n^{LL}_t, n^{LH}_t)$ where $n^{ij}$ is the number of observation of transition from state $i$ to state $j$. For example, $n^{LH}_t - n^{LH}_0$ denotes the numbers of observation of shrinking lending. Counters $N_0 = \left[ n^{HH}_0, n^{HL}_0, n^{LL}_0, n^{LH}_0 \right]^T$ summarize prior belief, and counter will be updated afterward after any new data sample is observed, these counters $N_t$ is updated as

$$
N_{t}^{ij} = \begin{cases}
N_{t-1}^{ij} + 1 & \text{if } F_{t-1} = i \text{ and } F_t = j \\
N_{t-1}^{ij} & \text{otherwise}
\end{cases}
$$
It is clear that $F_t$ and $(F_t, N_t)$ is a Markovian process. For our purpose, we do not need to derive out the full transition matrix of $(F_t, N_t)$. We only care about $Pr(F_{t+1} = h|F_t, N_t)$, it can be calculated as

$$Pr(F_{t+1} = H|F_t, N_t) = Pr(F_{t+1} = H|N_t) = \begin{cases} 
\int_0^1 pf(p|N_t)dp & \text{if } F_t = H \\
\int_0^1 (1-q)f(q|N_t)dq & \text{if } F_t = L
\end{cases}$$

The first equality is because $N_t$ summarizes the information of $F_t$. And $N_t$ is also evolving as previously stated. Hence the belief $B_t$ can be summarized as

$$\Pi_t = \begin{bmatrix}
\int_0^1 pf(p|N_t)dp & 1 - \int_0^1 pf(p|N_t)dp \\
1 - \int_0^1 qf(q|N_t)dq & \int_0^1 qf(q|N_t)dq
\end{bmatrix} = \begin{bmatrix}
\frac{N_{HH}^{fLH}}{N_{HH}^{fLH} + N_{HL}^{fLH}} & \frac{N_{HL}}{N_{HH}^{fLH} + N_{HL}^{fLH}} \\
\frac{N_{HH}^{fLH}}{N_{HH}^{fLH} + N_{HL}^{fLH}} & \frac{N_{HL}}{N_{HH}^{fLH} + N_{HL}^{fLH}}
\end{bmatrix}$$

The first equality comes from Bayesian learning; the second follows properties of the beta distribution. For prior belief, we set $N_0 \approx 0$, equivalently, $p_0 = q_0 = \frac{1}{2}$, this prior implicitly assume that agents have no prior information about the financial innovation process. Thus no previous transition was observed, $N_0 \approx 0$.

As Cogley and Sargent(2008) show, this updated belief will asymptotically converge to the true process, a numerical example is reserved in Appendix. In previous sections, the expectation operator bases on belief $B_t$, which now can be summarized by $\Pi_t$ and transition matrix of $z_t$. 

Figure 13: Structure replacement cost over GDP, Quarterly data, 1975Q1-2016Q4, source: Davis and Heathcote (2007)
Endogenous Money Creation: Loan is facilitating transactions

As we argued in previous sections, transaction-oriented money demand is not robust enough to justify QTM: $M_1$ is not event a robust index, co-moving with inflation on a one-to-one basis. We can understand $MV = PY$ in another way; money is generated by credit issuing: bank issues loans with the amount proportional to the value of goods, “money” here can be unsecured credit, e.g., credit card, or collateral loans. Surely, and there is a cost associated with credit, more precisely, there is a yield curve for credit. Hence, $V$ should represent the level and slope of this yield curve: as interest increases more with longer maturity, agents tend to keep loan shorter and trade more frequently. This is how we interpret velocity under our this money creation process. Along with a stationary interest rate assumption, we view $V$ as a constant in the long run. In another word, we assume the amount of money generated is proportional to the value of goods, and now we have more goods to take into consideration. Surely, this is merely a simplification to model money creation.

To make this simplification aligned with our previous data analysis, we propose a setup with loans, where loans can also implement a transaction. Furthermore, there is another transaction: house purchasing. We make canonical assumptions on this Monetary economy: people are working but only get paid for the units of cash after all consumption, borrowing (repayment) and asset investment transaction happened. People cannot enjoy output from own labor; they have to purchase from other people. There is no monitoring system except the bank. People can use cash or borrow from the bank to implement a transaction. Of course, the maximum credit is restricted by collateral constraint. The Central bank will use a helicopter to drop cash at the very beginning of every period. House is necessary for a consumption process to take place, no further utility it will offer. Homogeneous houses are offered by a competitive real estate company. We adopt this timeline because of liquidity of collateral asset.

Besides cash, all transactions can be implemented by borrowing too. Borrowing itself is

---

33\textsuperscript{This is the simplest timeline to solve. One interesting modification to this timeline is to allow asset market open after all transaction happens. In other words, under our original timeline, the only output from labor need to be transacted through money, and there is no inflation risk for asset holding. Under this modified timeline, collateral asset faces inflation risk.}
still under collateral constraint. And we will call the balance of cash and “money” created by borrowing as broad money, $M_2$ for example.

Concerning to house, we assume competitive real estate companies\textsuperscript{34} are to max$_{c_H,l}[c_H^\rho + l^\rho]^{\frac{1}{\rho}}$, $\rho < 0$ and $pc_H + pq_l \leq \text{Cost}$. Real estate company, owned by economy agents, will offer the houses to households at its cost: $p_t[1 + q_t^{\rho_{-1}}]^{\frac{c_H^\rho}{\rho}}$, the total quantity of house is $[c_H^\rho + l^\rho]^{\frac{1}{\rho}}$, where $c_H$ is the amount of final goods embedded in-house, i.e. house structure. Furthermore, from data we can tell that structure-replacement cost is almost a constant fraction of GDP, as documented in Figure-13. We have no government expenditure and private investment in our model; we thus restrict $c_H$ is a fixed fraction of total consumption, $\theta$ exogenously: more consumption needs a larger house. To make our micro-foundation simple, we assume that factory are all built underground, and a house is built on the ground. Since agents act atomically, they take the price and size of house as given:

\[
\frac{M_t + G_t}{p_t} + b_t + s_t(d_t + q_t) \geq c_t + [1 + q_t^{\rho_{-1}}]^{\frac{c_H^\rho}{\rho}}[c_H^\rho + l^\rho]^{\frac{1}{\rho}} + \frac{b'}{R} + s_{t+1}q_t
\]  

(10)

Where $\bar{c}_H$ and $\bar{l}$ is the final goods and land used in house construction, people take them as given. $G_t$ is cash transfer from government. But at equilibrium: $\bar{l} = 1$, $\bar{c}_H = \theta c$ and $\bar{c}_H + c = y$. From this equation, we embed in the pattern we found in data: right hand-side -Loan(broad money) is issued(generated) to purchase real estate, left hand-side-broad money can be generated by using collateral.

Furthermore, new budget constraint will become:

\[
c_t + b_{t+1}/R + l_{t+1}q_t + \frac{M_{t+1}}{p_t} + [1 + q_t^{\rho_{-1}}]^{\frac{c_H^\rho}{\rho}}[c_H^\rho + l^\rho]^{\frac{1}{\rho}} \leq b_t + l_t(d_t + q_t) + L_t + \frac{M_t + G_t}{p_t} + \Pi_t
\]

where $\Pi_t = 0$ is profit from real estate company. Again with borrowing constraints, our

\textsuperscript{34}Real Estate Companies are owned by households.
equilibrium condition consists of

\[ U'(c_t) = E^{E_t}[\beta RU'(c_{t+1})] + \mu_t \]  
(1)

\[ q_t = \frac{E^{E_t}[\beta U'(c_{t+1})(q_{t+1} + d_{t+1})]}{U'(c_t) - \phi_t \mu_t} \]  
(2)

\[ c_t = b_t + z_t f(1) - b_{t+1}/R \]  
(3)

\[ b_{t+1} \geq - (\psi_t + \phi_t q_t l_{t+1})|_{l_{t+1}=1} \]  
(5)

\[ M_t^b = p_t (1 - \theta) c_t + p_t[(\theta c_t)^\rho + l^\rho]^{\frac{1}{\rho}}[1 + q_t^{\frac{\rho - 1}{\rho}}]^{\frac{\rho - 1}{\rho}} \]  
(6)

Where \( M_t^b \) represents the stock of broadly defined money including credit.

The main mechanism can be understood through a simplification to Equation-(6). We simplify Equation-(6) to be \( M_t \approx p_t[(\theta c_t)^\rho + l^\rho]^{\frac{1}{\rho}}[1 + q_t^{\frac{\rho - 1}{\rho}}]^{\frac{\rho - 1}{\rho}} \), since consumption value is relatively small. We use this specification to emphasize the channel through which QTM breaks. It is straightforward to see that from this simplified money generating function:

\[ \frac{M_{t+1}^b}{M_t^b} \approx \frac{p_{t+1}[1 + q_{t+1}^{\rho}]^{\frac{\rho - 1}{\rho}}}{p_t[1 + q_t^{\frac{\rho - 1}{\rho}}]^{\frac{\rho - 1}{\rho}}} \]

From this equation, one can tell growth of broad money is not merely associated with inflation of final goods \( \frac{p_{t+1}}{p_t} \), but also with real price inflation of the collateral asset \( \frac{q_{t+1}}{q_t} \).

This intuition applies to our original setup too. Not only money is necessary to implement more kinds of transaction, but money can also be generated through (collateral) borrowing. Empirically, to estimate the effect of money growth onto inflation under this framework, one can take a log-difference and read the formula reversely, in a regression form: regress inflation onto growth rate of money. To be more explicit, from period \( t \) to \( t + N \) the coefficient will be:

\[ \beta_{t,t+N} = \frac{[1 + q_t^{\rho}]^{\frac{\rho - 1}{\rho}}}{[1 + q_{t+N}^{\rho}]^{\frac{\rho - 1}{\rho}}} \]  
\[ \rho < 0 \]

This is a formula nesting canonical QTM and broader QTM: when the relative price of collateral asset stays stable \( \frac{q_{t+1}}{q_t} \approx 1 \), as what happens during Lucas’ Period 1955-1980, \( \beta \) will be close to 1. When relative price of land is increasing \( \frac{q_{t+1}}{q_t} > 1 \), as more money is
generating from or chasing after asset; we will get a $\beta$ significant less than 1. The reason is that more money is generated by (because of) collateral asset instead of final goods under financial innovation era.

To solve this model at period $t$, we take $B_t$ as given, as if this belief will last forever, all usual optimization process goes through, we will get $b_{t+1}(b_t, z_t, F_t)$, $c_t(b_t, z_t, F_t)$ and $q_t(b_t, z_t, F_t)$ to solve equation-(1) to (6). The solution, policy functions $b_{t+1}(b_t, z_t, F_t)$, $c_t(b_t, z_t, F_t)$ and $q_t(b_t, z_t, F_t)$ thus merely determine the optimal plan for period $t$. People will act according to these policy function only during period $t$. People will have another group of policy functions in following periods. Furthermore, different realization paths of $\{z_t\}$ and $\{(\psi_t, \phi_t)\}$ will form different belief, sequence of policy function will be thus different accordingly.

To summarize, financial innovation is a process in which asset becomes better collateral, alleviating borrowing friction in this economy. To have a progressive process, leave a role for real shocks, we deviate from rational expectation framework and model financial innovation as a learning process, following Boz and Mendoza (2014) and Cogley and Sargent (2008). The financial innovation is a process in which people learn that their asset become better collateral, at least they believe so. One can skip this learning and exogenously change the Markovian process governing parameters $(\psi, \phi)$ in financial innovation regime. According to our calculation, the QTM implications out of these two methods are very similar: both processes generate increasing sequence of real asset price, narrowly defined QTM breaks.

**Numerical Algorithm and Calibration**

In this section, we will go through our numerical method briefly in the first part. Detailed procedure can be found in numerical algorithm Appendix. Then we calibrate our model to match U.S economy and financial innovation process. In the end, we will use our calibrated model to simulate a (bunches of) sequence(s) data and use the simulated data to run a QTM regression. As one will tell, we will have a slope significantly less than 1. Since we have more than one simulation, we have a distribution of the loading of Money onto inflation, $\beta$. 36
Numerical Solution Method

As we can tell from equilibrium conditions Equation-(1) to (6). The first four equations can be solved by policy functions $\mu_t(b_t, z_t, F_t)$, $b_{t+1}(b_t, z_t, F_t)$, $c_t(b_t, z_t, F_t)$ and $q_t(b_t, z_t, F_t)$. Given a exogenous money supply, $p_t(b_t, z_t, F_t)$ can be solved from the last condition. In this section, we focus on explaining the idea of how to solve the first four equations.

We combine the implication of complementary slackness condition and policy function iteration with endogenous grids. One can still use a usual value function iteration procedure to solve it. From slackness condition, we can have following equation-(\Theta). And We will use policy function iteration to solve equations-(2) to (5) with equation-(\Theta).

$$c_t = \min \{(U')^{-1}(E^{B_t}[\beta RU''(c_{t+1})]), \psi_t + \phi_t q_t s_{t+1} + b_t + z_t g(1)\} \quad (\Theta)$$

Where $c_t$ will take the left branch if collateral constraint is loose ($\mu_t = 0$) since $U'(c_t) = \beta RU'(c_{t+1}) + \mu_t$, and take the right branch if collateral constraint is binding ($\mu_t > 0$, then consumption should be solved from budget constraint\textsuperscript{35}). When collateral constraint is loose $\mu = 0$, normal pricing kernel applies and asset pricing formula goes through. Otherwise, when collateral constraint is binding, $c_t$ will be determined by $\psi_t + \phi_t q_t s_{t+1} + b_t + z_t g(1)$. In this case, binding constraint $b_{t+1} = -\psi_t - \phi_t q_t$ gives us an asset price. Then we can explore asset pricing equation-2 to recover consumption policy function. Policy function $\mu$ follows equation-1.

We then explore theoretical characteristics of solution: for every given $z_t$ and $F_t$, there is a threshold $\bar{b}$ such that, when $b_t > \bar{b}$ collateral constraint goes slack, when $b_t < \bar{b}$ collateral constraint goes binding, and when $b_t = \bar{b}$ collateral constraint is marginally binding. Hence $\bar{b}$ is the value of $b_t$, at which $c^{binding}(b_t, z_t, F_t) = c^{slack}(b_t, z_t, F_t)$, $q^{binding}(b_t, z_t, F_t) = q^{slack}(b_t, z_t, F_t)$, and $\mu^{binding}(b_t, z_t, F_t) = 0$.

On the top of all these procedures, we simulate a realization of $F_t$ from “True” process. People will then update their belief $B_t$ and combine it with knowledge of process $\{z_t\}$ to form the expectation, from which our optimization process starts. More details are left in Numerical algorithm Appendix.

\textsuperscript{35}This structure can be explored more in a more general context, see Korinek and Mendoza (2014)
Figure 14. The path of consumption, borrowing and asset price under one realization path of \( \{z_t\} \) with simulated Belief against Financial Structure Parameters \( F_t \)

Calibration

First, utility function and production technology are specified to be

\[
U(c_t) = \frac{c_1^{1-\gamma} - \gamma}{1-\gamma}, \quad f(l_t) = l_t^\alpha.
\]

We need to set the value of \( \gamma, \alpha, \rho, \phi^H, \phi^L, \psi^H, \psi^L \), and prior belief \( N_0 \) as long as the process of \( \{z_t\} \).

We calibrate the model to U.S during 1984-2006, excluding the crisis period. The date \( t = 1 \) is set to 1984, \( t = T \) is set to be 2006 since it is believed that 2007 is the start of financial crisis. We will then set \( F_t = H \) during \( t = 1 \) to \( t = T \). Learning period is 23 years.

The value of \( \gamma, \alpha, \rho, \phi^L, \psi^L \) will be set to match U.S annual data 1955-1984. And there is only one regime for financial structure for the pre-innovation period. The real interest rate is set to 2% since the average annual real return of three months and one year bill between 1984-2006 is 1.84% and 2.39%. TFP \( \{z_t\} \) is approximate use a log-AR(1) process, namely, \( \ln(z_t) = \theta \ln(z_{t-1}) + \epsilon_t \). We adopt Tauchen Method to discretize this AR(1) process. Routinely estimation give us value of \( \theta = 0.98 \) and \( \sigma_e = .01575 \). For \( \gamma \), we set it equal to 5, a middle value of the acceptable range from Mehra and Prescott(1984). We set the mean of output to be 1.

Furthermore, during pre-1984 era, residual land accounts for 43% GDP(Davis and Heathcote(2007) Dataset) and borrowing from abroad accounts for 29.7% GDP. Hence, we set
ψ_L = 0.297, φ_L = 0.0001, φ_H = 0.1 and ψ_H = 0.35 to match the leverage ratio at the end of 2006. We set β = 0.95 and \( N_0 = N_{0}^{HH} = N_{0}^{HL} = N_{0}^{LH} = N_{0}^{LL} = 0.015 \) (Boz and Mendoza (2014)). Furthermore, consumption contributes 60.5% of GDP during 1955-1984. Since there is no government and private investment here in our model, we assume there is a lump-sum loss \( A \), budget constraint is \( c + \frac{\nu}{R} + A = b + 1 \), hence \( A = 1 - 0.605 - 0.297(1 + 2\% - 1)/(1 + 2\%) = 0.389 \). And the true process of \( F_t \) is not quite important for our purpose, since our calibration process. The calibrated parameter is summarized in Table-1. The prior belief parameter is critical. For example, Boz and Mendoza (2014) set this parameter to match option-adjusted spread on Fannie Mae RMBS with 30-year maturity over T-bill. Under our set-up, \( N_t \) counts the changes. Since there is no regime change before the mid-1980s, it is no-harmful to set a small \( N_0 \approx 0 \). We take no stands on which excess return we should match; we thus implement numerical experiments with \( N_0 = 0.01, 0.005, 0.001 \). Our results are robust to this parametrization.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>Discounting Factor</td>
<td>.95</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Relative Risk Aversion Coefficient</td>
<td>5</td>
</tr>
<tr>
<td>( A )</td>
<td>lump-sum loss</td>
<td>0.389</td>
</tr>
<tr>
<td>( \rho )</td>
<td>Elasticity of substitution between land and structure</td>
<td>-2</td>
</tr>
<tr>
<td>( \sigma_e )</td>
<td>Standard deviation of Log-TFP</td>
<td>.01575</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Persistence of Log-TFP</td>
<td>.98</td>
</tr>
<tr>
<td>( L )</td>
<td>Land supply</td>
<td>1</td>
</tr>
<tr>
<td>( \psi_L )</td>
<td>Financial structure</td>
<td>0.297</td>
</tr>
<tr>
<td>( \psi_H )</td>
<td>Financial structure</td>
<td>0.35</td>
</tr>
<tr>
<td>( \phi_L )</td>
<td>Financial structure</td>
<td>.0001</td>
</tr>
<tr>
<td>( \phi_H )</td>
<td>Financial structure</td>
<td>.1</td>
</tr>
<tr>
<td>( N_0 )</td>
<td>Prior Belief</td>
<td>.015</td>
</tr>
<tr>
<td>( R )</td>
<td>Annualized interest rate</td>
<td>1.2</td>
</tr>
</tbody>
</table>

*Table 1: Calibration Parameter*
Quantitative findings

In this section, we first discuss the path of borrowing, consumption and land price. Then we apply our QTM regression to this simulated data. As one can tell from Figure-16, asset price is higher under a better financial structure.

In this subsection, we simulated 10000 sequences. For each seed, we draw \( \{ z_t \} \) from the Markovian process under the parameters in the previous section. \( b_0 \) will be set at \(-0.3\), following Boz and Mendoza (2014). Agents’ belief \( \{ B_t \} \) towards \( F_t \) will be the same for each seed. The belief is formulated under the prior in the previous section and a sequence of \( F^H \) realization. Transition matrix of \( F_t \) is a 2-by-2 matrix; it thus can be summarized by two parameters \( \phi_{HH} \) and \( \phi_{LL} \), which denote the probability of future(one period later) parameter stay in high-regime or low-regime conditional on current realization is high or low. We plot this transition matrix in Panel D of Figure-14.

For illustration purpose, we pick up a path of \( z_t \) and plot out the path of consumption, borrowing and asset price in Figure-14 (implied \( \beta = 0.5745 \)). As one can tell from Figure-14, as optimism builds up, consumption increase, while borrowing keeps increase though then converge to a steady state. Furthermore, the Land price keeps going up during the whole period. Under some parametrization of “true” process, people borrow too much. This over-borrowing is a result of non-rational expectation and will make the sudden stop more damaging. We can tell that consumption excesses new steady state 7%, at the peak, then converges to steady state quickly (less than ten periods). However, the land price keeps increasing during the whole sample period.

Thus, we then calibrate an ARMA(2,2) process\(^{36}\) to match \( M_2 \) growth of U.S., And we will use equation-5 to solve out the price. Then we run QTM regression to estimate the effect of Money growth onto inflation; we focus on income adjustment here, the results without income adjustment have very close distribution. Under our baseline parameter, from our 10000 simulations, slope estimator \( \beta \) of QTM is 0.572, as plotted in Figure-15, where we plot out the simulation distribution of \( \beta \). As one can tell from Figure-15, our money loading ranges from 0.54 to 0.6, significantly less than 1. Hence our mechanism and model can offer an alternative explanation to collapsing of QTM.

\(^{36}\)ARMA(2,2) is used just for a parsimonious setting. One can use an AR(2) process too.
Conclusion

In this article, we reviewed and discussed the historical performance of the Quantity Theory of Money (QTM). We re-evaluate the one-to-one relationship between money growth and inflation. By adopting the same statistical and economic criteria as Lucas (1980), with a much larger data set covering both a longer period and many more countries, we found QTM breaks in U.S, and this collapsing of QTM is universal.

It appears that the period 1955-1980 is the only period during which QTM fits data well in most of our sample countries. It starts to break down when we go beyond this period. Furthermore, the recent breaking down of QTM coincides with a process of financial innovation.

To explain this breaking down for U.S, we use a money generating theory instead of money demand. To be more precise, the loan is an important source of money generation. And we decompose the loan structure of U.S market. We found real estate is the major collateral asset of Households and Firms. We thus propose money chases after real estate, a bundle of final goods and land. To confirm our judgment, we use a long historical data of nominal house price and find that (long-run) growth of nominal house price co-moves with (leads) growth of broad money more robustly.
We then propose a framework under which financial innovation can affect our estimation of QTM. And as our quantitative exercise shows our model is capable of explaining the collapsing of QTM.
Appendix

Data source, Cross-Country

We obtain different measures of money stock from various sources. These sources are summarized as follows:

Norway: Historical statistics of Norge Bank\(^{37}\). For M1/2, 1870-1919 are from Jan T. Klovland, Monetary Aggregates in Norway 18192003, chapter 5, Page 208-210. 1919-2006 are from Table a2a. Nominal and real GDP are from Table-c6-table5 and table6. For deflator, data are taken from Table-c6-table7. Data from 1940-1945 are missing. For 2004-2015, Statistics Norway(M1 and M2 are from Table-08253\(^{38}\) and Table-10945). Data of real and nominal GDP from 2000-2015 are from Eurostat. CPI data from 1924(1864)-2015 are from table 08184 of Consumption index, issued by Statistics Norway. After 1924, we use the growth rate of CPI to represent inflation.

Switzerland: 1907-2006, M1/3 is from The monetary base and the M1, M2, and M3 monetary aggregates, Swiss National Bank, 2007. Before 1907, we take the data from Schularick and Taylor(2012) For Price, 1870-1992 are from H.1 of historical statistics of Switzerland online. After that, we adopt CPI from Swiss Federal Statistical Office. For GDP, 1870-1913 are from Table Q.1a; 1914-2005 are from Q.16a, b; 1980-2016 are from Eurostat. From 1984-2016, data of M1/2/3 are available from the database of Swiss National Bank.

Sweden: Historical Monetary and Financial Statistics for Sweden, Volume 1-2, Sveriges Riksbank. For 1872-2006, CPI and GDP deflator are taken from Table A8.1, Volume 1, Page 443-447. M0/3 during 1846-2012 are taken from Table A7.2, and 7.3, Volume 1, Page 325-332, since No data on M1 was presented before 1999. After 2012, we use data offered by Statistics Sweden. GDP is taken from Table A4.3, Page 164-169 from 1872-2000. After that, we take data from Statistics Sweden. CPI during 1980-2016 is taken from Statistics Sweden(Consumer Price Index (CPI)/Living Cost Index, July 1914=100).

Australia: M1/3: 18701983 Page 5771 from David Pope, Australian Money and Banking Statistics. M3 after 1960 is from Reserve bank of Australia(RBA, ID: DMAM3N), M1 after


\(^{38}\) data of M1/2 stops at Feb, 2015
1975 is from RBA(ID: DMAM1N), between 1960-1975, we use data from OECD. CPI from 1922-2016 from RBA(Table G1, GCPIAG)

U.S: Real and nominal GDP (M1, M2 stock) are available from the FRED database since 1947: I (1959: I). Before that growth rates on the real GNP and M1, M2 series constructed by Balke and Gordon (1986). Since there is no M1 data until 1914, we use the growth rate of money base to approximate the growth rate of M1 from 1871-1914.

U.K: M0/1/3/4/4x and nominal GDP is from Money creation in the modern economy, by Michael Mcleay, Amar Radia and Ryland Thomas. Consumption price and Real GDP data is from The UK recession in context what do three centuries of data tell us? by Bank of England. After 1997, we use M4x(excluding intermediate other financial corporations) to represent broad money index. According to Bank of England, “…modify the measurement of UK M4 by excluding the money holdings of some OFCs in order to obtain a better measure of those money holdings that are likely to be used as a medium of exchange.”

Spain: From 1870-1998, M1/2 and Broad index(Disponibilidades liquidas, similar to M2 definition of Fed). Table 9.16, Page697. Carreras and X. Tafunell (eds.), Estadisticas Historicas De Espana, Madrid 2005 http://www.fbbva.es/TLFU/dat/autores.pdf. After that, we get them from IFS. 1936-1940 are missing for Money. Nominal GDP(El PIB a precios corrientes) from 1870-2000, Table 17.7, Page 1339-1340. After that we can have GDP from Eurostat(1995-2006). GDP deflator, from 1870-2000, Table 17.16(Deflactores implcitotos del PIB a precios de mercado y sus componentes de gasto). 1995-1997, we can use data from OECD. After 1998, we use (M1/3)data from Table 1.13 from Banco de Espana.

Denmark: From 1995-2016, M1/3 are from Danmarks Nationalbank(1995-2013 from Table DNM1KOR, after 2005 from table DNMINOGL). From 1970-1995, we use data from OECD. Nominal Gdp is from Barro-Ursua Macroeconomic Data (2010). From 1971-2016, we take from OECD. Real GDP (1995-2016) and Harmonized Index of Consumer Prices(2001-2016) are taken from Eurostat. Monetary data before 1971 are from narrow(broad) monetary index of Schularick and Taylor(2012)\(^\text{39}\).

Germany: (M0/M2, Price and output, Pre-1960) are from Rolnick and Weber (1997)\(^\text{40}\).  

\(^{39}\)They construct the monetary index by sum liability of the central bank, commercial banks and saving banks, by using data from Hans Chr. Johansen(1986)  

\(^{40}\)As robust check we also used data from Schularick and Taylor(2012), they construct M1 based on
Then for M1/3 during 1969-1998, we take data from IFS, after 1999 from Deutsche Bundesbank (we add "German contribution to the monetary aggregate M3 and its components in the euro area" with "Banknotes in circulation / Deutsche Bundesbank."). For real GDP and CPI from 1971-2000, we use data from OECD. Then we use data from Deutsche Bundesbank.

Netherlands (M1/2/3), from 1982-2015, Table 5.4 from De Nederlandsche Bank. They do not directly provide M1, M2. So we use the definition of Monetary stock according to ECB: $M1 = \text{Currency in circulation} + \text{Overnight deposits}$. $M2 = M1 + \text{Deposits with an agreed maturity up to 2 years} + \text{Deposits redeemable at a period of notice up to 3 months}^{41}$. For pre-1982, we adopt data from Schularick and Taylor (2012).

France (M1/2) are from Banque De France-M1,2^{42} after 1980, before which we adopt data from Rolnick and Weber (1997).

Italy (M1, M2 plus/3): For the whole period, GDP index is taken from Alberto Baffig (2011) Italian National Accounts, 1861-2011. M1/2, 1861-1939, Monetary aggregates in Italy 1861-2014, Banca d’Italia. After 1999, we take data from Banca d’Italia. We construct M3(M1) by adding “Italian contribution to euro-area M3(M1) excluding currency”^{43} and “currency held by public”^{44}. From 1940-1998, M1^{45} Data is offered by Banca d’Italia. M2plus^{46} offered by Banca d’Italia) was used before 1998 (Historic data offered by Banca d’Italia). After 1998, we use M3.

For Canada, M1/2 series: M1 before 1952 is from Metcalf, Redish, and Shearer (1996) after 1980 we use data from Statistics Canada, M2 before 1967 is from Metcalf, Redish, and Shearer (1996). All data between can be downloaded from Statistics Canada. And Metcalf, Redish, and Shearer (1996) offer monthly data; we convert the monthly series to the annual frequency by taking value in December. GDP and GDP deflator data are from Statistics

---


---

41 M3=M2+Repurchase agreements+Money market fund (MMF) shares/units+Debt securities up to 2 years. This debt is restricted to liabilities of the money-issuing sector and central government liabilities with a monetary character held by the money-holding sector


43 $BAM-AGGM.M.1020001.M3XC.3.101.EMU814.SBI138.1000$

44 $BAM-AGGM.M.1010001.AM01.0.101.WRDBI2.S0.EUR$

45 $SST-STSMB.M.M1ST.101$

46 $SST-STSMB.M.M2PLST.101$
Canada\textsuperscript{47}.
Numerical Solution

For completeness, we first derive equation-(1) to (6). Then we went through the solution details.

\[
\mathcal{L} = \sum_t \beta^t \left\{ U(c_t) + \lambda_t \left[ (q_t + d_t) s_t + b_t + L_t - s_{t+1} q_t - c_t - \frac{b_{t+1}}{R} \right] + \mu_t \left[ \frac{b_{t+1}}{R} + \phi_t q_t s_{t+1} + \psi_t \right] \right\}
\]

First order conditions of \(c_t, s_{t+1}\) and \(b_{t+1}\) will be

\[
U'(c_t) = \lambda_t + \phi_t q_t \Rightarrow U'(c_t) = RE_t^B[\beta U(c_{t+1})] + \mu_t \tag{A.1}
\]

\[
\lambda_t q_t = E_t^B[\beta \lambda_{t+1} (q_{t+1} + d_{t+1})] + \mu_t \phi_t q_t \Rightarrow q_t = \frac{E_t^B[\beta \lambda_{t+1} (q_{t+1} + d_{t+1})]}{U'(c_t) - \phi_t \mu_t} \tag{A.2}
\]

Combine equation-(A.1) with budget constraint and binding borrowing constraint will have:

\[
c_t = \min \left\{ \left(U'(.)^{-1}(E_t^B[\beta RU'(c_{t+1})])\right), \psi_t + \phi_t q_t s_{t+1} + b_t + z_t g(1) \right\} \tag{A.3}
\]

For any period \(t\), exogenous state variables are \(z_t\) and \(F_t\), endogenous state variable is \(b_t\) and \(s_t\). Since market clear condition restricts \(s_t\) to be 1, we simply state that \(b_t\) is the only endogenous variable. Given any realization of exogenous state variables, there exists a \(\bar{b}\), such that borrowing constraint will be binding if \(b_t < \bar{b}\). \(48\)

Given any realization of \(z_t, F_t, \bar{b}\) and policy functions from last iteration(\(k\)), \(c^k(z, F, B)\), \(q^k(z, F, B)\), and \(\mu^k(z, F, B)\):

When \(b_t \geq \bar{b}\), borrowing constraint is loose, we denote the current case as Case.1. Consumption is thus determined by the left branch of equation-A.3. If agent choose a bond holding \(B'\) next period, we denotes the consumption policy function as \(c_t(z_t, F_t, B')\), which will defined as following:

\[
c_{1,t}(z_t, F_t, B') = (U')^{-1}(E_t^B[\beta RU'(c_{t+1}, F_{t+1}, B')]) \tag{A.5}
\]

where \(U')^{-1}\) is inverse function of \(U'(.)\) and \(c^k(z, F, B)\) is policy function from iteration \(k\). This policy function is defined on state variable space. It is worthy to point out that

\(48\) It is possible that \(\bar{b} = \infty\), so borrowing constraint will be binding regardless the realization of exogenous variable
$c_{1,t}(z_t, F_t, B')$ is not a policy function yet, since $B'$ is not a state variable, at least not for period $t$. And when $b_t \geq \bar{b}$, borrowing constraint is loose, thus $\mu_t(z_t, F_t, B') = 0$.

From equation-(A.2), we can solve out the value of collateral asset given people choose bond holding of next period to be $B'$ as below:

$$q_{1,t}(z_t, F_t, B') = \frac{E^B_t[\beta U'(c^k(z_{t+1}, F_{t+1}, B'))(q^k(z_{t+1}, F_{t+1}, B') + d_{t+1})]}{U'(c_{1,t}(z_t, F_t, B'))} \quad (A.6)$$

where $q^k(z, F, B)$ is policy function from iteration $k$. $d_{t+1}$ is dividend from land. And the exception in equation-(A.5) and (A.6) are based on belief $B_t$ towards $z_{t+1}$ and $F_{t+1}$. And $q_{1,t}(z_t, F_t, B')$ is an increasing function of the first and third input.

Furthermore, $b_t$ can be backed out through budget constraint if people choose bond holding of next period to be $B'$, given $s_{t+1} = s_t = 1$:

$$c_t + \frac{B'}{R} = d_t + L_t + b_t \quad \Rightarrow$$

$$b_t(z_t, F_t, B') = d_t + L_t - c_t(z_t, F_t, B') - \frac{B'}{R}$$

Hence, for any $z_t$ and $F_t$, we define policy functions in iteration $k + 1$ for $B \geq \bar{b}(z_t, F_t)$ or Case 1 as:

$$c^{k+1}_1(z, F, B) = c_{1,t}(z, F, d_t(z) + L_t(z) - R c_t(z, F, B') - RB)$$

$$q^{k+1}_1(z, F, B) = q_{1,t}(z, F, d_t(z) + L_t(z) - R c_t(z, F, B') - RB)$$

$$\mu^{k+1}_1(z, F, B) = 0$$

where $c_{1,t}(z_t, F_t, B')$ and $q_{1,t}(z_t, F_t, B')$ are defined in equation-(A.5) and (A.6).

Otherwise, when $b_t < \bar{b}$, borrowing constraint is binding, we denote the current case as Case 2. Thus value of collateral asset is thus determined by borrowing constraint-*.

Given agent choose a bond holding $B'$ next period, we denotes asset price policy function as $q_t(z_t, F_t, B')$, which will defined as following:

$$q_{2,t}(z_t, F_t, B') = \frac{1}{\phi_t}[-\psi_t - \frac{B'}{R}] \quad (A.7)$$

where both of $\phi_t$ and $\psi_t$ are determined by $F_t$. And we substitute $mu_{2,t} = U'(c_t) -$
$E^{B_t}[RU''(c_{t+1})]$ into asset pricing equation-A.2 to solve consumption, we will get:

$$c_{2,t}(z_t, F_t, B') = U''^{-1}(\frac{1}{1 - \phi_t} \left\{ \frac{U'(c^k(z_{t+1}, F_{t+1}, B'))(d_{t+1} + q^k(z_{t+1}, F_{t+1}, B'))}{q_{2,t}(z_t, F_t, B')} 
- \phi_t R_t^{B_t}[RU''(c^k(z_{t+1}, F_{t+1}, B'))] \right\}) \tag{A.8}$$

where $q_{2,t}(z_t, F_t, B')$ is determined in Equation-A.8. Furthermore, according to equation-A.1:

$$\mu_{2,t}(z_t, F_t, B') = U'(c_{2,t}(z_t, F_t, B')) - E^{B_t}[RU''(c^k(z_{t+1}, F_{t+1}, B'))] \tag{A.9}$$

where $(U')^{-1}$ is inverse function of marginal utility $U'(\cdot)$, $c^k(z, F, B)$ and $q^k(z, F, B)$ are consumption and asset price policy function from the $k$th iteration. And finally in following formula, we back out the endogenous state variable $b_t$ at which given exogenous state variables $F_t$, agent is to choose $B'$ as bond holding for next period.

From budget constraint:

$$b_{t}(z_t, F_t, B') = d_t + L_t - c_t(z_t, F_t, B') - \frac{B'}{R}$$

The policy functions for the $k + 1$th iteration can be defined as following when borrowing constraint is binding:

$$c_{2,t}^{k+1}(z, F, B) = c_{2,t}(z, F, d_t(z) + L_t(z) - Rc_t(z, F, B') - RB)$$

$$q_{2,t}^{k+1}(z, F, B) = q_{2,t}(z, F, d_t(z) + L_t(z) - Rc_t(z, F, B') - RB)$$

$$\mu_{2,t}^{k+1}(z, F, B) = \mu_{2,t}(z, F, d_t(z) + L_t(z) - Rc_t(z, F, B') - RB)$$

In the end, we need to define a $\bar{b}$ for each $z_t$ and $F_t$. As we know, $\bar{b}$ is the threshold, if $b_t \geq \bar{b}$ agent stays unconstrained. In another words, $\bar{b}$ is the level at which agent will be marginal binding: For any realization of $z_t$ and $F_t$, say $z_t = z$, $F_t = F$, $\bar{b}(z_t, F_t)$\textsuperscript{49} is the solution of

$$c_{2}^{k+1}(z, F, \bar{b}) = c_{1}^{k+1}(z, F, \bar{b})$$

\textsuperscript{49}Since $\bar{b}$ will be different for different exogenous variables. Hence it is a function of $z_t$ and $F_t$. 

49
where \( c^k_i(z, F, b) \), \( i = 1, 2 \) represents consumption policy function under case \( i \). As our policy functions converge, threshold level will converge too.

Furthermore, since we are forcing agents to choose \( B' \) for next period, then we should restrict the choice of \( B' \) is a reasonable set. For example, for a borrowing constraint to be bind, \( B' < -\psi_t \), or agent will face no restriction at all, even when the collateral has zero value.

The range of \( B' \) is of course given by Equation-*. Under the constraint, \( B' \geq -\psi_t - \phi_t q_{1,t}(z_t, F_t, B') \geq -\psi_t - \phi_t q_{1,t}(z_t, F_t, B') \), the second inequality follows from Equation-A.2 with \( \mu = 0 \). Or we can present borrowing constraint as \( \psi_t + \phi_t q_{1,t}(z_t, F_t, B') + B' \geq 0 \), the left hand side of this equation is increasing function of \( B' \). The lower bound of \( B' \) should be thus solved from equation below:

\[
\psi_t + \phi_t q_{1,t}(z_t, F_t, B_{lower}) + B_{lower} = 0
\]

Hence, we will restrict \( B' \in [B_{lower}, -\psi_t] \) and solve case 2. If backed \( b_t \) stays below \( \bar{b} \), the policy function serves as constraint branch. Similarly, we restrict \( B' \geq B_{lower} \) when solving case 1. If backed \( b_t \) stays above \( \bar{b} \), the policy function serves as unconstrained branch. Since \( b_t \) are backed through budget constraints in both cases and consumption and borrowing is lower in case 2, backed \( b_t \) has no overlapping. We can patch these two branch together and form the policy function for next iteration.
A numerical solution under parameter $\beta = 0.96$, $R = 1.03$, $\gamma = 3$, $\alpha = .2$, $\phi^{H(L)} = 0.070(.05)$, $\psi^{H(L)} = 0.2(.2)$, $z^{H(L)} = 1.1(.969)$, is displayed in Figure-16.
References


52


