Cost of Inflation in Inventory Theoretical Models

Fernando Alvarez
University of Chicago
f-alvarez1(at)uchicago.edu

Francesco Lippi
University of Sassari, EIEF
f.lippi(at)uniss.it

Roberto Robatto
University of Wisconsin-Madison
robatto(at)wisc.edu

February 9, 2017

Abstract

We show that the area under the long-run demand curve for money measures the welfare cost of inflation for a very large class of inventory theoretical models of money demand. The class of inventory models considered has a general stochastic structure of the net cash expenditures as well as of the fixed/variable cost of withdrawing and depositing money. Thus, our framework includes, as special cases, a large number of models that have been studied in the literature. The most important feature that is responsible for our result is the fact that private agents fully internalize all the costs and benefits associated with managing their inventory of money. As a result, the social costs and benefits of holding money, which are related to the welfare cost of inflation, are equal to the private costs and benefits of holding money, which are in turn captured by the area under the money demand curve.

*We thank Lars Hansen, Harald Uhlig, and the students of Econ 33502 at the University of Chicago, in particular Yoshio Nozawa, for excellent comments.
1 Introduction and Summary

This paper contributes to the analysis of the welfare cost of inflation in inventory theoretical models of money demand. This class of monetary models is of interest because of its ability to account for actual households’ cash management behavior in the last two decades, notwithstanding major financial innovations in the payment instruments (see e.g. Alvarez and Lippi (2009, 2013) for some evidence).

Classic analyses model the welfare cost of anticipated inflation as a tax on money holdings. Bailey (1956) applies this approach to a static model of money demand. In static demand theory, if agents have quasi-linear preferences, taxation implies a deadweight loss that is measured as the reduction in consumer surplus, net of the revenues generated by the tax, which are simply transfers to other individuals in the society. The marginal cost of producing money is approximately zero because it can be freely printed by the government. Thus, surplus is the area under the Marshallian money demand curve as a function of its price, that is, the nominal interest rate, which represents the opportunity cost of holding money. According to this theory, which uses the one-to-one relationship between the nominal interest rate and inflation in the long run, the deadweight loss of a nonzero nominal interest rate is then equal to the welfare cost of inflation.

The simplicity of measuring the welfare cost of inflation as the area under the money demand curve is an important advantage for policy analysis and implementation. But this result is based on a static model, and it is not clear a priori whether it can be applied to dynamic inventory models of money. In these models, agents must finance an exogenous stream of consumption subject to a cash-in-advance constraint. This constraint imposes two costs that agents minimize: the cost of “making trips to the bank” in order to adjust their stock of money, and the opportunity cost of holding money, which is proportional to the nominal interest rate.

In this paper, we provide a positive answer to the above question. That is, we show that for a very large class of inventory models of money, if agents have a small discount rate, the welfare cost of the inflationary tax can be measured as the area under the average money demand curve.

The link between our result and the analyses of the welfare cost of inflation in static demand theory follows from three arguments. First, although inventory theoretical models are dynamic, the problem faced by agents as the discount rate tends to zero has the same solution as solving the steady-state problem, which is essentially static. The logic of this result is the same as in the neoclassical growth model, in which the steady-state level of capital tends to the golden-rule level as the discount rate tends to zero, and solving for the
golden-rule level of capital is simply a static problem. Second, the trade-off between the costs of adjusting the stock of money and the opportunity cost of holding money gives rise to an interior solution for the choice of average money holdings, as if agents had concave utility in average money holdings. This result, together with the assumption of linear disutility in the costs paid by the agents to adjust their stock of money, creates a parallel with the assumption of quasi-linear utility used in static demand theory. Third, there is no other wedge in the model, and thus the area under the money demand curve, which measures the private benefits of holding money, is also equal to the social benefits of holding money. Thus, this area measures the welfare cost of inflation.

The third argument distinguishes our result from Lagos and Wright (2005) and several other search-theoretical models of money. In those models, setting zero nominal interest rates is not sufficient to achieve the first-best allocation, because of a bargaining friction. As explained by Lagos, Rocheteau, and Wright (2015), the bargaining friction creates a wedge between the private and social benefits of holding money. Since the area under the money demand curve includes the private benefits of holding money, it measures only a fraction of the total welfare cost of inflation. As a result, in those models, inflation has both first and second-order effects on welfare.\(^1\) In contrast, in our model, inflation has only second order effects. Formally, the envelope theorem applies when we study the effect of a small change of inflation on welfare; that is, the first-order effect of a small change in inflation is zero because it corresponds to the private agents’ first-order conditions evaluated at the equilibrium.

The strength of our result is that it is independent of many details of the model. If there is some inventory-theoretical model (within the class we consider) that is a good approximation of how agents manage money and make transactions in practice, the welfare cost of inflation can simply be measured as the area under the money demand curve in the data. In this sense, the spirit of the analysis is in line with the “sufficient static” approach used in other area of economics. For instance, Chetty (2009) discusses such approach in the context of the public economic literature; Arkolakis, Costinot, and Rodríguez-Clare (2012) use it to provide robust estimates for the welfare gains from trade that holds in a large class of models; and Alvarez, Le Bihan, and Lippi (2014) apply a similar idea to quantify the effects of monetary policy shocks on output.

In Section 3, we provide a simple example based on the model of Baumol (1952) and Tobin (1956) in which we prove our result using both a closed-form solution and the envelope theorem argument. We also provide the same analysis for another simple special case, namely, the model of Miller and Orr (1966). We then apply the same logic to a very general

\(^1\)In the calibration of Lagos and Wright (2005), the welfare cost of inflation is an order of magnitude higher than the one measured using the approach of static demand theory.
setup, presented in Section 4. In our stochastic environment, we allow for both fixed and proportional transaction costs to adjust the stock of money, and for the possibility of receiving free exogenous adjustment opportunities, which have proven to be important in rationalizing the behavior of households facing the financial innovation that has taken place in the last two decades, as shown by Alvarez and Lippi (2009). A more precise comparison of our general setup with the literature of inventory models of money is postponed to Section 5, so that we can more clearly relate the features of our framework with other papers. The main result is proven in Section 6.

Our result is closely related to Lucas (2000), who shows that the area under the money demand curve is approximately equal to the welfare cost of inflation for the Sidrauski model of money in the utility function, and for the McCallum-Goodfriend shopping time model. We comment further on the result of Lucas (2000) in the next section.

2 The Area under the Money Demand Curve

Following Lucas (2000), let $w(R)$ be defined as the area under the inverse money demand curve, net of the interest rate cost:

$$w(R) = \int_0^R M(x)dx - M(R)R,$$

where $M(R)$ is the demand for money evaluated at the (net) nominal interest rate $R$. This equation is equivalent to the ordinary differential equation (O.D.E.) and boundary condition

$$w'(R) = -M'(R)R \text{ and } w(0) = 0.$$  

Lucas (2000) defines $\tilde{w}(R)$ as the percentage of income compensation that an agent requires to be indifferent between a steady state with interest rate $R$ and one with interest rate zero. Lucas (2000) shows that $\tilde{w}(R) \approx w(R)$ for the Sidrauski money-in-the-utility function and McCallum-Goodfriend shopping time models. In his setup, there is a simple relation with two-goods consumer theory that can be easily explained by focusing on the Sidrauski model. While the Sidrauski setup is fully dynamic, the measure of the cost of inflation is conducted by Lucas (2000) for steady states or balanced growth paths with constant inflation. The analysis is thus effectively reduced to a static model with two goods – consumption and money – because real money balances enter directly in the utility function. Since preferences in Lucas (2000) are not quasi-linear, the income compensation is not exactly equal to the area under the money demand curve. However, the two are approximately the same, especially
for small nominal interest rates,\(^2\) because the income effect of the inflation tax is small.\(^3\) The income effect is small because the cost of holding money, which is measured by the forgone interest on money balances, is a small fraction of income, in particular if the nominal interest rate is small. In fact, it is exactly zero if the nominal interest rate is zero. In terms of interpretation, the steady-state assumption is then matched with low-frequency US data.

We seek to generalize the result of Lucas (2000) to models of money demand based on the aggregation of the decisions of agents solving dynamic inventory problems in possibly stochastic environments. Since our setup is different, we have to define the analogous object in these models to the compensated variation \(\tilde{w}(R)\) defined in Lucas (2000). We define \(\tilde{w}(R)\) in inventory theoretical models of money to be the real resources used by agents in transactions aimed at adjusting their own money balances; we do not include the opportunity cost of holding money because it is simply a transfer between agents and thus does not represent any waste from the point of view of the society as a whole. We then proceed to show that \(\tilde{w}(R) = w(R)\) if agents do not discount the future very much. The result we develop has a “long-run” interpretation because the average resources will be computed under the long-run distribution of money holding implied by the model. The logic of the result can be explained first with a very simple but important example, namely, the Baumol-Tobin model of money demand, which we turn to next.

### 3 The Cost of Inflation in Two Classic Inventory Models

In this section we analyze two classic inventory models of money: the model of Baumol (1952) and Tobin (1956), and the model of Miller and Orr (1966). A third example, based on the model of Eppen and Fama (1969), is provided in Appendix A.

Each one of these models has an analytical solution for which we can define the welfare cost of inflation as the average adjustment cost paid by agents. This does not include the interest rate cost, which is the private opportunity cost of holding cash, because it is simply a transfer between agents for the society as a whole. We can also solve explicitly for the area under the money demand curve. For each model, we then compare the two expressions and verify that they are identical. We also use the model to provide guidance for the intuition

---

\(^2\)In particular, Lucas (2000) shows that the ODEs for \(w(R)\) and \(\tilde{w}(R)\) as functions of \(R\) have exactly the same functional form when evaluated at \(R = 0\), and hence by continuity, they are very similar for small values of \(R\). Furthermore, he shows that for commonly used and empirically plausible numerical examples, and for ranges of values of interest rates \(R\) going from low to moderate, these two functions are indistinguishable.

\(^3\)The income effect of a tax is zero for quasi-linear preferences because it only affects the good with constant marginal utility.
behind the proof of the more general result of Section 6.

3.1 The Result in the Baumol-Tobin Model

We describe the classic model of Baumol (1952) and Tobin (1956) to show that the cost of inflation is indeed the area under the demand curve. Our description and argument are done in a way to facilitate the argument for the general case, where the notation is necessarily more complex.

Time is continuous. Let $c$ be the constant deterministic rate of consumption per unit of time that an agent must finance with cash. Consider policies of withdrawing $W$ unit of cash at equally spaced intervals, so that there are $n$ withdrawals in a unit of time, where we ignore integer constraints. Between withdrawals, cash is spent at rate $c$ (i.e. $dm/dt = -c$), and cash hits zero every $1/n$ periods, just before a withdrawal (i.e., withdrawals occur at time $\tau_i = i/n$), where cash balances jump from zero to $m^* = W$. Thus, cash balances $m(t)$ follow the familiar saw-tooth of the Baumol-Tobin model. Each of these policies implies a different average number of withdrawals per unit of time $n$ and a different size of the withdrawal $m^* = W$, but if cash hits zero at the time of withdrawal and finances consumption $c$, the policies must satisfy $W \times n = c$. The average cash balance across time implied by these policies is thus $M = W/2$. Each cash withdrawal entails a fixed cost $K$ and average cash balances have an opportunity cost $R$. Notice that there is no discounting; that is, the agent evaluates the average cost under the invariant or long-run distribution.⁴

The problem for the agent is

$$v(R) = \min_{M,n} MR + nK \quad \text{subject to} \quad 2Mn = c$$

(3)

The solution to this problem is

$$n(R) = \sqrt{\frac{Rc}{2K}} \quad \text{and} \quad M(R) = \sqrt{\frac{cK}{2R}}.$$  

(4)

Then, define the welfare cost associated with the nominal interest rate $R$, denoted by $\tilde{w}(R)$, as

$$\tilde{w}(R) = K \cdot n(R).$$

(5)

While the problem defined in (3) considers both transaction costs and interest rate costs, for the society as a whole, the interest lost by some person $MR$ is a transfer to some other agent, and thus it is not included in (5).

⁴There is also no consideration of the impact of inflation on the use of cash balances (i.e., $c$ will be kept fixed as we vary $R$).
We can verify, simply by integrating the corresponding expression, that indeed \( \tilde{w}(R) = w(R) \); that is, the cost of inflation is equal to the area under the money demand curve. We then have \( \tilde{w}'(R) = K' n'(R) = \frac{1}{2} \sqrt{\frac{K}{2R}} \) and \( w'(R) = -M'(R) R = \frac{1}{2} \sqrt{\frac{K}{2R}} \).

The preceding argument uses the closed-form solution of the Baumol-Tobin model, hiding its logic. To understand the general argument, we turn to a different proof, which we will extend to the general case. First notice that our definition of the cost of inflation, (5), implies that

\[
\tilde{w}(R) = v(R) - M(R) R. \tag{6}
\]

Differentiating (6), we have \( \tilde{w}'(R) = v'(R) - M(R) - M'(R) R \). Using the envelope theorem on the problem defined by (3), we have \( v'(R) = M(R) \), and hence \( \tilde{w}'(R) = -M'(R) R = w'(R) \). This finishes the proof.

The alert reader may have noticed that this proof does not use many of the particular assumptions of the Baumol-Tobin model, except that the objective function trades off the opportunity cost \( M R \) with some form of resource cost, and this trade-off gives rise to a money demand \( M(R) \). To confirm this intuition, we repeat the analysis for an inventory model of money demand in which there is uncertainty about the cash flow to be financed, namely, the Miller and Orr model.

### 3.2 Cost of Inflation in the Miller and Orr Model

In the basic version of Miller and Orr (1966), unregulated cash balances follow a Browning motion without drift:

\[
dm(t) = \sigma dB(t) \quad \text{and} \quad m(t) \geq 0, \tag{7}
\]

where \( B(t) \) is a standard Brownian motion, so that at time zero, the random variable \( B(t) \) is normally distributed with mean \( B(0) \) and variance \( t\sigma^2 \). The interpretation of (7) is that the agent has to finance expenditures that are subject to a cash-in-advance constraint, but the agent can also receive cash inflow. The fact that the agent can have both inflows and outflows of money makes this model more natural for a firm.

As in Baumol-Tobin, there is a fixed transaction cost \( K \) that must be paid to withdraw money, but in the Miller and Orr model agents choose to both withdraw and deposit. We consider policies of the following type. Agents let cash balances drift according to (7) as long as cash balances are nonnegative and smaller than a threshold value \( m^* \). If cash balances hit either \( m(t) = 0 \) or \( m(t) = m^* \), then a withdrawal of \( W = m^* \) or a deposit of \( D = m^{**} - m^* \) takes place, and cash balances are returned after the adjustment to the value of \( m^* \). Thus, a policy is described by two parameters: \( 0 < m^* \leq m^{**} \). Given parameters \( (m^{**}, m^*) \), or equivalently, given \( (W, D) \), we compute the average number of adjustments, that is, the
deposits and withdrawals per unit of time, denoted by \( n(m^*, m^{**}) \), as well as the average cash balances, denoted by \( M(m^*, m^{**}) \). As in the Baumol-Tobin model, we let \( R \) be the opportunity cost of cash balances per unit of time. We assume again that the agent wants to minimize the sum of the average opportunity cost of the cash balances and the average cost of adjustments per unit of time; that is, there is no discounting:

\[
v(R) = \min_{0 \leq m^* \leq m^{**}} M(m^*, m^{**}) R + n(m^*, m^{**}) K. \tag{8}
\]

In Appendix B we repeat the derivation of Miller and Orr (1966), showing that the optimal choice of money \( M(m^*, m^{**}) \) and deposits and withdrawals \( n(m^*, m^{**}) \) is:

\[
M(m^*, m^{**}) = \frac{m^* + m^{**}}{3}, \tag{9}
\]

\[
n(m^*, m^{**}) = \frac{\sigma^2}{m^*(m^{**} - m^*)}, \tag{10}
\]

and the solution is

\[
M(R) = \frac{4}{3} \left( \frac{3\sigma^2 K}{4R} \right)^{\frac{1}{3}} \quad \text{and} \quad n(R) = \frac{\sigma^2}{2} \left( \frac{3\sigma^2 K}{4R} \right)^{-\frac{2}{3}}.
\]

As for the Baumol-Tobin model, we can verify that the welfare cost of inflation in Miller-Orr, which is defined as in (5), is equal to the area under the money demand curve. Using the above closed-form results, we have

\[
w'(R) = -M(R) R = \frac{1}{3} \left( \frac{\sigma^2 K}{R} \right)^{\frac{1}{3}} \left( \frac{4}{3} \right)^{\frac{2}{3}} \quad \text{and}
\]

\[
\tilde{w}'(R) = K n'(R) = \frac{1}{3} \left( \frac{\sigma^2 K}{R} \right)^{\frac{1}{3}} \left( \frac{4}{3} \right)^{\frac{4}{3}},
\]

and thus \( w(R) = \tilde{w}(R) \) using \( w(0) = \tilde{w}(0) \).

Moreover, we can also verify that the approach using the envelope theorem leads to the same conclusion because the results derived for the Baumol-Tobin model \( \tilde{w}'(R) = v'(R) - M(R) - M'(R) R \) and \( v'(R) = M(R) \) hold for Miller-Orr too.

4 General Setup for Inventory Models of Money

In this section we describe a general setup for an inventory theoretical model of money. We formulate the problem in continuous time because we believe it is more natural for an
inventory problem in which agents choose the time to act.\(^5\) Yet we believe that the same result about the welfare cost of inflation applies to discrete-time formulations.

Let \((x_t, m_t)\) be the state of an agent at time \(t\), where \(x_t \in X\) is an exogenous Markov process that carries information about shocks and \(m_t\) are the real cash balances. The exogenous state \(x_t\) is governed by the diffusion

\[
dx_t = \mu_x(x_t) \, dt + \sigma_x(x_t) \, dB.
\]

The problem for the agent is to minimize the expected discounted cost of withdrawals and deposits needed to finance an exogenous stream of consumption. Let \(C_t\) be the real value of the cumulative cash expenditures of the agent. We want to emphasize that \(C_t\) represents the cumulative value of consumption financed between time zero and time \(t\), while \(dC_t\) represents the value of consumption that is financed in a period of time of length \(dt\) (if \(dC_t\) is negative, it represents an inflow of cash).\(^6\) The process for \(C_t\) is a mixed jump diffusion process. The jump component of \(C_t\) captures lumpy expenditures (or inflows) and is described by a Poisson counter \(N_t\), with Poisson intensity rate \(\kappa(x_t)\) at time \(t\), with jumps at time \(t_i = 1, \ldots, N_t\) denoted by \(z_{t_i}\) and distributed with c.d.f. \(F(\cdot; x)\) conditional on the realization of \(x_t = x\). Except for the presence of \(x_t\), the distribution of the jump component \(z_t\) is independent of all past histories. The diffusion part has drift \(\mu_c(x_t)\) and volatility \(\sigma_c(x_t)\). Thus,

\[
C_t = \int_0^t \mu_c(x_s) \, ds + \int_0^t \sigma_c(x_s) \, dB_s + \sum_{i=0}^{N_t} z_{t_i}.
\]

The cumulative consumption between time zero and time \(t\) described by (11) is the solution to the stochastic differential equation:

\[
dC_t = \mu_c(x_t) \, dt + \sigma_c(x_t) \, dB_t + z_t \, dN_t.
\]

We can interpret this differential as the following. In a short period of time of length \(\Delta\), the expenditures to be financed with cash are \(C_{t+\Delta} - C_t\) (recall that \(C_{t+\Delta}\) is the cumulative consumption from time zero to time \(t + \Delta\)) and can be approximated by

\[
C_{t+\Delta} - C_t \approx \mu_c(x_t) \Delta + \sigma_c(x_t) \sqrt{\Delta} \, \epsilon_t + \chi_t \, z_t,
\]

where \(\epsilon_t\) is a standardized normal random variable, \(z_t\) is drawn from \(F(\cdot; x_t)\), and \(\chi_t\) is a

\(^5\)See, for instance, Bar-Ilan (1990a) on the lack of optimality of \(sS\) rules in a discrete-time version of a model for which \(sS\) rules become optimal if formulated in continuous time.

\(^6\)The approach of using the cumulative value of consumption is also sometimes used in continuous-time models that are not directly related to the inventory theory of money, such as Brunnermeier and Sannikov (2014).
Bernoulli distributed random variable, which takes the value $\chi_t = 1$ with probability $\Delta \kappa(x_t)$ and zero otherwise.

Given the process $dC_t$, the evolution of unregulated real cash balances is

$$dm_t = -m_t \pi dt - dC_t \equiv -(m_t \pi + \mu_c(x_t))dt - \sigma_c(x_t) dB_t - z_t dN_t$$

and $m_t \geq 0$ (13) for all $t$, where $m_t$ are the real balances and $\pi$ is the inflation rate. In a period of length $dt$, real cash balances decrease because of inflation $\pi$ and because of consumption expenditure $dC_t$. Note that Equation (13) is the unregulated process of money, in the sense that it does not include deposits or withdrawals.

The agent withdraws from and deposits to an asset account with real rate of return $r$. We denote $m(t^+) - m(t^-)$ as the jump in cash holdings that is due to the withdrawals and deposits the agent makes at time $\tau$. If the exogenous state is $x_\tau = x$, making a withdrawal of size $W > 0$ entails a fixed cost $K(x)$ and a variable cost $k(x) \geq 0$ (and either $K(x) > 0$ or $k(x) > 0$ or both), so the total cost is $K(x) + k(x) W > 0$. Likewise, for a deposit of size $D$, there are fixed and variable costs $\bar{K}(x)$ and $\bar{k}(x)$, respectively, so the total cost is $\bar{K}(x) + \bar{k}(x) D$. We summarize these costs as

$$K (m(t^+) - m(t^-), x) = \begin{cases} 
K(x) + k(x) W & \text{if } m(t^+) - m(t^-) = W > 0, \\
0 & \text{if } m(t^+) - m(t^-) = 0, \\
\bar{K}(x) + \bar{k}(x) D & \text{if } m(t^+) - m(t^-) = D > 0.
\end{cases} \quad (14)$$

and we assume that they are bounded above by constants $\hat{K} < \infty$ and $\hat{k} < \infty$:

$$\bar{K}(x), K(x) \leq \hat{K} \text{ for all } x \in X, \text{ and } \bar{k}(x), k(x) \leq \hat{k}. \quad (15)$$

In addition, we assume that there are occurrences in which the agent has the opportunity to withdraw or deposit at no cost. Such free adjustment opportunities are governed by a Poisson process with arrival rate $\lambda(x_t)$.

We can write the sequence problem for the agent as follows:

$$G(m, x) = \min_{\{\tau, m_t\}} \mathbb{E} \left\{ \sum_{i=0}^{\infty} e^{-\gamma \tau_i} [I_{\tau_i} K (m(t^+)_{\tau_i} - m(t^-)_{\tau_i}, x_{\tau_i}) + m(t^+)_{\tau_i} - m(t^-)_{\tau_i}] \right\} \quad (16)$$

subject to Equation (13), where $\{\tau_i\}_{i=0}^\infty$ is a sequence of stopping times indicating when adjustment takes place (i.e., withdrawals or deposits), and where $I_{\tau_i}$ is an indicator that takes the value of zero at the time of a free adjustment opportunity and the value of one otherwise. That is, the agent minimizes the total cost incurred to finance the exogenous
consumption stream. This total cost has two components: the costs paid to adjust the stock of cash, \( K(\cdot) \), and the real value of the cash withdrawn from the asset account.

In Appendix C.1, we show that (16) can be represented as an equivalent problem in which the agent minimizes the shadow cost:

\[
V(m, x; R, r) = \min_{\{\tau, m_0\}} \mathbb{E} \left\{ \sum_{i=0}^{\infty} e^{-r \tau_i} \left[ I_{\tau_i} K(m_{\tau_i}^+) - m_{\tau_i}^- , x_{\tau_i} \right] \right\} + \int_0^{\infty} e^{-r t} R m_t \, dt \tag{17}
\]

subject to (13). Problem (17) is closer to the formulation of standard inventory theoretical problems. The shadow cost faced by the agent has two components: the cost \( K(\cdot) \) incurred at the time of an adjustment, as in Problem (16), and the (discounted) opportunity cost of holding cash balances \( Rm_t \), where \( R \equiv r + \pi \) is the opportunity cost of holding one unit of cash per period and can be interpreted as the nominal interest rate. To emphasize the dependence of problem (17) on the parameters \( R \) and \( r \), we have included them in the argument of the value function; that is, we have written it as \( V(m, x; R, r) \).

We also have that cash has to be nonnegative; therefore,

\[
V(0, x; R, r) = \min_{\hat{m} \geq 0} \left\{ K(\hat{m}, x) + V(\hat{m}, x; R, r) \right\} \quad \text{and} \quad V(m, x; R, r) = +\infty \text{ if } m < 0 . \tag{18}
\]

We assume that the optimal policy is of the \( sS \) form, and we refer to it as \( p(R, r) = \{ p(x; R, r) \}_{x \in X} \); where, as in (17), we emphasize the dependence on the parameters \( R \) and \( r \). For each state \( x \), the optimal policy \( p(x; R, r) \) is described by five functions:

\[
p(x; R, r) = ( m_s(x; R, r) , \bar{m}_{\bar{s}}(x; R, r) , \bar{m}_*(x; R, r) , m_{**}(x; R, r) , \bar{m}^*(x; R, r)) , \tag{19}
\]

where \( m_s(x; R, r) \) is the value of cash chosen after a free adjustment opportunity, \( \bar{m}_{\bar{s}}(x; R, r) \) is the value of cash that triggers a deposit, \( \bar{m}_*(x; R, r) \) is the value of cash after the agent has made a deposit, \( m_{**}(x; R, r) \) is the value of cash that triggers a withdrawal, and \( \bar{m}^*(x; R, r) \) is the value of cash after the agent has made a withdrawal.\footnote{We omit the specification of the Hamilton-Jacobi-Bellman equation that describes the evolution of the state in the inaction region because we do not use it for the derivation of our results.}

We define three additional endogenous objects: \( H(m, x; p(R, r)) \), \( M(R, r) \), and \( P(y, x; p(R, r)) \). We denote \( H(m, x; p(R, r)) \) to be the invariant distribution of cash \( m \) and of shock \( x \) implied by the optimal decision rules \( p(R, r) \), and \( \Sigma(R, r) \) to be the support of the invariant distribution, that is, the smallest closed set whose complement has probability zero. We denote \( M(R, r) \) to be the expected value of cash holdings under the invariant
distribution \( H(m, x; p(R, r)) \):

\[
M(R, r) \equiv \int m \, H(dm, dx; p(R, r)). \tag{20}
\]

In other words, \( M(R, r) \) is the “long-run” money demand of this economy when the opportunity cost of cash holdings is \( R \) and agents discount the future at rate \( r \). We denote \( P(y, x; p(R, r)) \) to be the invariant distribution of a withdrawal or deposit \( y \in \mathbb{R} \) and of shock \( x \) implied by the optimal decision rules \( p(R, r) \), conditional on a non-free adjustment taking place.

The objective of this paper is not to formally show the existence of a solution with the characteristics proposed above. The model stated here is written to encompass the models in the literature that we survey below, and where several papers formally analyze such solutions. Indeed, the nature of the contribution of this paper is mostly independent of the exact details of the characterization. Nevertheless, we briefly comment on some assumptions to ensure that we have a well-defined stationary problem. We let \( r > 0 \) and assume that \( \mu_x \) and \( \sigma_x \) are such that \( \{x_t\} \) is stationary and has a unique invariant distribution with density \( h_x(\cdot) \). We assume that the intensity rates for the two Poisson processes, \( \kappa(x) \) and \( \lambda(x) \), are bounded above uniformly on \( x \) by a constant \( L \). Finally, the jump component of the cumulative consumption process has a finite mean: the distribution of \( z \) conditional on \( x \), with the c.d.f. given by \( F(\cdot; x) \), has finite first moments for all \( x \), and these first moments are integrable with respect to the invariant distribution of \( x \):

\[
\int \left[ \int |z| F(dz; x) \right] h_x(x) dx < \infty. \tag{21}
\]

5 Models Considered in the Literature

This section reviews different assumptions made in several of the inventory theoretical models of money in the literature. We look separately at four categories of assumptions (the process for real consumption flows, the discount rate, adjustment costs, and constraints on money holding), and for each category we describe how the general model of Section 4 can be mapped into related papers.

While most of the papers that we survey are characterized by continuous time and, if there is uncertainty, by a continuous state space as in our setup, we also compare our framework to papers with different structures. In particular, we compare our model with Milbourne (1983), in which time is discrete, and with Song and Zipkin (1993), in which the state \( x \) is a finite Markov chain rather than a diffusion.
In Section 5.5, we comment on the importance of deriving the shadow cost formulation (in our model, (17)) from the total cost minimization problem (in our model, (16)), rather than directly stating the problem of an agent as the minimization of the shadow cost. The latter approach is sometimes used in the literature and, if the problem is not properly formulated, an inconsistency may arise, with consequences for the analysis of the welfare cost of inflation.

5.1 Process for Real Consumption Flows

The simplest special case of the process for consumption in (12) is provided by the constant drift of the seminal model of Baumol (1952) and Tobin (1956), \( \mu_c(x_t) = \mu > 0 \), without any uncertainty, \( \sigma_c(x_t) = 0 \) and \( \kappa(x_t) = 0 \), and thus the cumulative consumption is \( C(t) = ct \). Other models such as Jovanovic (1982) and Alvarez and Lippi (2009) share the same features.

A random component to the cash flow is introduced by Miller and Orr (1966), Miller and Orr (1968), Eppen and Fama (1969), and Weitzman (1968) using no drift, \( \mu_c(x_t) = 0 \), no jumps, \( \kappa(x_t) = 0 \), and constant volatility, \( \sigma_c(x_t) = \sigma > 0 \), resulting in the cumulative process \( C(t) = \sigma B_t \). This process allows for both outflows and inflows of cash, and it is thus more appropriate for firms.

The combination of constant drift and constant volatility, \( C(t) = \mu t + \sigma B_t \) where \( \mu, \sigma > 0 \), is explored by Constantinides and Richard (1978), Constantinides (1978), Frenkel and Jovanovic (1980), Harrison, Sellke, and Taylor (1983), Harrison and Taskar (1983), Sulem (1986), and Bar-Ilan (1990a). The model by Baccarin (2009) is similar, but it allows for stochastic drift and stochastic volatility, thus, \( C(t) = \mu_c(x_t)t + \sigma_c(x_t)B_t \).

The jump component \( zdN_t \) in the consumption process is used in Alvarez and Lippi (2013) to model lumpy purchases (note that the size of each jump is a constant, \( z_t = z > 0 \)), in combination with a constant drift \( \mu_c(x_t) = \mu \). Alvarez and Lippi (2013) also document that the jump component is important in the description of actual cash consumption for a sample of Italian and Austrian households, and of a broader liquid asset (close to M2) for a sample of Italian customers of a large commercial bank. In Archibald and Silver (1978) and Song and Zipkin (1993), both drift and volatility are zero, but the size of each jump \( z \) is instead a random variable drawn respectively from a constant c.d.f. \( F(z) \) and from a state-dependent c.d.f. \( F(z; x_t) \). Bar-Ilan, Perry, and Stadje (2004) allow for the possibility of constant drift and constant volatility, and the \( z \) and \( N \) of the jump component are generated by a compounded Poisson process.

Since \( \mu_c(x_t), \sigma_c(x_t), \kappa(x_t), \) and \( F(\cdot; x_t) \) are in general a function of \( x_t \), our model allows

8Jovanovic (1982) is a general equilibrium setup, so one has to consider the dual of the maximization problem to obtain an inventory problem as considered here and set \( u(\cdot) \) as specified in (1) in Jovanovic (1982) to be linear.
for a very general time dependence of the process for consumption flow.

5.2 Discount Rate


5.3 Adjustment Costs

The third set of assumptions refers to the adjustment cost described by (14) and by the free adjustment opportunity with arrival rate $\lambda(x_t)$.

The simplest possibility is to allow for costs that are symmetric for withdrawals and deposits, $K(x) = \overline{K}(x)$ and $k(x) = \overline{k}(x)$. Within this class, there are models with only a fixed cost (such as Baumol (1952) and Tobin (1956), Miller and Orr (1966), Jovanovic (1982), and Alvarez and Lippi (2009)), models with only a proportional cost (Harrison and Taskar (1983), Eppen and Fama (1969)), and models with both fixed and proportional costs (Miller and Orr (1968), Harrison, Sellke, and Taylor (1983)).

Several models allow instead for costs that are not symmetric between withdrawals and deposits. Weitzman (1968) has asymmetric fixed costs only, and Eppen and Fama (1969) have asymmetric proportional costs only. Constantinides and Richard (1978), Constantinides (1978), Song and Zipkin (1993), Bar-Ilan, Perry, and Stadje (2004), and Baccarin (2009) have both asymmetric fixed costs and asymmetric proportional costs. Milbourne (1983) and Archibald and Silver (1978) have symmetric fixed costs but asymmetric proportional ones. Frenkel and Jovanovic (1980), Sulem (1986), and Bar-Ilan (1990a) do not allow for the possibility of deposits (which can be captured by $K(x) = +\infty$ and $k(x) = +\infty$ in our framework). The former model has fixed withdrawals costs only, while the latter two papers allow for both fixed and proportional withdrawal costs.
The free adjustment opportunity in Alvarez and Lippi (2009, 2013), with a constant arrival rate $\lambda(x_t) = \lambda$, creates some instances in which there are zero adjustment costs. This feature produces precautionary behavior in the sense that the optimal adjustment takes into account the possibility of a future free withdrawal or deposit opportunity, a behavior that is in line with the empirical evidence presented by Alvarez and Lippi (2009).

### 5.4 Constraints on Money Holding

We restrict our analysis to the case $m \geq 0$, but $m < 0$ is allowed at a higher holding cost in Eppen and Fama (1969), Constantinides and Richard (1978), Archibald and Silver (1978), Constantinides (1978), Milbourne (1983), Harrison and Taskar (1983), Sulem (1986), Bar-Ilan (1990a), and Baccarin (2009). A value of $m < 0$ is usually interpreted as an overdraft, while Bar-Ilan (1990a) has an interesting discussion on the interpretation of this as a cash-credit model, that is, on interpreting the purchases that happen when $m < 0$ as “credit” goods in the language of Lucas and Stokey (1987).

### 5.5 Derived Opportunity Cost and Welfare Cost of Inflation

In several inventory theoretical models of money, the minimization of an opportunity cost (in this paper, (17)) is often stated as the primitive problem, rather than being derived from an explicit total cost minimization (in this paper, (16)). This approach has been adapted from general inventory problems outside the monetary literature, but in inventory models of money, it has implications for the specification of the law of motion of cash (in this paper, (13)) and for the interpretation of the discount rate. If the model does not address such implications correctly, an inconsistency may arise. This issue is particularly important if one wants to conduct comparative statics with respect to the rate of inflation, for moderate or large inflation rates (the inconsistency, if any, will be minor if inflation is low and disappears if inflation is zero). Since the topic under consideration is the cost of inflation, we think that it is appropriate to emphasize this potential problem.

We argue that, if the model is set in terms of the real value of cash (as in the model in this paper), then the law of motion of cash must include inflation eroding the real value of cash (in this paper, the term $-\pi m(t)$ in Equation (13)). Otherwise, if the model is set in nominal terms, discounting should take place using the nominal interest rate, and nominal cash balances should not be changing with inflation.

If this link between discounting and nominal versus real specification of cash is not respected, the inconsistency may arise. The effect of this inconsistency is compounded by the fact that many models are set in steady state (thus, the future is not discounted) and use
money in nominal terms. For instance, Frenkel and Jovanovic (1980) use cash in nominal terms, fix the nominal interest rate, and then compare the result of minimizing the present value of the opportunity cost using the nominal interest rate as the discount rate, and of minimizing the steady-state cost (with no discounting). Not surprisingly, the two methods yield identical results when the nominal interest rate is zero.

To better understand our remark, consider the following simple example. Assume that cumulative consumption evolves according to $dC(t) = \sigma B(t)$, inflation $\pi$ is positive, and the real discount rate is $r = 0$, thus the nominal interest rate is $R = r + \pi = \pi$. If in an interval $t, t + s$ the realization of the Brownian component is zero ($B(\tau) = 0$ for all $t \leq \tau \leq t + s$), then nominal cash is constant but real cash decreases exponentially at the rate of inflation. Discounting future nominal cash using the nominal interest rate (i.e., inflation in this example) is thus equivalent to discounting the real value of cash using the real discount rate $r = 0$. If $\pi = 0$, then $R = r = 0$, and thus the discount rate is zero in the model both in nominal terms and in real terms.

Finally, as noted by Bar-Ilan (1990b), the opportunity cost of holding money $R$ does not necessarily have to be interpreted as the nominal interest rate. If the focus is on a monetary aggregate broader than money, then $R$ must be interpreted as the opportunity cost of holding a unit of such monetary aggregate, in comparison to holding another safe asset (e.g., bonds) that cannot be used for transactions.

6 Cost of Inflation and Area under the Money Demand Curve

In this section, we present and prove our main result. We first define the cost of inflation (Section 6.1), and then we list and discuss some regularity conditions (Section 6.2). Finally, we show that, for a small discount rate $r$, the cost of inflation in the model of Section 4 can be computed as the area under the money demand curve (Section 6.3).
6.1 Cost of Inflation: Definition

The cost of inflation \( \tilde{w}(R, r) \) is defined by:

\[
\tilde{w}(R, r) \equiv r \int \mathbb{E} \left\{ \sum_{i=0}^{\infty} e^{-r \tau_i} I_{\tau_i} K \left( m_{(\tau_i^+)} - m_{(\tau_i^+)} \right), x_{\tau_i} \right\} \bigg| x_0 = x, m_0 = m, p(R, r) \right\} H(dx, dm; p(R, r)) \quad (22)
\]

The term inside the integral is the present value of the adjustment costs \( K(\cdot) \) discounted at rate \( r \). The expectation is taken with respect to the paths of the exogenous variables, conditional on initial conditions \((x_0, m_0) = (x, m)\). This expression is then integrated with respect to the initial conditions \(m\) and \(x\) using the invariant distribution \( H(m, x; p(R, r)) \). If the economy is composed of a continuum of agents whose initial conditions are distributed according to the invariant distribution, then the integral in (22) averages the present discounted value of the adjustment costs \( K(\cdot) \) across all agents in the economy. Importantly, the cost of inflation (22) does not include the interest payments because such payments are transfers between agents and thus do not represent losses for the economy as a whole. Finally, the cost is expressed as a flow (i.e., it is multiplied by \( r \)).

6.2 Regularity Conditions

To prove our main result, we will impose the following regularity conditions about the equilibrium process for money \( m_t \), the value function \( V(m, x; R, r) \), the invariant distribution \( H(m, x; p(R, r)) \), and the expected value of cash holdings \( M(R, r) \). In the models that we know, these regularity conditions are typically satisfied.

i. For all \( r > 0 \), the process for money balances and the number of adjustments per unit of time are essentially bounded; that is, \(|m_t| < B(R)\) and \( \text{card} \left( \{\tau_i : t \leq \tau_i < t + 1\} \right) < \tilde{B}(R) \) with probability one, for some finite \( B(R) \) and \( \tilde{B}(R) \).

ii. For all \( r > 0 \) and \( R \), the process \( \{m_t\} \) has a unique invariant distribution, for which Birkhoff’s ergodic theorem applies.

iii. For all \( r > 0 \), the invariant distribution \( H(m, x; p(R, r)) \) has a density \( h(m, x; p(R, r)) \), which is differentiable with respect to \( R \), for almost all \( R \). For all \( r > 0 \) and each \( R > 0 \), the density \( h(m, x; p(R, r)) \) and whenever it is differentiable, its derivative \( \partial h(m, x; p(R, r)) / \partial R \), are bounded by functions that are integrable with respect to \((m, x)\) on the support of the invariant distribution \( H \).
iv. For almost all \( R > 0 \), the invariant distributions \( H(m, x; p(R, r)) \) and \( P(y, x; p(R, r)) \) converge weakly as \( r \downarrow 0.\) Whenever it exists, the derivative \( \partial h(m, x; p(R, r)) / \partial R \) converges pointwise as \( r \downarrow 0 \) for all \((m, x) \in \Sigma(R, r)\).

v. For almost all \( R, r > 0 \), the value function \( V(m, x; R, r) \) and, whenever it exists, \( \partial h(m, x; p(R, r)) / \partial R \) are jointly continuous in \((R, r)\).

vi. For all \( r > 0 \), the expected value of cash holding \( M(\hat{R}, r) \), defined in Equation (20), is integrable with respect to \( \hat{R} \) for almost all intervals \([0, \hat{R}]\).

The boundedness of Item (i) is usually satisfied because decision rules have an upper threshold for a deposit of cash and cash is nonnegative, and also because the number of adjustments is finite, due to the fixed cost.

The uniqueness of the invariant distribution in Item (ii) is also typically satisfied, since decision rules given by thresholds generate stochastic processes that are recurrent.

In most problems analyzed, the invariant distribution is assumed to have a density, since it is typically characterized as the solution of a Kolmogorov forward equation. We view Item (iii), Item (iv), and the joint continuity of \( \partial h / \partial R \) in Item (v) as regularity conditions on the density of the invariant distribution. In simple models like Baumol-Tobin, the continuity of the objective function evaluated at the optimal choice can be verified directly, or it can be established using the maximum theorem. The regularity condition about the value function \( V(m, x; R, r) \) in Item (v) extends this idea to the general case of Section 4.

Item (vi) guarantees that the area under the money demand curve is well defined. This assumption is typically satisfied, since the elasticity of the average money holding with respect to the nominal interest rate is usually less than one (for instance, it is one-half for Baumol-Tobin).

Finally, we show that Item (i), together with the general setup described in Section 4, can be used to derive the following two additional regularity conditions.

**Proposition 1.** Assume that the regularity condition in Item (i) holds. Then the following regularity conditions hold as well:

vii. For almost all \( R > 0 \), the value function \( V(m, x; R, r) \) is differentiable with respect to \( R \).

viii. For all \( r > 0 \) and each \( R > 0 \), the function \( V(m, x; R, r) \) and whenever it is differentiable, its derivative \( \partial V(m, x; R, r) / \partial R \), are bounded by functions that are integrable under the invariant \( H(\cdot; p(R, r)) \).

---

\(^9\)This type of convergence is sometimes denoted as weak-* convergence.
These results follow from Theorem 1 and 2 in Milgrom and Segal (2002). The differentiability of $V(m, x; R, r)$ with respect to $R$ follows from two observations. First, the argument of the minimization problem in (17) is an elementary function of $R$ (and thus differentiable with respect to $R$). Second, the boundedness in Item (i) guarantees that the differentiability can be extended to the left-hand side of (17), despite the max operator. The boundedness of $V(m, x; R, r)$ and of its derivative follows directly from the bounds imposed in Item (i) and the fact that future costs are discounted at a positive rate, $r > 0$. The detailed proof is provided in Appendix C.2.

6.3 The Result in the General Inventory Theoretical Model

We are now ready to state our main result.

**Proposition 2.** Assume that the regularity conditions (i) to (vi) hold. For almost all $R > 0$, the cost of inflation is well approximated by the area under the money demand curve for small real interest rates; that is,

$$\lim_{r \downarrow 0} \tilde{w}(R, r) = \lim_{r \downarrow 0} w(R, r) \equiv \lim_{r \downarrow 0} \left( \int_0^R M(\hat{R}, r) d\hat{R} - M(R, r) R \right).$$

(23)

The logic of Proposition 2 and of its proof are essentially the same as the simple results illustrated in Section 3.1 for the Baumol-Tobin and Miller-Orr models. To clarify this link, we now highlight the main steps of the proof of Proposition 2. The detailed proof is provided in Appendix C.3.

To prove Proposition 2 we derive an O.D.E. of the form:

$$\lim_{r \downarrow 0} \left( \frac{\partial \tilde{w}(R, r)}{\partial R} \right) = \lim_{r \downarrow 0} \left( R \frac{\partial M(R, r)}{\partial R} \right) = -\lim_{r \downarrow 0} \left( \frac{\partial}{\partial R} \left( \int_0^R M(\hat{R}, r) d\hat{R} - M(R, r) R \right) \right)$$

(24)

and a boundary condition at $R \downarrow 0$. That is, similar to Baumol-Tobin and Miller-Orr, we show that the welfare cost of inflation, as $r \downarrow 0$, is equivalent to the O.D.E. that describes the area under the money demand curve, presented in Section 2.\(^{10}\)

\(^{10}\)In order to derive Equation (23) from the O.D.E. (24), we need to interchange the limit as $r \downarrow 0$ with the derivative with respect to $R$ in (24), and then integrate with respect to $R$. Interchanging the limit as $r \downarrow 0$ with the derivative requires showing that the convergence in (24) is uniform, so that the proof is slightly more complicated than for the Baumol-Tobin and Miller-Orr models.
key steps that are required to obtain Equation (24) and their connections to the proof the welfare cost of inflation in Baumol-Tobin and Miller-Orr.

In the first step, we generalize Equation (6), derived for the Baumol-Tobin model, to the model of Section 4. That is, we link the welfare cost of inflation defined in Equation (22) to the private costs that each agent pays to manage its money holdings, represented by \( V(m, x; R, r) \), and to forgone interest payments on average money holdings, \( RM(R, r) \). That is, we show that the welfare cost of inflation, defined in Equation (22) can be written as:

\[
\tilde{w}(R, r) = r \int V(m, x; R, r)H(dm, dx; p(R, r)) - RM(R, r)
\]  

Recall that the welfare cost of inflation differs from the private costs \( V(m, x; R, r) \); that is, the private costs \( V(m, x; R, r) \) includes both the costs paid to adjust the stock of money, \( K(\cdot) \), and the forgone interest payments on average money holdings, \( RM(R, r) \), whereas the welfare cost of inflation includes only the former. We consider the integral of \( V(m, x; R, r) \) because the welfare cost of inflation is measured as the average cost with respect to the state \((m, x)\), and \( V(m, x; R, r) \) is multiplied by \( r \) because we express all costs as flows.

In the second step, we differentiate Equation (25) with respect to \( R \), similar to the approach used for the Baumol-Tobin model:

\[
\frac{\partial}{\partial R} \tilde{w}(R, r) = \frac{\partial}{\partial R} \left[ r \int V(m, x; R, r)H(dm, dx; p(R, r)) \right] - \frac{\partial}{\partial R} [RM(R, r)]
\]

\[
= \frac{\partial}{\partial R} \left[ r \int V(m, x; R, r)H(dm, dx; p(R, r)) \right] - \left[ M(R, r) + R \frac{\partial}{\partial R} M(R, r) \right].
\]  

(26)

The third and last step of the proof is slightly different from the analysis in Baumol-Tobin and Miller-Orr. In those models, we used the envelope theorem to simplify the term that corresponds to the first entry on the right-hand side of (26). As a result, we were able to rewrite the expression corresponding to the right-hand side of (26) as the O.D.E. that describes to the area under the money demand curve. In contrast, the envelope theorem in the general model of Section 4 implies:

\[
\int \frac{\partial}{\partial R} V(m, x; R, r)H(dm, dx; p(R, r)) = M(R, r).
\]  

(27)

Thus, the envelope theorem is not sufficient to simplify the first term on the right-hand side

---

11 Recall that \( M(R, r) \) is the average money holdings, defined in Equation (20).

12 That is, we used the the envelope theorem to show that a marginal change of \( R \) on the total cost faced by an the agent, \( v(R) \), is equal to the average money holdings: \( v'(R) = M(R) \).
of (26) because

\[
\frac{\partial}{\partial R} \left[ r \int V(m, x; R, r) H (d m, d x; p(R, r)) \right] \neq \int \frac{\partial}{\partial R} r V(m, x; R, r) H (d m, d x; p(R, r))
\]

\[
= M(R, r),
\]

where the inequality is due to the dependance of the endogenous invariant distribution of cash \(m\) and shock \(x\), \(H (m, x; p(R, r))\), on \(R\). However, if we take the limit as \(r \downarrow 0\) of (26), we are able to apply to the envelope theorem as in Baumol-Tobin and Miller-Orr. Taking the limit \(r \downarrow 0\), (26) becomes:

\[
\lim_{r \downarrow 0} \frac{\partial}{\partial R} \tilde{w} (R, r) =
\lim_{r \downarrow 0} \frac{\partial}{\partial R} \left[ r \int V(m, x; R, r) H (d m, d x; p(R, r)) \right] - \lim_{r \downarrow 0} \left[ M(R, r) + R \frac{\partial}{\partial R} M(R, r) \right]. \tag{28}
\]

When the discount rate converges to zero, agent are de facto minimizing the steady-state cost of managing cash holdings, rather than the present-discounted value. That is, taking \(r \downarrow 0\) creates an analogy to Baumol-Tobin and Miller-Orr, in which there is no discounting. As a result, we are able to apply the envelope theorem:

\[
\lim_{r \downarrow 0} \frac{\partial}{\partial R} \left[ r \int V(m, x; R, r) H (d m, d x; p(R, r)) \right] = \lim_{r \downarrow 0} M(R, r). \tag{29}
\]

and Equation (28) collapses to (24).

Let us discuss further the role of \(r \downarrow 0\). By taking this limit, it is as if we were considering the average long-run cost faced by agents to manage cash holdings, because of the ergodicity of the decision rules in our model.\(^{13}\) That is, when \(r \downarrow 0\), agents’ decisions regarding money holdings depend only on \(R\), and are thus independent of the current state \(x\) or \(m\). This idea is formalized by the result of the following proposition, which we will invoke to prove Proposition 2. The proof of Proposition 3 is provided in Appendix C.4.

**Proposition 3.** Assume that the regularity conditions (i), (ii), (iii), and (iv) hold. For almost all \(R > 0\) and all \((m, x) \in \Sigma (R, r)\),

\[
\lim_{r \downarrow 0} r V (m, x; R, r) = \bar{V} (R), \tag{30}
\]

where \(\bar{V} (R)\) is a bounded function of \(R\) and is independent of \(m\) and \(x\).

\(^{13}\)On the equivalence of average expected cost and expected discounted cost as the discount factor goes to zero, see, for example, Flynn (1976) and Archibald and Silver (1978).
7 Conclusions

We have provided a general dynamic inventory theoretical model of money that encompasses several papers in the literature. We have shown that, under some regularity conditions that are typically satisfied by these papers, the welfare cost of inflation can be measured as the area under the money demand curve, provided the discount rate is small.

References

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi. 2014. “Small and large price changes and the propagation of monetary shocks.”


### Appendix

#### A Cost of Inflation in the Eppen and Fama Model

This appendix discusses another seminal model of money demand, due to Eppen and Fama (1969). This model has the same process as Miller and Orr for (unregulated) cash balances but assumes that adjustment costs are proportional instead of fixed. We follow the formulation in Miller and Orr (1968), which leads to a closed-form solution for the minimized problem and the associated money demand. We also verify that the welfare cost of inflation (i.e. $\tilde{w}(R)$), is equal to the area under the money demand curve in this model as well.

The setup of the model is the same as in Section 3.2 except that instead of a fixed cost $K$, there is a variable cost $k$ of adjustment. In particular, the adjustment cost equals $kD$ if
there is a deposit of size $D$ and equals $kW$ for a withdrawal of size $W$. The type of policy considered is one in which the agent keeps the real balances controlled in the interval $[0, m^*]$. If real balances hit $m^*$, the agent makes a small deposit to keep the balances in $[0, m^*]$. Likewise if balances hit zero, the agent makes a small withdrawal to keep them in $[0, m^*]$. Given such a policy, one can compute the average cash balances, $M(m^*)$, as well as the expected value of the absolute value of adjustment per unit of time, denoted by $A(m^*)$. This quantity has the interpretation of the expected value of the sum of all the deposits plus all the withdrawals over a large period, divided by the length of the period. The agent solves

$$v(R) = \min_{m^* \geq 0} M(m^*) R + A(m^*) k.$$  \hspace{1cm} (A-1)

We first show that

$$M(m^*) = \frac{m^*}{2} \text{ and } A(m^*) = \frac{\sigma^2}{m^*}. \hspace{1cm} (A-2)$$

First, we argue that the density of the invariant distribution of $m$, denoted by $q$, is uniform in $[0, m^*]$. This can be seen by considering the equivalent discrete-time discrete-state-space model, where each period is of length $\Delta$ and where $m$ increases (decreases) with probability 1/2 by the amount $\sqrt{\Delta} \sigma$ in each period, implying that

$$q(m) = \frac{1}{2} q \left( m + \sqrt{\Delta} \sigma \right) + \frac{1}{2} q \left( m - \sqrt{\Delta} \sigma \right) \text{ for } 0 < m < m^*,$$

$$q(0) = \frac{1}{2} q \left( \sqrt{\Delta} \sigma \right) + \frac{1}{2} q(0) \text{ and}$$

$$q(m^*) = \frac{1}{2} q \left( m^* - \sqrt{\Delta} \sigma \right) + \frac{1}{2} q(m^*).$$

Therefore, taking a second-order Taylor approximation, dividing by $\Delta$, and taking the limit as $\Delta \rightarrow 0$, we have

$$0 = q''(m) \frac{\sigma^2}{2} \text{ and } q'(0) = q'(m^*) = 0.$$  

We can also see directly that (A-3) has a uniform distribution as its solution for any $\Delta > 0$. Since the distribution is uniform, then $M(m^*) = m^*/2$.

Second, we derive $A(m^*, \Delta)$. In a period of length $\Delta$, the probability of an adjustment is equal to $\sqrt{\Delta} \sigma/m^*$. This is the case because, to have an adjustment, the value of $m$ must be in either one of the two extremes, and in each case it will trigger an adjustment with probability 1/2. Since the size of each of the adjustments is $\sqrt{\Delta} \sigma$, and $A$ is the product of

14Technically, cash balances follow a reflecting Brownian motion with barriers 0 and $m^*$. A thoroughly rigorous analysis of this type of problem is given in Harrison and Taskar (1983).
the probability times the size of the adjustment,

\[ A(m^{**}, \Delta) = \frac{\sqrt{\Delta \sigma}}{m^{**}} \times \sqrt{\Delta \sigma} = \frac{\Delta \sigma^2}{m^{**}}. \]

Dividing this expression by the length of the time period \( \Delta \), we obtain the desired formula, \( A(m^{**}) = \sigma^2/m^{**} \).

Using (A-2), we can derive \( m^{**}, M(R) \), and \( A(R) \):

\[ m^{**} = \sqrt{\frac{2 \sigma^2 k}{R}}, \quad M(R) = \sqrt{\frac{\sigma^2 k}{2R}} \quad \text{and} \quad A(R) = \sqrt{\frac{\sigma^2 R}{2k}}. \]

We define the welfare cost of inflation to be \( \bar{w}(R) \equiv k A(R) \), which implies that

\[ \bar{w}'(R) = k A'(R) = \frac{1}{2} \sqrt{\frac{\sigma^2 k}{2R}} \]

and using Equation (2), we conclude that \( \bar{w}(R) = w(R) \). The same result can be obtained by applying the envelope theorem to problem (A-1): the result \( \bar{w}'(R) = w'(R) \) follows from the same steps as in Baumol-Tobin.

### B Derivation of \( n \) and \( M \) for Miller and Orr Model

To derive the average number of adjustments per unit of time for arbitrary parameters \((m^*, m^{**})\), we use the fundamental theorem of renewal theory and compute the reciprocal of the expected time between adjustments. To compute the expected time between adjustments, we define the following differential equation for the expected time until the first occurrence that \( m(t) = m^* \) or \( m(t) = m^{**} \):

\[ 0 = 1 + \frac{\sigma^2}{2} T''(m) \]  \hspace{1cm} (B-1)

for \( m \in (0, m^{**}) \), with boundary conditions \( T(0) = T(m^{**}) = 0 \), since adjustment happens immediately at each of these boundaries. This differential equation can be derived heuristically from taking limits and assuming differentiability from the following discrete-time (of length \( \Delta \)), discrete-state approximation:

\[ T(m) = \Delta + \frac{1}{2} T(m + \sqrt{\Delta \sigma}) + \frac{1}{2} T(m - \sqrt{\Delta \sigma}) \]

where we use the approximation that, in the interior, \( m \) increases (decreases) by the amount \( \sqrt{\Delta \sigma} \) with probability \( \frac{1}{2} \) each period. The solution of the ODE (B-1) is a quadratic function,
which, with the required boundaries, gives
\[
T(m^*) = \frac{m^*(m^{**} - m^*)}{\sigma^2} = \frac{1}{n(m^*, m^{**})}.
\]
Therefore, \(n(m^*, m^{**}) = \sigma^2 / [m^{**}(m^{**} - m^*)]\).

To derive the average cash balances for arbitrary parameters \((m^*, m^{**})\), we start by deriving the density of the invariant distribution of cash holdings. The forward Kolmogorov equation for this density is
\[
0 = q''(m) \frac{\sigma^2}{2} \text{ for } m \in (0, m^*) \cup (m^*, m^{**}), \quad q(0) = q(m^{**}) = 0 ,
\]
and it can be derived as the limit of the discrete-time, discrete-state law of motion where each period is of length \(\Delta\) and where \(m\) increases (decreases) by the amount \(\sqrt{\Delta} \sigma\) with probability \(1/2\) in each period:
\[
q(m) = \frac{1}{2} q\left(m + \sqrt{\Delta\sigma}\right) + \frac{1}{2} q\left(m - \sqrt{\Delta\sigma}\right) \text{ for } m \neq m^* \quad \text{(B-3)}
\]
\[
q(m^*) = \frac{1}{2} q\left(m^* + \sqrt{\Delta\sigma}\right) + \frac{1}{2} q\left(m^* - \sqrt{\Delta\sigma}\right) + \frac{1}{2} q(0) + \frac{1}{2} q(m^*)
\]
\[
q(0) = \frac{1}{2} q\left(\sqrt{\Delta\sigma}\right) \text{ and } q(m^{**}) = \frac{1}{2} q\left(m^{**} - \sqrt{\Delta\sigma}\right)
\]
for \(0 < m < m^{**}\). The solution for (B-2) is a density \(q(m)\) with a zero second derivative in the two segments, \((0, m^*)\) and \((m^*, m^{**})\). Imposing that this density integrates to one, that it is upward sloping for \((0, m^*)\) and downward sloping in \((m^*, m^{**})\), and that it equals zero in the extremes and is continuous at \(m^*\), we obtain the following triangular distribution:
\[
1 = h(m^*) m^*/2 + h(m^*) (m^{**} - m^*)/2 \implies h(m^*) = \frac{2}{m^{**}} ,
\]
\[
h(m) = \frac{h(m^*)}{m^*} m = \frac{2}{m^{**} m^*} m, \text{ for } m \in (0, m^*) ,
\]
\[
h(m) = h(m^*) - \frac{h(m^*)}{m^{**} - m^*} m = \frac{2}{m^{**}} - \frac{2}{(m^{**})^2} m, \text{ for } m \in (m^*, m^{**}).
\]

Since a triangular distribution with parameters 0, \(m^*, m^{**}\) has an expected value of \((m^* + m^{**})/3\), we obtain that
\[
M(m^*, m^{**}) \equiv \int_0^{m^{**}} m h(m) \, dm = \frac{m^* + m^{**}}{3} .
\]

Problem (8) can be solved by taking first-order conditions with respect to withdrawals.
\[ W \equiv m^* \text{ and deposits } D \equiv m^{**} - m^* : \]

\[
\min_{w,d \geq 0} R \frac{D + 2W}{3} + K \frac{\sigma^2}{WD} + \frac{\sigma^2}{W} \]

which gives that

\[
0 = -K \frac{\sigma^2}{WD^2} + \frac{1}{3} R, \quad \text{and} \quad 0 = -K \frac{\sigma^2}{W^2 D} + \frac{2}{3} R
\]

or, rearranging,

\[
D = \frac{3 \sigma^2}{WD R} \quad \text{and} \quad W = \frac{1}{2} \frac{3 \sigma^2}{WD R}
\]

which have the solutions

\[
D = 2W \quad \text{and} \quad W = \left( \frac{3 \sigma^2 K}{4 R} \right)^{\frac{1}{3}}.
\]

Therefore,

\[
M(R) = \frac{D + 2W}{3} = \frac{4W}{3} = \frac{4}{3} \left( \frac{3 \sigma^2 K}{4 R} \right)^{\frac{1}{3}},
\]

\[
n(R) = \frac{\sigma^2}{WD} = \frac{\sigma^2}{2W^2} = \frac{\sigma^2}{2} \left( \frac{3 \sigma^2 K}{4 R} \right)^{-\frac{2}{3}}.
\]

C. Proofs

C.1 Equivalence of Problems (16) and (17)

In this appendix we formalize the relationship between problem (16) and problem (17). To compare the opportunity cost value function \( V(\cdot) \) with the cost \( G(\cdot) \), it is useful to define the present value of the cash expenditures as follows:

\[
\mathbb{C}(x) = \mathbb{E}\left\{ \int_0^\infty e^{-rt}dC_t \mid x_0 = x \right\}.
\]

We are now ready to relate the two value functions as follows. For simplicity, we omit \( R \) and \( r \) from the arguments of the value function \( V(\cdot) \) and of the policy function \( p \).
Proposition 4. If the opportunity cost is $R = r + \pi$, the functions $G(\cdot)$ and $V(\cdot)$ satisfy
\[ G(m, x) = V(m, x) - m + C(x) \tag{C-1} \]
for all $m \geq 0$ and all $x \in X$, and the optimal policies associated with $G(\cdot)$ and $V(\cdot)$ are both of the sS form described by Equation (19), then $p = \{p(x)\}_{x \in X}$ and $G(\cdot)$ solve the Bellman equation for the total cost problem (16) if and only if $p = \{p(x)\}_{x \in X}$ and $V(\cdot)$ solve the Bellman equation for the shadow cost problem (17).

Proof. Let $G(\cdot)$ be a function satisfying Equation (C-1) that solves the Bellman equation for the total cost problem (16) and let $p$ be the optimal policy. We show that this implies that $p$ and $V(\cdot)$ solve the Bellman equation for the shadow cost problem (17).

Adding and subtracting $m_0 = m$ to the right-hand side of (16) evaluated at the path of money and at the stopping times implied by the optimal policy $p$, we have

\[ G(m, x) = \mathbb{E} \left\{ \sum_{i=0}^{\infty} e^{-rt_i} \left[ I_{\tau_i} K \left( m(\tau_i^+) - m(\tau_i^-), x_{\tau_i} \right) + m(\tau_i^+) - m(\tau_i^-) \right] + m - m \right\} \]

or, defining $\tau_{-1} \equiv 0$ and rearranging, we have

\[ G(m, x) + m = \mathbb{E} \left\{ \sum_{i=0}^{\infty} e^{-rt_i} \left[ I_{\tau_i} K \left( m(\tau_i^+) - m(\tau_i^-), x_{\tau_i} \right) \right] 
- \sum_{i=0}^{\infty} e^{-rt_i} \left[ m(\tau_i^-) - e^{-r(\tau_i-1-\tau_i)} m(\tau_i^+) \right] \right\}. \tag{C-2} \]

We now derive an expression for the last term on the right-hand side of the previous expression. Consider the budget constraint, Equation (13), and multiply both sides by $e^{-rt}$:

\[ e^{-rt} dm_t = -\pi e^{-rt} m_t dt - e^{-rt} dC_t. \]

Integrating both sides from $\tau_{i-1}^+$ to $\tau_i^-$, we have

\[ \int_{\tau_{i-1}^+}^{\tau_i^-} e^{-rt} dm_t = -\pi \int_{\tau_{i-1}^-}^{\tau_i^-} e^{-rt} m_t dt - \int_{\tau_{i-1}^-}^{\tau_i^-} e^{-rt} dC_t. \]

The integral on the left-hand side is a Riemann-Stieltjes one, and thus it satisfies

\[ \int_{\tau_{i-1}^+}^{\tau_i^-} e^{-rt} dm_t = e^{-r\tau_i} m(\tau_i^-) - e^{-r\tau_{i-1}} m(\tau_{i-1}^+) + r \int_{\tau_{i-1}^+}^{\tau_i^-} e^{-rt} m_t dt \]
Therefore,

\[ e^{-r\tau_i} \left[ m(\tau_i) - e^{-r(\tau_i-\tau_{i-1})} m(\tau_{i-1}^+) \right] = - (\pi + r) \int_{\tau_{i-1}}^{\tau_i} e^{-rt} m_t dt - \int_{\tau_{i-1}}^{\tau_i} e^{-rt} dC_t. \]

Plugging the last expression into Equation (C-2), using \( r + \pi = R \), and rearranging, we have

\[ G(m, x) + m = \mathbb{E} \left\{ \sum_{i=0}^{\infty} e^{-r\tau_i} \left[ I_{\tau_i} K \left( m(\tau_i^+) - m(\tau_i^-), x_{\tau_i} \right) \right] + \int_0^\infty R e^{-rt} m_t dt + C(x) \right\}. \]

Using Equation (C-1) and rearranging, the result follows. The converse can be established similarly.

C.2 Proof of Proposition 1

Item (vii) follows from Theorem 2 in Milgrom and Segal (2002). This theorem applies because, for all \( \{\tau_i, m_i\} \), the argument of the minimization in Equation (17) is absolutely continuous in \( R \) and because its derivative with respect to \( R \) is bounded:

\[ \mathbb{E} \left\{ \int_0^\infty e^{-rt} m_t dt \right\} \leq B(R) \mathbb{E} \left\{ \int_0^\infty e^{-rt} dt \right\} = \frac{B(R)}{r}, \quad \text{(C-3)} \]

where the inequality follows from Item (i) and the equality uses \( \int_0^\infty e^{-rt} dt = 1/r \). Under these assumptions, Theorem 2 in Milgrom and Segal (2002) concludes that \( V \) is absolutely continuous and thus differentiable almost everywhere.

To prove Item (viii), we use the differentiability of \( V \) that we have just established and apply Theorem 1 in Milgrom and Segal (2002):

\[ \frac{\partial V(m, x; R, r)}{\partial R} = \mathbb{E} \left\{ \int_0^\infty e^{-rt} m_t dt \mid x_0 = x, m_0 = m, p(R, r) \right\} \]

which is bounded, as established by Equation (C-3).

To complete the proof of Item (viii), we still need to show that the first term inside the expectation on the right-hand side of Equation (17) (the discounted sum of adjustment costs) is also bounded. We first define

\[ \tilde{K}_t = \sum_{\{\tau_i : t \leq \tau_i < t+1\}} e^{-r(\tau_i-t)} \left[ I_{\tau_i} K \left( m(\tau_i^+) - m(\tau_i^-), x_{\tau_i} \right) \right] \quad \text{(C-4)} \]
and using Item (i) and the assumption in Equation (15), we have

\[ \tilde{K}_t \leq \left[ \tilde{K} + B(R) \hat{k} \right] \sum_{\{\tau_i: t \leq \tau_i < t+1\}} e^{-r(\tau_i - t)} \]

\[ \leq \left[ \tilde{K} + B(R) \hat{k} \right] \tilde{B}(R) < \infty, \]

where the second inequality uses the fact that \( \sum_{\{\tau_i: t \leq \tau_i < t+1\}} e^{-r(\tau_i - t)} \) is bounded by the number of adjustments that take place between time \( t \) and time \( t + 1 \). Therefore,

\[
\mathbb{E} \left\{ \sum_{i=0}^{\infty} e^{-rt} \left[ I_{\tau_i} K \left( m_{(\tau_i^+)} - m_{(\tau_i^-)}, x_{\tau_i} \right) \right] \right\} = \mathbb{E} \left\{ \sum_{t=0}^{\infty} e^{-rt} \tilde{K}_t \right\} 
\]

\[ \leq \sum_{t=0}^{\infty} e^{-rt} \left( \left[ \tilde{K} + B(R) \hat{k} \right] \tilde{B}(R) \right) < \infty, \]

completing the proof.

### C.3 Proof of Proposition 2

In order to derive Equation (24), we begin by noting that\(^{15}\)

\[ \int [rV(m, x; R, r)] H(dm, dx; p(R, r)) \]

\[ = \int \mathbb{E} \left\{ \sum_{i=0}^{\infty} re^{-rt} I_{\tau_i} K \left( m_{(\tau_i^+)} - m_{(\tau_i^-)}, x_{\tau_i} \right) \mid x_0 = x, m_0 = m, p(R, r) \right\} H(dm, dx; p(R, r)) \]

\[ + R \int \mathbb{E} \left\{ \int_{0}^{\infty} re^{-rt} m_t dt \mid x_0 = x, m_0 = m, p(R, r) \right\} H(dm, dx; p(R, r)) \]

\[ = \tilde{w}(R, r) + R \int_{0}^{\infty} re^{-rt} \mathbb{E} \left[ m_t \mid x_0 = x, m_0 = m, p(R, r) \right] H(dm, dx; p(R, r)) dt \]

\[ = \tilde{w}(R, r) + R \int_{0}^{\infty} re^{-rt} M(R, r) dt = \tilde{w}(R, r) + RM(R, r). \]

The first equality uses the definition of the value function (17). The second equality uses the definition of \( \tilde{w}(R, r) \) in Equation (22) and interchanges expectations, using the assumption \(|m_t| < B(R)\) for all \( r > 0 \) in Item (i). The third equality uses the ergodicity assumption of Item (ii), which implies that the unconditional expectation of money demand is equal to the average money demand \( M(R, r) \) under the invariant distribution (see Equation (20)). The last equality uses \( \int_{0}^{\infty} re^{-rt} dt = 1.\)

\(^{15}\)Where relevant, results should be interpreted in the almost sure sense.
Then, differentiating with respect to $R$ and rearranging, we have

$$\frac{\partial}{\partial R} \int [rV(m, x; R, r)] H(dm, dx; p(R, r)) - M(R, r) = \frac{\partial \tilde{w}(R, r)}{\partial R} + R \frac{\partial M(R, r)}{\partial R}. \quad (C-6)$$

The first term in Equation (C-6) can be rewritten as

$$\frac{\partial}{\partial R} \int rV(m, x; p(R, r)) H(dm, dx; p(R, r)) \quad (C-7)$$

$$= \int \int \left( \frac{\partial rV(m, x; R, r)}{\partial R} \right) h(m, x; p(R, r)) \, dx \, dm$$

$$+ \int \int rV(m, x; R, r) \left[ \frac{\partial h(m, x; p(R, r))}{\partial R} \right] \, dx \, dm.$$

In (C-7), we have used the fact that $H$ has a density, Item (iii), and we have interchanged the integration with respect to $(m, x)$ with the derivatives with respect to $R$ using Leibniz’s rule. Leibniz’s rule can be applied because $\partial h(m, x; R, r)/\partial R$ and $\partial V(m, x; R, r)/\partial R$ exist by Item (iii) and Item (vii), and the functions $h(m, x; R, r)$ and $rV(m, x; R, r)$ and their derivatives with respect to $R$ are bounded by integrable functions by Item (iii) and Item (viii).

We now focus on the term $\partial [rV(m, x; R, r)]/\partial R$ in Equation (C-7). By the envelope theorem, for almost all $R > 0$, we have

$$\frac{\partial r V(m, x; R, r)}{\partial R} = \mathbb{E} \left\{ \int_0^\infty r e^{-rt} m_t \, dt \mid x_0 = x, m_0 = m, p(R, r) \right\}, \quad (C-8)$$

where we have used Theorem 1 in Milgrom and Segal (2002) because the objective function is affine in the parameter $R$ and satisfies Item (vii). Using the result in Equation (C-8), we can thus rewrite the first term on the right-hand side of Equation (C-7):

$$\int \frac{\partial r V(m, x; R, r)}{\partial R} h(m, x; p(R, r)) \, dx \quad (C-9)$$

$$= \int \mathbb{E} \left\{ \int_0^\infty e^{-rt} m_t \, dt \mid x_0 = x, m_0 = m, p(R, r) \right\} H(dm, dx; p(R, r))$$

$$= M(R, r),$$

where the second equality follows from the same steps as in (C-5). Thus, combining Equation (C-6), Equation (C-7), and Equation (C-9), we have

$$\int \int \left( r V(m, x; R, r) \left[ \frac{\partial h(m, x; p(R, r))}{\partial R} \right] \right) \, dx \, dm = \frac{\partial \tilde{w}(R, r)}{\partial R} + R \frac{\partial M(R, r)}{\partial R}. \quad (C-10)$$

Next, we show that the left-hand side of Equation (C-10) converges uniformly to zero as
\( r \downarrow 0 \), as does the right-hand side, implying that Equation (24) holds uniformly with respect to \( R \). To do so, we first show that the left-hand side of Equation (C-10) converges pointwise to zero as \( r \downarrow 0 \):

\[
\lim_{r \downarrow 0} \int \int \left( r \ V (m, x; R, r) \left[ \frac{\partial h (m, x; p (R, r))}{\partial R} \right] \right) \, dx \, dm
\]  

(C-11)

\[
= \int \int \left( \left[ \lim_{r \downarrow 0} r \ V (m, x; R, r) \right] \left[ \lim_{r \downarrow 0} \frac{\partial h (m, x; p (R, r))}{\partial R} \right] \right) \, dx \, dm
\]

\[
= \int \int \left( \nabla (R) \left[ \lim_{r \downarrow 0} \frac{\partial h (m, x; p (R, r))}{\partial R} \right] \right) \, dx \, dm
\]

\[
= \nabla (R) \lim_{r \downarrow 0} \left[ \frac{\partial}{\partial R} \int \int h (m, x; p (R, r)) \, dm \, dx \right]
\]

\[
= \nabla (R) \lim_{r \downarrow 0} \left[ \frac{\partial}{\partial R} (1) \right] = 0.
\]

The first equality interchanges the limit with respect to \( r \) with the integration with respect to \((m, x)\) using the Lebesgue dominated convergence theorem, which holds given the boundedness assumptions in Item (iii) and Item (viii), and given the pointwise convergence of \( r \ V (m, x; R, r) \) shown in Proposition 3 and of \( \partial h / \partial R \) in Item (iv). The second equality uses the result of Proposition 3. The third equality rearranges by noting that \( \nabla (R) \) does not depend on \( m \) or \( x \) and thus can be taken out of the integrals, it again interchanges the limit with respect to \( r \) and the integration, and it uses Leibniz’s rule to interchange the integrals with the derivative with respect to \( R \); the argument that allows the application of Leibniz’s rule is analogous to Equation (C-7). The fourth equality uses the fact that \( h \) is a density and thus integrates to one, and the last equality uses \( \frac{\partial}{\partial R} (1) = 0 \).

We can then apply Lemma 1 (see Appendix C.5) to show that the convergence of Equation (C-11) as \( r \downarrow 0 \) is not just pointwise but also uniform. Lemma 1 applies if we consider a closed and bounded interval containing \( R \) because the argument of the limit in Equation (C-11) is continuous. Continuity follows from Item (v) and is preserved under the integral sign because of the boundedness assumptions in Item (iii) and Item (viii). Therefore, the convergence of the left-hand side of Equation (C-10) and thus of Equation (24) is uniform.

Next, we want to exchange the limit and the derivative in Equation (24), which further requires showing that \( \tilde{w} (R, r) \) and \( \left( \int_{0}^{R} M (\hat{R}, r) d\hat{R} - M (R, r) \right) \) converges as \( r \downarrow 0 \). The convergence of the former can be established similarly to the convergence of \( r V (m, x; R, r) \) in Proposition 3. To show the convergence of the latter, we first note, using Equation (20), that \( M (R, r) \) converges because of the weak convergence of \( H \) in Item (iv). To establish the convergence of \( \int_{0}^{R} M (\hat{R}, r) d\hat{R} \) as \( r \downarrow 0 \), we note that \( M (R, r) \) is integrable because of the
assumption in Item (vi) and its pointwise limit as \( r \downarrow 0 \) exists as just established; therefore we can apply the Lebesgue dominated convergence theorem, and the result follows. Thus, we can exchange the limit and the derivative in Equation (24):

\[
0 = \frac{\partial}{\partial R} \lim_{r \downarrow 0} \left[ \tilde{w}(R, r) - \left( \int_0^R M(\bar{R}, r)d\bar{R} - M(R, r)R \right) \right],
\]  \hspace{1cm} (C-12)

and it follows that the convergence of the term in square brackets is uniform.

Then, integrating Equation (C-12) from \( \bar{R} > 0 \) to \( R \) and taking the limit as \( \bar{R} \downarrow 0 \), we have

\[
0 = \lim_{r \downarrow 0} \left[ \tilde{w}(R, r) - \left( \int_0^R M(\bar{R}, r)d\bar{R} - M(R, r)R \right) \right] - \lim_{\bar{R} \downarrow 0} \lim_{r \downarrow 0} \left[ \tilde{w}(R, r) - \left( \int_0^R M(\bar{R}, r)d\bar{R} - M(R, r)R \right) \right].
\]  \hspace{1cm} (C-13)

Next, we show that the second line of Equation (C-13) (i.e., the boundary condition) is zero, and thus the result of the proposition follows from the fact that the limits of the two terms \( \tilde{w}(R, r) \) and \( \left( \int_0^R M(\bar{R}, r)d\bar{R} - M(R, r)R \right) \) exist as derived before. First, using Equation (22), we note that for any \( r > 0 \), \( \lim_{R \downarrow 0} \tilde{w}(R, r) = 0 \), since as \( R \downarrow 0 \) the agent can make the cost arbitrarily close to zero by holding a very large quantity of money without facing any opportunity cost and thus avoid paying any adjustment costs as well. Second, for any \( r > 0 \) and as \( R \downarrow 0 \), both \( \int_0^R M(\bar{R}, r)d\bar{R} \) and \( M(R, r)R \) converge to zero. The latter result must hold, otherwise the expected value of cash holding \( M(\bar{R}, r) \) would not be integrable in \([0, R]\), violating Item (vi). Third, because of the existence of the limit as \( R \downarrow 0 \) just established, and because of the uniform convergence of the term in square brackets in the second line of Equation (C-13) as \( r \downarrow 0 \), we can exchange the limits in the second line of Equation (C-13).\(^{16}\) Thus, the result follows.

C.4 Proof of Proposition 3

Using Equation (17) evaluated at the optimal choices, we note that \( rV(m, x; R, r) \) is the sum of two terms:

\[
V(m, x; R, r) = \mathbb{E} \left\{ \sum_{i=0}^{\infty} re^{-rt_i} \left[ I_{t_i} K \left( m_{(r_i)}^+ - m_{(r_i)}^-, x_{r_i} \right) \right] \right\} + R \mathbb{E} \left\{ \int_0^{\infty} re^{-rt} m_t dt \right\},
\]

\(^{16}\)This result is sometimes referred to as the Moore-Osgood theorem.
where the expectation is taken with respect to each possible realization of the exogenous path \( \{x_t, t \geq 0\} \). Thus, we need to analyze
\[
\lim_{r \downarrow 0} r V(m, x; R, r) = \\
\lim_{r \downarrow 0} \mathbb{E} \left\{ \sum_{i=0}^{\infty} re^{-rt_i} \left[ I_{\tau_i} K \left( m(\tau_i^+), x_{\tau_i} \right) - m(\tau_i^-), x_{\tau_i} \right] \right\} + R \lim_{r \downarrow 0} \mathbb{E} \left\{ \int_0^{\infty} re^{-rt} m_t \, dt \right\}. 
\] (C-14)

The rest of the proof is organized as follows. We first analyze \( \lim_{r \downarrow 0} \int_0^{\infty} re^{-rt} m_t \, dt \), and we show that it converges to a bounded limit that depends only on \( R \) and, in particular, is independent on the exogenous path \( \{x_t, t \geq 0\} \); therefore \( \mathbb{E} \left\{ \lim_{r \downarrow 0} \int_0^{\infty} re^{-rt} m_t \, dt \right\} = \lim_{r \downarrow 0} \int_0^{\infty} re^{-rt} m_t \, dt \). We then argue that we can exchange the limit with the expectation with respect to the exogenous path \( \{x_t, t \geq 0\} \); therefore, \( \lim_{r \downarrow 0} \mathbb{E} \left\{ \int_0^{\infty} re^{-rt} m_t \, dt \right\} \) is bounded and a function of \( R \) only, as well. Finally, we use the same approach to show that the first expectation on the right-hand side of Equation (C-14) is also bounded and a function of \( R \) only, completing the proof.

To compute \( \lim_{r \downarrow 0} \int_0^{\infty} re^{-rt} m_t \, dt \), we use Item (ii) and Equation (20), and note that
\[
\lim_{T \to \infty} \frac{1}{T} \int_0^{T} m_t \, dt = M(R, r) \, . 
\] (C-15)

Taking the limit of Equation (C-15) as \( r \downarrow 0 \), we have
\[
\lim_{r \downarrow 0} \lim_{T \to \infty} \frac{1}{T} \int_0^{T} m_t \, dt = \lim_{r \downarrow 0} M(R, r) \\
= \lim_{r \downarrow 0} \int \int m H(dm, dx; p(R, r)) . 
\] (C-16)

The second equality uses Equation (20), and the expression in (C-16) is well defined due to Item (iv), bounded due to Item (i), and independent of \( m \) and of \( \{x_t, t \geq 0\} \).

Then, we note that
\[
\lim_{T \to +\infty} \int_0^{T} re^{-rt} \, dt = \lim_{T \to +\infty} \frac{1}{T} \int_0^{T} dt = 1 \, , 
\] (C-17)

so we can regard these two as distributions. The square of the norm of their difference, as
\( r \downarrow 0, \) is

\[
\lim_{r \downarrow 0} \lim_{T \to \infty} \int_0^T \left( re^{-rt} - \frac{1}{T} \right)^2 dt = \lim_{r \downarrow 0} \lim_{T \to \infty} \int_0^T \left( r^2 e^{-2rt} + \frac{1}{T^2} - 2 \frac{re^{-rt}}{T} \right) dt
\]

\[
= \lim_{r \downarrow 0} \lim_{T \to \infty} \left( \frac{r}{2} - \frac{re^{-2rT}}{2} \right) + \frac{1}{T} - \frac{2}{T} (1 - e^{-rT}) \tag{C-18}
\]

\[
= \lim_{r \downarrow 0} \frac{r}{2} = 0.
\]

Thus, since \( m_t \) is bounded by Item (i),

\[
\lim_{r \downarrow 0} \lim_{T \to \infty} \int_0^T re^{-rt} m_t dt = \lim_{r \downarrow 0} \lim_{T \to \infty} \frac{1}{T} \int_0^T m_t dt
\]

\[
= \int \int m \lim_{r \downarrow 0} h (m, x; p(R, r)) dm dx, \tag{C-19}
\]

where the first equality follows from (C-17) and (C-18), and the second equality uses Equation (C-16).

Then, we note that we can exchange the limit as \( r \downarrow 0 \) with the expectation with respect to \( \{x_t, t \geq 0\} \); that is,

\[
\lim_{r \downarrow 0} \mathbb{E} \left\{ \int_0^T e^{-rt} m_t dt \right\} = \mathbb{E} \left\{ \lim_{r \downarrow 0} \int_0^T e^{-rt} m_t dt \right\} \tag{C-20}
\]

because the argument of the expectation converges as a function of \( r \), as just shown, and because it is bounded. Therefore, we can apply the Lebesgue dominated convergence theorem. Boundedness, for a fixed \( R \) and for all \( r > 0 \), can be shown using Item (i):

\[
\int_0^\infty e^{-rt} m_t dt \leq B(R) \int_0^\infty e^{-rt} dt = B(R),
\]

where the last equality uses \( \int_0^\infty e^{-rt} dt = 1. \)

We now focus on the first term on the right-hand side of Equation (C-14). Using \( \bar{K}_t \), defined in Equation (C-4), we can write

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \bar{K}_t = \lim_{T \to \infty} \frac{1}{T} \sum_{\{t \leq T\}} \left[ I_{\tau_i} K \left( m_{(\tau_i^+)} - m_{(\tau_i^-)}, x_{\tau_i} \right) \right]
\]

\[
= \int \int K(y, x) P(dy, dx; p(R, r)) ,
\]

where the first equality uses Equation (C-4), and the second equality follows from Item (ii).
We can then follow the same steps as in equations (C-16) to (C-20), and the result follows.

C.5 Joint Continuity and Uniform Convergence

We state and prove a lemma that we use in the proof of Proposition 2.

**Lemma 1.** Let $\mathcal{R}$ be a compact space and let $g : \mathcal{R} \times (0, +\infty) \to \mathbb{R}$ such that $g$ is jointly continuous in its arguments. If $g(R, r)$ converges pointwise to zero for all $R \in \mathcal{R}$ as $r \downarrow 0$, then $g(R, r)$ converges uniformly to zero.

**Proof.** By assumption of pointwise convergence to zero, for all $\varepsilon > 0$ and for all $R \in \mathcal{R}$, there exists a $\delta(\varepsilon, R) > 0$ such that $|g(R, r)| < \varepsilon$ for all $0 < r < \delta(\varepsilon, R)$. The function $\delta(\cdot, \varepsilon) : \mathcal{R} \to \mathbb{R}$ can be chosen to be continuous in $R$ because $g$ is continuous. Let $\delta^*(\varepsilon) = \min_{R \in \mathcal{R}} \delta(R, \varepsilon)$, which exists and is finite because the function is continuous on the compact space $\mathcal{R}$, and $\delta^*(\varepsilon) > 0$ because $(\arg\min_{R \in \mathcal{R}} \delta(R, \varepsilon)) \in \mathcal{R}$ and $\delta(R, \varepsilon) > 0$ for all $R \in \mathcal{R}$. Then $|g(R, r)| < \varepsilon$ for all $0 < r < \delta^*(\varepsilon)$ and for all $R \in \mathcal{R}$; that is, $g(R, r)$ converges to zero as $r \downarrow 0$ uniformly with respect to $R$. ■