An Analysis of Coordination Mechanisms for the U.S.
Cash Supply Chain

Milind Dawande, Mili Mehrotra, Vijay Mookerjee, Chelliah Sriskandarajah
School of Management, University of Texas at Dallas, Richardson, TX 75083,
milind@utdallas.edu, mili.mehrotra@student.utdallas.edu, {vijaym, chelliah}@utdallas.edu

The overuse of its currency processing facilities by depository institutions (DIs) has motivated the Federal Reserve (Fed) to impose its new cash recirculation policy. This overuse is characterized by the practice of cross-shipping, where a DI both deposits and withdraws cash of the same denomination in the same business week in the same geographical area. Under the new policy, which came into effect from July 2007, the Fed has imposed a recirculation fee on cross-shipped cash. The Fed intends to use this fee to induce DIs to effectively recirculate cash so that the societal cost of providing cash to the public is lowered.

To examine the efficacy of this mechanism, we first characterize the social optimum and then analyze the response of DIs under a recirculation fee levied on cross-shipped cash. We show that neither a linear recirculation fee, which is the Fed’s current practice, nor a more sophisticated non-linear fee is sufficient to guarantee a socially optimal response from DIs. We then derive a fundamentally different mechanism that induces DIs to self-select the social optimum. Our mechanism incorporates a fairness adjustment that avoids penalizing DIs that recirculate their fair share of cash and rewards DIs that recirculate more than this amount. We demonstrate that the mechanism is easy to implement and tolerates a reasonable amount of imprecision in the problem parameters. We also discuss a concept of welfare-preserving redistribution wherein the Fed allows a group of DIs to reallocate (amongst themselves) their deposits and demand if such a possibility does not increase societal cost. Finally, we analyze the impact of incorporating the custodial inventory program, another component of the Fed’s new policy.

1. Introduction

Despite an increase in the use of electronic payment mechanisms, the total value and volume of physical currency (cash) that is in circulation in the United States (U.S.) has increased rapidly over the past three decades. Cash transactions continue to be favored, primarily for their ease and the anonymity they offer. According to a recent report by Transoft (Transoft International 2007), a company specializing in currency supply chain management solutions, the total value of
cash in circulation in the U.S. has increased from $93.4 billion in 1977 to $783.2 billion in 2006; over the same period, the volume of cash has increased from 7.5 billion to 26.4 billion notes. The Federal Reserve of the U.S. (hereafter referred to as the Fed) estimates that the value of cash in circulation will exceed $1000 billion by 2010; it was $690 billion in 2004, up from $492 billion five years ago (Blacketer and Evetts 2004). A similar increase has been observed in other countries as well. For example, in the United Kingdom, nearly two-thirds of all payments are made in cash, and it is estimated that this will remain the case until the foreseeable future (Association for Payment Clearing Services 2005).

Corresponding to this increasing use of cash, there is an increasing need for fit cash: currency in a suitable condition to be circulated to the public. For instance, in the U.S., the number of ATMs increased by about 200% between 1996 and 2003 (Blacketer and Evetts 2004). We briefly examine the classification of cash based on its physical condition.

1.1. Fit, Used, and Unfit Cash
In the U.S., the Bureau of Engraving and Printing is the sole authority that can produce new cash; this money is then introduced into the currency supply by the Fed. Once in supply, the new cash becomes part of used cash. The term “used cash” broadly refers to all cash that is currently in circulation and is of two types: fit and unfit cash. As mentioned earlier, fit cash refers to cash that is in suitable condition to remain in circulation. There are two main types of fit cash: ATM-Fit cash and Non-ATM-Fit cash. As their names imply, ATM-Fit cash can be dispensed via ATMs while Non-ATM-Fit cash is appropriate for most transactions, but not for ATMs. Finally, unfit cash – soiled, defaced, or torn cash – is not suitable for circulation and is destroyed by the Fed.

The Fed – through its 12 regional branches – is the central component of the currency supply chain. Depository institutions (DIs) demand fit cash from the Fed to meet their customers’ demand and deposit used cash to the Fed. The Fed fit-sorts (i.e., separates fit and unfit cash) the deposited used cash, recirculates fit cash back to the DIs, and destroys unfit cash.

1 Coins are significantly less perishable and have lesser value. Therefore, the issues arising in the circulation of coins are relatively insignificant and not considered in this paper.

2 The term “depository institution” denotes a financial institution that obtains its funds mainly through deposits from the public. This includes commercial banks, savings and loan associations, savings banks, and credit unions.
1.2. The Fed’s Concern and its New Regulations

Succinctly put, the Fed is concerned that DIs are overusing its cash processing services. For their part, DIs view cash on hand as a non-earning asset and deposit as much of the used cash they collect (from their customers) to the Fed as soon as possible and order fit cash from the Fed only when needed. This action of the DIs is understandable: they want to hasten their deposits with the Fed as the credited deposits immediately become available to earn a return – for example, by lending to other institutions (Transsoft International 2007). A result of this behavior by DIs is the increased frequency of their transactions – deposits and withdrawals – with the Fed.

Apart from the fact that deposits to the Fed immediately become earning assets, another reason for the DIs to “dump” most of their deposits of used cash to the Fed and withdraw fit cash as and when needed is to avoid the cumbersome task of sorting the deposited cash into fit and unfit notes. This process, known as fit-sorting, is a sophisticated operation involving expensive machinery and highly skilled labor. Naturally, the DIs are reluctant to invest in fit-sorting and simply avoid it by depositing the collected used cash to the Fed.

As early as 1998, the Fed noticed that roughly 30% to 50% of the currency a DI deposits with the Fed is reordered by the same DI in the same denomination within five days. Such actions by DIs led the Fed to define cross-shipping as depositing fit or non-fit-sorted currency and ordering the same denomination during the same business week within a Federal Reserve zone (Federal Reserve 2003). To fully understand the Fed’s desire to curtail the practice of cross-shipping, we discuss the impact of cross-shipping.

• Increased fit-sorting costs for the Fed: A direct result of increased cross-shipping is a corresponding increase in fit-sorting cost. Over the years, the Fed has had to significantly increase the resources dedicated to fit-sorting used cash and bundling fit cash. Also, the Fed processes a vast majority of notes at facilities that have a near-constant return to scale, i.e., the Fed’s cost increases almost linearly with the volume of cash (Bohn et al. 2001). Consequently, a significant fraction of its annual budget is consumed by these activities. In 2003, for example, $387 million (or about 15%) of the Fed’s total budget of $2.63 billion was dedicated to currency management operations (Blacketer and Evetts 2004).

• Blocked cash and inflationary effect: Excessive deposits of used cash to the Fed (by DIs) result
in cash being locked in an unproductive state (waiting to be fit-sorted or being fit-sorted). When this cash cannot be actively used, productive activities are denied, causing costs to society. On the other hand, the Fed partly addresses the problem by printing extra currency. For example, in 2006, the Fed received and processed 38 billion used notes, destroyed seven billion unfit notes, and paid out 39 billion in new and (used) fit notes (Transoft International 2007). While printing extra cash alleviates the “blocked cash” problem, this activity has an inflationary effect. These two effects – blocking delays and inflationary effects resulting from the printing of new cash to alleviate these delays – are indeed related to each other. On the one hand, the Fed can reduce the blocking delays by printing new cash and, consequently, increase the inflationary effect. On the other hand, if printing of new cash is eliminated or reduced, then there is an increase in the blocking delays. Presumably, the Fed carefully manages these two effects. In other words, the Fed tolerates some inflation as well as some blocking delays. In this context, it is instructive to note that blocked cash consists of deposits held by customers/retailers at various DIs. Therefore, blocked cash is included as part of currency in circulation. Both blocked cash and newly printed cash contribute to currency in circulation and, thereby, increasing the money supply. Hence, it is reasonable to consider the combined effect of fit-sorting followed by extra printing as an inflationary burden on the financial system. A third, but relatively minor, component is the physical cost of printing extra cash. The root cause of all these effects is the amount and time for which cash is locked in an unproductive state.

To summarize, an increase in cross-shipping not only increases the Fed’s cost of physically fit-sorting cash (direct cost) but also results in the blocking of cash in an unproductive state and unnecessary printing of new cash (indirect cost). This combined impact is indicated by the manner in which the Fed has chosen to define the notion of cross-shipping. If the Fed were concerned only about the cost of fit-sorting used cash, an obvious solution would have been to impose a fit-sorting fee on deposits of used cash (to the Fed) by a DI. Such a fee has been argued for in the literature (Lacker 1993), although in a different context. Instead, by asking DIs to avoid or minimize cross-shipping, the Fed has aimed for a broader goal of encouraging DIs to re-circulate cash themselves by fit-sorting. If the DIs were to invest in building fit-sorting capabilities, then both of the Fed’s concerns mentioned above – increased fit-sorting costs and increased volume of blocked cash – will
be addressed.

Unfortunately, the Fed’s earlier warnings about reducing cross-shipping were largely ignored by the DIs, primarily because it was not clear on exactly how to reduce cross-shipping. The DIs were essentially getting fit-sorting services for free from the Fed and following the Fed’s guidelines involved significant investments in developing these capabilities themselves (Transoft International 2007). During 2004, deposits of 7.2 billion fit USD 10 and USD 20 notes were followed or preceded by orders of the same denomination by the same institution in the same business week in the same geographical area (Federal Reserve 2006a). This pattern suggests that some DIs are relying on the Fed for a substantial amount of currency processing.

Eventually, with the intention to “encourage private-sector behavioral changes that would lower the overall societal costs of cash processing and distribution by curtailing overuse of a free governmental service” (Federal Reserve 2003), the Fed proposed its new cash re-circulation policy that came into effect from July 2007. We describe the two main components of this new policy:

1. Recirculation Fee: A DI will pay a per-bundle fee (a bundle consists of 1000 notes of the same denomination) on cross-shipped cash. The fee is not applied to deposits of unfit cash. Also, the fee is not applicable to deposits of $50 and $100 notes because these notes are a relatively minor component of cross-shipped currency and, more importantly, because of the risk that DIs might recirculate high-denomination counterfeit notes (Federal Reserve 2003).

2. Custodial Inventory Program (Federal Reserve 2003): This program allows DIs to deposit cash into custodial inventories without subjecting the withdrawals made in the same week to recirculation fees. A custodial inventory for a DI contains fit cash deposited to the Federal Reserve (and is, therefore, an earning asset for the DI) but is located within the DI’s secured facility and is segregated from its operating cash. An additional benefit is that custodial inventories may allow DIs to avoid the costs of preparing and transporting their temporarily surplus currency to and from Federal Reserve offices (Federal Reserve 2003). Because DIs may use custodial inventory to meet demand, this program provides an incentive for DIs to fit-sort used cash at their expense and to keep the resulting fit cash in custodial inventory. The program further encourages fit-sorting by the DIs by forbidding cash withdrawn from the Fed from being deposited into custodial inventory. In the first few years of its introduction, participation (by DIs) in the custodial inventory program
is expected to be slow as it first requires developing some infrastructure such as segregated storage of custodial inventory, acquiring fit-sorting capability, etc. Therefore, it is the recirculation fee that will be the primary focus of our discussion. Later, in Section 5.3, we will briefly analyze the impact of custodial inventories.

To summarize our discussion thus far, the Fed intends to use the recirculation fee on cross-shipped cash as a mechanism for inducing DIs to behave in a manner so that the societal cost of providing currency to the public (which is the Fed’s eventual goal) is minimized. In this paper, we argue that the mechanism of imposing such a recirculation fee on cross-shipped currency will, in general, not induce DIs to behave in a socially optimal manner. Furthermore, we show that the percentage gap between the optimal social cost and the social cost under the Fed’s mechanism can be high. We also illustrate situations where the Fed’s mechanism can perform well. We propose a new mechanism that is fundamentally different from the Fed’s fee. In our mechanism, a DI is charged for deposits (if any) of non-fit-sorted used cash to the Fed only if it also makes positive withdrawals from the Fed in the same business week. We show that such a “cross-shipping adjusted fit-sorting fee” induces socially optimal behavior from DIs. Our mechanism decides the adjusted fit-sorting fee for a DI based on its inherent ability to recirculate cash. This property ensures that the fee is viewed as fair by the DIs. We argue against the possibility of collusion and discuss a concept of welfare-preserving redistribution, where the Fed knowingly allows a coalition of DIs to internally reallocate deposits and demand if such a possibility does not increase societal cost. We report on a comprehensive numerical study (based on data gleaned from public sources) conducted to analyze the impact of uncertainty on the optimal decisions. Our results show that, while uncertainty does have an impact, the magnitude of this impact is small even under a considerable amount of uncertainty. We also analyze the impact of the Fed’s custodial inventory program and argue that custodial inventory incentivizes DIs to fit sort more than they otherwise would have under the fit-sorting fee alone.

The rest of the paper is organized as follows. Section 2 presents the relevant literature. In Section 3, we first set up the social planner’s problem and then obtain the social optimum. Section 4 argues the need for a coordinating mechanism and then explore the Fed’s current mechanism. We present and analyze our new mechanism in Section 5. In Section 5.3, we discuss the impact of the custodial inventory – the second component of the Fed’s newly instituted policy. Section 6 offers
further discussions, identifies topics for further study, and concludes the paper.

2. Literature Review

The cash supply chain is a closed-loop supply chain (Rajamani et al. 2006, Geismar et al. 2007). Unlike typical consumer products, cash re-circulates in the economy. After new cash is printed by the Bureau of Engraving and Printing and released into the economy by the Federal Reserve, it is used in the transactions between the various stakeholders – DIs, businesses, consumers, etc – of the economy. Used cash (both fit and unfit cash) eventually makes its way back to the Fed via deposits from the DIs. While unfit cash is destroyed by the Fed, fit cash is re-circulated. Thus, the cash supply chain is conceptually similar to the flow of goods in reverse logistics and remanufacturing (Fleischmann 2001, Guide and Van Wassenhove 2003).

While early research in closed-loop supply chains focused more on operational issues (see, e.g. van der Laan et al. 1999, Toktay et al. 2000, Guide 2000, Ferrer and Ketzenberg 2004), recent work has investigated strategic interactions among the participants of closed-loop supply chains. Heese et al. (2005) and Webster and Mitra (2007) analyze the competitive advantage for a firm engaged in the activity of product take-back. Savaskan et al. (2004) address the issue of channel choice and coordination in a closed-loop supply chain with competing retailers. Majumder and Groenevelt (2001) study the competition between an original equipment manufacturer and a remanufacturer, and suggest incentives to increase the fraction of remanufacturable products. Webster and Mitra (2008) analyze a similar situation under government subsidies. Most of these studies consider product returns that occur at the end of a product’s life cycle or due to overstocking. In a recent study, Guide et al. (2006) examine the time-sensitivity of product returns and suggest that firms with high return rates should focus on responsiveness (speed and decentralization). Another recent work on false-failure returns by Ferguson et al. (2006) is relatively close to our work. False-failure returns refer to those goods that are returned to the retailer but have no functional or cosmetic defects. They propose a target rebate contract to incentivize a retailer to increase her effort and, thus, reduce the number of false-failure returns. Structurally, the primary cause of mis-coordination in the cash supply chain – overuse of the Fed’s fit-sorting facilities resulting from deposits of non-fit-sorted used cash – is similar to that of faulty returns of goods by consumers.

The above literature considers coordination between two or more for-profit organizations (e.g.,
suppliers and retailers, manufacturer and remanufacturer). As such, (i) the analysis is often game-theoretic and (ii) the scope of the coordination is largely limited to the parties involved. In contrast, the coordination problem in our case involves a societal organization (i.e., the Fed) and a (typically) for-profit firm (i.e., a DI). Therefore, the coordination scheme assumes the form of a policy statement. Furthermore, the impact of the mis-coordination is borne by all sectors of the economy, and not just the banking sector.

In its spirit, the coordination issue addressed in this paper is similar to that of Emissions Trading of greenhouse gases (Barrett et al. 1992, Edmonds et al. 1999). The negative effects of the pollutants emitted by an entity (a firm, a group of firms, or a country) impact the global environment and the basic idea is to incentivize the entity to reduce this negative effect (Barrett 1994, Hoel 1997, 2001). However, the reporting, verification, and enforcement of the emissions (e.g., the carbon footprint) from an entity is a complex process (Climate Registry 2009). In comparison, in our case, the Fed has precise knowledge of the used-cash deposits from each DI.

3. The Model

In this section, we describe the flow of cash, develop the social planner’s problem, and obtain the social optimum. We start by formalizing the relevant notation.

3.1. Notation

Our analysis considers cash of a single denomination. There are three main reasons: First, a DI’s weekly demand for fit cash (from the Fed) and the deposits of used-cash (received from its customers) are denomination-specific. Second, all the transactions between the Fed and a DI are in bundles, with each consisting of 1000 notes of a single denomination. Third, as we will see shortly, all relevant costs are calculated per-bundle. Tables 1-2 summarize our notation. We will develop supplemental notation as and when required later in this and the subsequent sections.

Based on our discussions with practitioners, we assume that the weekly demand-to-deposits ratio \( \rho_i \) for a DI \( i \in I \) is a constant. Thus, the demand can be expressed as a function of deposits. Also, a typical DI has strictly positive used-cash deposit each week. Hence, our results are based on the assumption that the weekly used-cash deposit \( x_i \) at DI \( i \) is strictly positive. The amount of fit-sorting a DI \( i \in I \) needs depends on the relationship between its demand-to-deposits ratio \( \rho_i \) and the constant \( g \). If \( \rho_i < g \), then fit-sorting a \( \frac{\rho_i}{g} \) fraction of its used-cash deposits generates the fit
### Table 1 Notation and Model Parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>Number of DIs in the system.</td>
</tr>
<tr>
<td>(I)</td>
<td>Index set for DIs.</td>
</tr>
<tr>
<td>(x_i)</td>
<td>Weekly deposit of used cash (in number of bundles) to a DI (i \in I).</td>
</tr>
<tr>
<td>(\rho_i)</td>
<td>Ratio of fit-cash demand to used-cash deposits for a DI (i \in I).</td>
</tr>
<tr>
<td>(\tau)</td>
<td>The function representing fit-sorting cost for the Fed with the number of bundles processed as the argument.</td>
</tr>
<tr>
<td>(\Gamma)</td>
<td>The function representing indirect cost (in the system) with the number of bundles of used-cash deposited to the Fed as the argument.</td>
</tr>
<tr>
<td>(\kappa)</td>
<td>The function representing cash processing cost for a DI with the number of bundles processed as the argument.</td>
</tr>
<tr>
<td>(e)</td>
<td>The function representing the cross-shipping fee with the number of cross-shipped bundles as the argument.</td>
</tr>
<tr>
<td>(T)</td>
<td>The function representing the transportation cost for a DI with the number of bundles as the argument.</td>
</tr>
<tr>
<td>(h)</td>
<td>Per-unit per-period fit-cash inventory holding cost.</td>
</tr>
<tr>
<td>(g)</td>
<td>Percentage of fit cash in the used-cash deposit. In practice, this value is typically around 75%.</td>
</tr>
</tbody>
</table>

### Table 2 Decision Variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\phi_i)</td>
<td>Fraction of used cash to be fit-sorted by a DI (i \in I) (as desired by the social planner).</td>
</tr>
<tr>
<td>(\psi_i)</td>
<td>Fraction of used cash to be fit-sorted by a DI (i \in I) (decision made by the DI).</td>
</tr>
</tbody>
</table>

Cash necessary to meet customers’ demand. On the other hand, if \(\rho_i \geq g\), then the total demand of fit cash from customers exceeds (or equals) the maximum amount of fit cash that can be generated from fit-sorting used-cash deposits. We, therefore, characterize a DI to be one of two types: (i) **Type 1** if \(\rho_i \geq g\), and (ii) **Type 2** if \(\rho_i < g\). Intuitively, a Type 2 DI has the potential to generate (by fit-sorting) more than its demand of fit cash. On the other hand, a Type 1 DI always faces a fit-cash deficit since it cannot generate enough fit cash (by fit-sorting) to meet its demand.

#### 3.2. The Flow of Cash

Figure 1 depicts the weekly flow of cash in the cash supply chain. A DI \(i\) receives a certain amount of used cash \((x_i\) bundles) from its customers, fit-sorts a \(\phi_i\) fraction of this used cash and sends (i) \((1-g)\phi_i x_i\) bundles of unfit cash and (ii) the remaining \((1-\phi_i)x_i\) bundles of used cash to the Fed. If \(\phi_i \leq \frac{\rho_i}{g}\), then to meet its demand for fit cash (which equals \(\rho_i x_i\)), the DI demands \((\rho_i x_i - g \phi_i x_i)\) bundles of fit cash from the Fed. If \(\phi_i > \frac{\rho_i}{g}\), then it deposits the additional \((g \phi_i x_i - \rho_i x_i)\) amount of fit cash to the Fed. Thus, the DI transacts (i.e., either deposits or withdraws) \(|\rho_i x_i - g \phi_i x_i|\) bundles of fit-cash with the Fed. The DI meets its customers’ demand by combining the cash it fit-sorts
3.3. Social Planner’s Problem

Recall from Section 1.2 our aim to analyze the recirculation fee as a mechanism to enforce DIs to behave in socially optimal manner. Thus, we first consider a social planner who envisions the total societal (or system) cost (i.e., the costs incurred by all the stakeholders) and optimizes the fit-sorting fractions ($\phi_i$) for the DIs. We now describe the components of the total system cost:

1. **Fed’s Fit-Sorting Cost and Indirect System Cost**: The Fed fit-sorts the used-cash deposited by all the DIs. As discussed earlier, the physical fit-sorting costs at the Fed’s facilities have a near-constant return to scale (Bohn et al. 2001). Therefore, we assume that the fit-sorting cost function $\tau$ for the Fed is linear in the number of bundles fit-sorted. A DI $i \in I$ deposits $(1 - \phi_i)x_i$ non-fit-sorted used cash to the Fed. Thus, the total fit-sorting cost at the Fed is $\sum_{i \in I} \tau((1 - \phi_i)x_i)$. Note that the DI’s total deposit of used cash to the Fed is $(1 - g\phi_i)x_i$, which is the sum of non-fit-sorted used cash ($(1 - \phi_i)x_i$) and unfit cash ($(1 - g)\phi_i x_i$). The Fed simply destroys the unfit cash. Hence, the Fed’s fit-sorting cost depends only on the amount of non-fit-sorted used cash deposited by the DI.

Recall from Section 1.2 our discussion about indirect costs for the system from used-cash deposits. Since the Fed faces demand of fit-cash from DIs continuously, a used-cash deposit by a DI immediately makes a corresponding contribution to the idle cash that is waiting to be fit-sorted; thus,
the negative externality (imposed on the system) from a used-cash deposit is such that its impact is immediate. Within a Federal Reserve zone (there are a total of 12 Fed zones in the U.S.), a DI and all its branches must transact with the Fed as a single entity (Federal Reserve 2003). Thus, all the branches of a DI in a single zone communicate (i.e., deposit/withdraw cash) with the Fed as a single entity. Therefore, for a single DI, the amount of weekly used cash deposits are typically significant. Each DI’s deposits are separately fit-sorted and released at the Fed’s facility (this helps the Fed to easily monitor and identify sources of counterfeit cash). Therefore, we consider the cost at the level of each deposit and add these costs to arrive at the total indirect cost. A used-cash deposit of \((1 - \phi_i)x_i\) by a DI \(i \in I\) results in an indirect cost of \(\Gamma((1 - \phi_i)x_i)\). Over all DIs, this cost is \(\sum_{i \in I} \Gamma((1 - \phi_i)x_i)\), where \(\Gamma\) is the indirect cost function.

2. Costs Incurred by a DI \(i \in I\):

- **Fit-Sorting Cost:** The DI fit-sorts a \(\phi_i\) fraction of its used-cash deposit \(x_i\). The corresponding fit-sorting cost is \(\kappa(\phi_i x_i)\), where \(\kappa\) is the fit-sorting cost function for DIs.

- **Transportation Cost:** The DI (1) deposits \((1 - \phi_i)x_i\) bundles of non-fit-sorted used-cash as well as the unfit cash (which equals \((1 - g)\phi_i x_i\) bundles) to the Fed and (2) demands \((\rho_i x_i - g\phi_i x_i)\) bundles of fit-cash from the Fed or deposits \((g\phi_i x_i - \rho_i x_i)\) bundles of fit-cash to the Fed. Thus, the DI incurs a transportation cost of \(T((1 - g\phi_i)x_i + |\rho_i x_i - g\phi_i x_i|)\).

- **Inventory Holding Cost:** Cash deposited with the Fed is an earning asset. A DI can earn a positive return on its cash, credited to its account with the Fed, by lending to some other institution via the Federal Funds Market (Federal Reserve 2006a). By holding fit cash at its own vault, the DI incurs an opportunity cost of \(h \min\{\rho_i x_i, g\phi_i x_i\}\): a DI’s demand for fit cash is \(\rho_i x_i\) and the amount of fit cash generated is \(g\phi_i x_i\). If the fit-sorted cash is less than the demand, then the DI holds only the fit-sorted cash \(g\phi_i x_i\) in its vault. If the fit-sorted cash is more than the required amount, the DI holds \(\rho_i x_i\) bundles of fit-sorted cash (which is enough to meet its demand); the remaining amount \(g\phi_i x_i - \rho_i x_i\) is transported back to the Fed. Here, it is important to note that the Fed does not pay any interest on the cash a DI deposits with it.

Over all DIs, the total incurred cost is:

\[
\sum_{i \in I} \left( \frac{\kappa(\phi_i x_i)}{\text{Fit-Sorting Cost}} + \frac{T((1 - g\phi_i)x_i + |\rho_i x_i - g\phi_i x_i|)}{\text{Transportation Cost}} + h \min\{\rho_i x_i, g\phi_i x_i\} \right)
\]
Let $\phi = (\phi_1, \phi_2, \ldots, \phi_n)$. The total system cost is

$$SC(\phi) = \sum_{i \in I} \tau((1 - \phi_i)x_i) + \sum_{i \in I} \Gamma((1 - \phi_i)x_i) + \sum_{i \in I} \left\{ \kappa_i(\phi_i x_i) + T((1 - g\phi_i)x_i + | \rho_i x_i - g\phi_i x_i |) + h \min\{ \rho_i x_i, g\phi_i x_i \} \right\}$$

Then, the social planner’s problem is $\min_{\phi} SC(\phi)$, where $0 \leq \phi_i \leq 1, i \in I$.

It is clear that the severity of indirect cost (i.e., $\Gamma(\cdot)$) increases with the volume of used-cash deposits to the Fed. The fit-sorting process itself involves several sub-processes, each of which may have internal waiting (Bohn et al. 2001). We draw from results in bulk arrival queueing systems (Shanthikumar 1994), where the total time (i.e., internal waiting time plus processing time) is a convex and increasing function of the batch size. By analogy, we propose that the processing time for a DI’s deposit is convex and increasing in the size of the deposit. Thus, from the systemic view, it is reasonable to conclude that the indirect cost $\Gamma(w)$ resulting from a used-cash deposit of $w$ by a DI is convex and increasing. The exact functional form for the indirect cost, however, is difficult to estimate because of the complex nature of the variety of effects that result from blocked cash. Broadly, given the lack of knowledge we are faced with, this function can be visualized as one of two types: one that grows in a “modest” manner versus one whose growth is “severe”. In the literature (e.g., in computational complexity theory), polynomials (resp., exponential functions) are used to exemplify modest (resp., severe) growth. Motivated by this classification, we consider two distinct forms for the indirect cost (as a function of the volume $w$): (i) $\Gamma(w) = \gamma w^p, p > 1$ and (ii) $\Gamma(w) = \gamma a^w, a > 1$. For convenience, we present our analysis using the first form\(^3\); later, in Section 3.4, we also summarize the results with the second form. Table 3 shows the forms of the various cost functions used in our analysis. As mentioned above, a DI’s per-bundle holding cost of fit cash is the cost of lost opportunity of lending via the Federal Funds Market and, therefore, does not depend on the index $i$ of the DI. The functions of transportation and fit-sorting are typically executed by third-party providers such as Brinks’ Inc., who have standard per-bundle

\(^3\)The polynomial assumption on $\Gamma$ is not necessary. Our results hold for any convex and increasing function $\Gamma$. We use this form to obtain an explicit solution and to obtain insights.
rates\textsuperscript{4}. Therefore, the per-bundle fit-sorting cost $\alpha$ and the per-bundle transportation cost $\beta$ are assumed to be independent of $i$. Substituting the forms in Table 3, the social planner’s objective is:

$$SC(\phi) = \sum_{i \in I} (\lambda(1 - \phi_i)x_i + \gamma((1 - \phi_i)x_i)^p + \alpha\phi_ix_i + \beta\{(1 - g\phi_i)x_i + |\rho_i x_i - g\phi_i x_i|\} + h \min\{\rho_i x_i, g\phi_i x_i\}) - \sum_{i \in I} \text{Fed's Inventory Holding Cost}$$

where $0 \leq \phi_i \leq 1, i \in I$. Next, we obtain the social optimum.

3.4. The Social Optimum

Let $\Upsilon(\phi_i) = [\lambda(1 - \phi_i)x_i + \gamma((1 - \phi_i)x_i)^p + \alpha\phi_ix_i + \beta\{(1 - g\phi_i)x_i + |\rho_i x_i - g\phi_i x_i|\} + h \min\{\rho_i x_i, g\phi_i x_i\}]$. Then, the social planner’s objective function is $SC(\phi) = \sum_{i \in I} \Upsilon(\phi_i)$, and is separable in $i \in I$. Thus, for each DI $i \in I$, we optimize $\Upsilon(\phi_i)$ with respect to $\phi_i$. Since $\Upsilon(\phi_i)$ is not differentiable at $\phi_i = \frac{\omega_i}{g}$, we intend to optimize it the following way: Analyze the objective function separately under two constraints, namely Case 1: $\phi_i \leq \frac{\omega_i}{g}$ and Case 2: $\phi_i \geq \frac{\omega_i}{g}$, obtain the optimum under each case, and then find the unique optimum for the social planner’s objective by comparing the two optima. For clarity, let us denote $\phi_i = \phi_{i1}$ (resp., $\Upsilon = \Upsilon_1$) and $\phi_i = \phi_{i2}$ (resp., $\Upsilon = \Upsilon_2$), respectively, for the two cases. Thus, $\Upsilon(\phi_i) = \Upsilon_1(\phi_{i1})$ (resp., $\Upsilon_2(\phi_{i2})$) for Case 1 (resp., Case 2).

The social planner’s unique optimum, say $\phi^*_{i}$, can be obtained by comparing the objectives for the two cases. Thus, $\phi^*_{i} = \arg\min\{\Upsilon_1(\phi_{i1}), \Upsilon_2(\phi_{i2})\}$. Based on this comparison, we specify the social planner’s unique optimum in Propositions 1 and 2. The proofs of all the results are provided in the Online Supplement.

**Proposition 1.** For a Type 1 DI $i \in I$, only Case 1 is relevant. The unique social optimum is

\textsuperscript{4}The problem discussed in the paper originates from our interaction with Brink’s Inc., a third-party logistics provider that traditionally offers secure-logistics services to DIs. The services that Brink’s typically offers include (a) collecting used cash from the branches, (b) counting and bundling cash, (c) distributing fit cash to the branches, (d) refilling ATM machines, and (e) transporting cash to and from the Fed. The Fed’s new guidelines have presented a new business opportunity – offering fit-sorting and custodial inventory services – for Brink’s and other similar providers such as Loomis, Fargo & Co., Dunbar Armored, etc.
\[ \phi^*_i = \begin{cases} 1, & \text{if } (\alpha + hg \leq 2\beta g + \lambda) \\ \max \{1 - \frac{\alpha^*}{x_i}, 0\}, & \text{otherwise.} \end{cases} \]

where \( c^* = \left( \frac{\alpha - 2\beta g + \lambda}{x_i \rho} \right)^{1/\tau}. \)

For a Type 2 DI, to find the unique optimum, we compare the objectives under the two cases. This comparison results in relationships among the various parameters and classifies the optimum solution under these relationships; Proposition 2 provides the complete classification. Figure 2 shows some of these possibilities. For example, case (c) represents the following parameter relationships:

1. when fit-sorting at the Fed and transportation of cash to and from the Fed is more expensive than fit-sorting at the DI’s facility and holding fit-sorted cash at the DI (i.e., \( \alpha + hg \leq 2\beta g + \lambda \))
2. the fit-sorting cost at the DI is less than that at the Fed (i.e., \( \alpha < \lambda \)).

Intuitively, under these two conditions, the social planner prefers that the DI do all the fit-sorting; hence \( \phi^*_i = 1. \)

![Figure 2: Optimum Fit-Sorting Fraction for Type 2 DI (Societal cost for a DI is in millions of dollars)](image)

**Proposition 2.** For a Type 2 DI \( i \in I \), the unique social optimum is specified by the following cases:

- If \( (\alpha + hg \leq 2\beta g + \lambda) \), then

\[ \phi^*_i = \begin{cases} 1, & \text{if } (\alpha \leq \lambda) \\ \max \{1 - \frac{\alpha^*}{x_i}, \frac{\rho}{\rho} \}, & \text{otherwise.} \end{cases} \]

- If \( (\alpha + hg > 2\beta g + \lambda) \) and \( (\alpha \leq \lambda) \), then
if $c^* < x_i$, then
\[
\phi_i^* = \begin{cases} 
1 & \text{if } (1 - \frac{\phi_i}{g})x_i \geq \min\{c^*, \frac{\gamma(p-1)x^p}{hg-2\beta g}\} \\
1 - \frac{c^*}{x_i} & \text{otherwise.}
\end{cases}
\]

if $c^* \geq x_i$, then
\[
\phi_i^* = \begin{cases} 
1 & \text{if } (1 - \frac{\phi_i}{g})x_i > \frac{\gamma(px_i(c^*p-1)-x^p)}{hg-2\beta g} \\
0 & \text{otherwise.}
\end{cases}
\]

If $(\alpha + hg > 2\beta g + \lambda)$ and $(\alpha > \lambda)$, then
\[
\text{if } (1 - \frac{\phi_i}{g})x_i \geq c^* \text{ then } \phi_i^* = \max\{1 - \frac{d^*}{x_i}, 0\}. 
\]
\[
\text{if } (1 - \frac{\phi_i}{g})x_i \leq \min\{c^*, d^*\}, \text{ then } \phi_i^* = \max\{1 - \frac{c^*}{x_i}, 0\}. 
\]
• if $d^* < (1 - \frac{\phi_i}{g})x_i < c^*$, and
  o if $c^* < x_i$, then
\[
\phi_i^* = \begin{cases} 
1 - \frac{d^*}{x_i} & \text{if } (1 - \frac{\phi_i}{g})x_i > \frac{\gamma(px_i(c^*p-1)-x^p)}{hg-2\beta g} \\
1 - \frac{c^*}{x_i} & \text{otherwise.}
\end{cases}
\]
  o if $c^* \geq x_i$, then
\[
\phi_i^* = \begin{cases} 
1 - \frac{d^*}{x_i} & \text{if } (1 - \frac{\phi_i}{g})x_i > \frac{\gamma(px_i(c^*p-1)-x^p)}{hg-2\beta g} \\
0 & \text{otherwise.}
\end{cases}
\]

where $d^* = (\frac{\alpha - \lambda}{p\gamma})\frac{1}{p-1}$.

The results above show that the social planner is sensitive to the characteristics of a DI (i.e., $x_i$ and $\rho_i$) in determining its optimal fit-sorting fraction. This is natural since the negative externality (imposed on society) of (non-fit-sorted) used cash deposited by a DI depends on its amount and not on what fraction of the DI’s total deposits it constitutes. In turn, for a DI, the optimal value of this amount is determined by its specific characteristics as well as the other system parameters. For a Type 1 DI, the optimum fit-sorting fraction $\phi^*_i$ can take one of the three values: 0, $1 - \frac{c^*}{x_i}$, 1 (Figure 3). Intuitively, if fit-sorting at the Fed and transportation of cash to and from the Fed is more expensive than the fit-sorting at the DI’s facility and holding the fit-sorted cash at the DI (i.e., $\alpha + hg \leq 2\beta g + \lambda$), then the social planner prefers that the DI do all the fit-sorting. Hence, in this case, $\phi^*_i = 1$. Otherwise, the DI either fit-sorts nothing (i.e., $\phi^*_i = 0$) or fit-sorts a $1 - \frac{c^*}{x_i}$ fraction of its used cash. For a Type 2 DI, the optimum fit-sorting fraction $\phi^*_i$ can take one of the five values: 0, $1 - \frac{c^*}{x_i}$, $\frac{\phi_i}{g}$, $1 - \frac{d^*}{x_i}$, 1 (Figure 3). Note that a Type 2 DI has the capability of generating more fit cash than its required amount. To meet its demand for fit cash, a Type 2 DI needs to fit
sort a $\frac{\rho_i}{g}$ fraction of its used-cash deposits. The social objective takes two different forms on the two sides (left-hand-side and right-hand-side) of $\frac{\rho_i}{g}$. On the left-hand-side of $\frac{\rho_i}{g}$, the behavior of a Type 2 DI is similar to that of a Type 1 DI; thus, its fit-sorting fraction can be either 0, $1 - \frac{\psi_i}{x_i}$, or $\frac{\rho_i}{g} < 1$. On the right-hand-side of $\frac{\rho_i}{g}$ (corresponding to the possibility of fit-sorting more than its requirement), the optimal fit-sorting fraction can be either $\frac{\rho_i}{g}$, $1 - \frac{\psi_i}{x_i}$, or 1.

Figure 4 depicts the behavior of the optimum fit-sorting fraction $\phi_i^*$ with respect to $x_i$ and $\rho_i$ for the case when $\alpha + hg > 2\beta g + \lambda$ and $\alpha > \lambda$. For a Type 1 DI, $\phi_i^*$ is independent of $\rho_i$ (Proposition 1; see Figure 4(a)). However, for a Type 2 DI, $\phi_i^*$ depends on $\rho_i$ and can take one of the following four values under the above conditions: 0, $1 - \frac{\psi_i}{x_i}$, $1 - \frac{\psi_i}{x_i}$, and $\frac{\rho_i}{g}$. We did not achieve an optimal fit-sorting fraction of 0 or $\frac{\rho_i}{g}$ for any of the Type 2 DIs considered in the figure. Thus, the jump in Figure 4(b) indicates a shift from $\phi^* = 1 - \frac{\psi_i}{x_i}$ to $\phi^* = 1 - \frac{\psi_i}{x_i}$ as $\rho_i$ and $x_i$ vary.

Remark 1. The preceding analysis assumed the “polynomial” form for the indirect system cost: $\Gamma(w) = \gamma w^p, p > 1$. Recall from Section 3.3 our mention of the “exponential” form $\Gamma(w) = \gamma a^w, a > 1$. Under this form, solutions to the first order conditions of the social planner’s problem when $(\alpha + hg > 2\beta g + \lambda)$ and $(\alpha > \lambda)$ are:

$$\phi_{i1} = 1 - \frac{\log_a\left(\frac{\alpha - 2\beta g + h g - \lambda}{\gamma \ln a}\right)}{x_i}; \quad \phi_{i2} = 1 - \frac{\log\left(\frac{\alpha - \lambda}{\gamma \ln a}\right)}{x_i}$$

The unique interior and boundary solutions under this and the other cases can be obtained as before.

In the next section, we analyze the efficacy of the Fed’s coordinating mechanism. Later, in Section 5, we will propose and analyze a structurally different kind of fee.
4. Analysis of the Fed’s Coordination Mechanism

We start this section by justifying the need for a coordinating mechanism between the Fed and the DIs to achieve a social optimum. Next, in Section 4.2, we consider the Fed’s chosen mechanism of levying a recirculation fee on cross-shipped cash, and develop a DI’s objective function under this mechanism. We show that the Fed’s current mechanism (as well as its non-linear extension) is not a coordinating one. Finally, in Section 4.3, we describe a numerical study to compare the performance of the Fed’s mechanism with the social optimum. Our results show that the social cost corresponding to the fit-sorting decisions under the Fed’s mechanism can be reasonably higher than the optimal social cost.

4.1. Need for a Coordinating Mechanism

The components of the total cost incurred by a DI \( i \in I \) are as follows:

1. **Fit-Sorting Cost:** The cost of fit-sorting a \( \psi_i \) fraction of its used-cash deposit \( x_i \) is \( \kappa(\psi_i x_i) \).

2. **Transportation Cost:** The DI deposits a total of \((1 - g \psi_i)x_i\) mixed (used and unfit) cash to the Fed: \((1 - g)\psi_i x_i\) unfit cash (resulting from the fit-sorting operation) and \((1 - \psi_i)x_i\) non-fit-sorted used-cash. Furthermore, the DI either withdraws \((\rho_i x_i - g \psi_i x_i)\) fit cash from the Fed or deposits \((g \psi_i x_i - \rho_i x_i)\) fit cash to the Fed. Thus, the total transportation cost incurred is \( T((1 - g \psi_i)x_i + |(\rho_i x_i - g \psi_i x_i)|) \).
3. Holding Cost: The DI holds \( \min\{g\psi ix_i, \rho_i x_i\} \) fit-cash in its own vault and incurs a corresponding cost of \( h \min\{g\psi ix_i, \rho_i x_i\} \).

The DI’s objective is

\[
\min_{\psi_i} \kappa(\psi_i x_i) + T[(1 - g\psi_i)x_i + |\rho_i x_i - g\psi_i x_i|] + h \min\{\rho_i x_i, g\psi_i x_i\}
\]

or

\[
\min_{\psi_i} \alpha \psi_i x_i + \beta[(1 - g\psi_i)x_i + |\rho_i x_i - g\psi_i x_i|] + h \min\{\rho_i x_i, g\psi_i x_i\}
\]

where \( 0 \leq \psi_i \leq 1 \).

We analyze the above objective function for each of the two types of DIs (a DI, of course, is aware of its type): Type 1 and Type 2 (see Section 3.1). For a Type 1 DI, \( \psi_i \leq 1 \leq \rho_i g \) and for a Type 2 DI, \( \psi_i \) can exceed \( \frac{\rho_i g}{E} \).

For a DI \( i \in I \), let \( \tilde{\psi}_i \) denote the optimum fraction. It is easy to observe that, without any coordination mechanism, \( \tilde{\psi}_i \in \{0, \frac{\rho_i g}{E}, 1\} \) for each DI. Comparing with the social optimum (Section 3.4), we conclude that this is a strictly suboptimal solution for the social planner’s problem. Thus, a coordination mechanism is indeed required to induce DIs to self-select the social optimum. The corollary below directly follows from the comparison of the fit-sorting fractions above and those in the social optimum (Propositions 1 and 2). The proof is straightforward and is therefore omitted.

**Corollary 1.** The optimal fit-sorting fraction chosen by an individual DI (absent any coordination mechanism) is less than or equal to its fit-sorting fraction in the social optimum. In other words, without any coordination mechanism, individual DIs always under-sort relative to the social optimum.

Next, we consider the Fed’s mechanism and develop a DI’s objective under this mechanism.

4.2. Fed’s Coordination Mechanism

Recall from Section 1.2 the Fed’s mechanism of imposing a recirculation fee on cross-shipped cash. Cross-Shipping occurs if a DI \( i \in I \) withdraws fit cash from the Fed and deposits (non-fit-sorted) used-cash to the Fed in the same business week. In this case, the cross-shipped amount \( y_i \) is defined as the minimum of (a) the fit-cash withdrawal from the Fed (equal to \( \rho_i x_i - g\psi_i x_i \)) and (b) the amount of fit-cash (equal to \( g(1 - \psi_i x_i) \)) in the used-cash deposited to the Fed. Thus,
Let $e(y_i)$ be the recirculation fee charged by the Fed on a cross-shipped amount of $y_i$ bundles. Then, the objective of the DI under this fee is

$$
\min_{\psi_i} e(y_i) + \alpha \psi_i x_i + \beta \left[ (1 - g \psi_i) x_i + |\rho_i x_i - g \psi_i x_i| \right] + h \min \{ \rho_i x_i, g \psi_i x_i \}
$$

We analyze the Fed’s current “linear” recirculation fee as a mechanism to induce the DIs to choose $\phi^*$ as their optimum. The Fed’s current recirculation fee is $5 for each cross-shipped bundle.

“From July 1st 2007, the cross shipping of fit USD 10 and USD 20 bills - that is the depositing at, and withdrawing from, the Federal Reserve fit USD 10 and USD 20 bills within the same week will be subject to a cross shipping penalty charge of USD 5 per bundle of 1,000 notes.” (De La Rue plc. 2008)

Consider a general linear recirculation fee $Ky$, where $y$ is the cross-shipped amount and the constant $K > 0$ is the per-bundle charge. Under such a fee, the objective of a DI $i$ is

$$
\min_{\psi_i} K \min \{ g(1 - \psi_i) x_i, (\rho_i - g \psi_i) x_i \} + \alpha \psi_i x_i + \beta \left[ (1 - g \psi_i) x_i + |\rho_i x_i - g \psi_i x_i| \right] + h \min \{ \rho_i x_i, g \psi_i x_i \}
$$

We analyze this objective for the two types of DIs. It is easy to see that $\psi_i^* \in \{0, \frac{\rho_i}{g}, 1\}$. We conclude that such a fee is unable to induce socially optimal behavior from DIs. Intuitively, the Fed’s mechanism is unable to achieve coordination as this fee does not capture the non-linear structure of indirect cost. We formally record this observation.

**Corollary 2.** A linear recirculation fee $Ky$, where $y$ is the number of cross-shipped bundles and the constant $K > 0$ is the per-bundle charge, is not a coordinating mechanism.

While it is clear that the Fed’s mechanism does not induce socially optimal behavior, one may wonder if the Fed’s mechanism closely approximates the social optimum. In such a case, the Fed may not be able to do much better by adopting any other mechanism. Next, we compare the performance of the Fed’s mechanism with the social optimum.

### 4.3. Performance of the Fed’s Mechanism

We now report on a numerical analysis conducted to evaluate the performance of the Fed’s mechanism. The results indicate that the gap between the optimum social cost and the social cost under
the Fed’s mechanism can be substantial. We also identify the cases where the Fed’s mechanism can perform well.

**The Test Bed:**

To anchor our experiments in reality, we generated our instances based on information from two sources: (i) our discussions with Brink’s Inc. and (ii) The Fed’s publicly available documents (Federal Reserve 2003, 2006a, 2006b). The Fed expects the recirculation policy to result in savings of approximately $250 million over the next ten years in its currency processing costs (Federal Reserve 2006a). In 2003, $387 million (or about 15%) of the Fed’s total budget of $2.63 billion was dedicated to currency management operations (Blacketer and Evetts 2004). Also, the Fed’s current cross-shipping fee is $5 per bundle of cross-shipped currency (Federal Reserve 2003). Using these guidelines, we estimate the weekly demand (in number of bundles) of fit cash of a DI to be between 2,500 and 7,500 bundles. We assume 50 DIs in a Federal zone. For experimental purposes, we fix weekly demand ($y_i$) at 5,000 bundles, for all $i$. For a DI, let $\rho_i = \frac{y_i}{x_i}$ be ratio of the demand to deposits. Our discussions with Brink’s Inc. suggest an important characteristic of the ratio $\rho_i$: there is little week-to-week variation in the value of $\rho_i$ for a specific DI. However, across DIs, the value of $\rho_i$ varies depending on several factors, including the size of a DI, the dominant locations of its branches (e.g., metro or rural area), the primary focus of its business (e.g., personal or business banking), etc. Thus, to represent a variety of DIs, we consider 50 values (to represent 50 different DIs) of $\rho_i$ randomly from $U[0.25, 1]$. Using this, for a DI $i$, the used cash deposit is set at $x_i = \frac{y_i}{\rho_i}$.

The other cost parameters are fixed to meet the goal of realism. To estimate the holding cost $h$, we considered the following facts: one reason that DIs engage in cross-shipping is to avoid incurring opportunity costs of holding currency in their vaults. A DI can earn a positive return on its cash, credited to its account with the Fed, by lending it to some other institution via the Federal Funds Market (Federal Reserve 2006a). Over the past 17 years (1990 to 2007), the intended annual Federal Funds Rate (for transactions in the Federal Funds Market) has varied between a minimum of 1.25% and a maximum of 8% (Federal Reserve 2008b). The median rate was 4.75% and the average rate was 4.54%. Therefore, assuming an annual rate of return of around 5% and notes of denomination $20 (a dominant component of cross-shipped currency), the holding cost per bundle per week ($h$) was set to $20. After comprehensive testing, we present the parameter settings shown in Table 4.
To analyze the behavior of the Fed’s mechanism, we varied a DI’s per bundle fit-sorting cost $\alpha$ for eight levels in increments of one unit, starting from $1 per bundle. Data on the volume of used cash processed and new cash printed by the Fed in recent years indicates that the indirect costs resulting from blocked cash (Section 1.2) are quite significant. Accordingly, the parameters for the indirect cost function $\Gamma$ (Section 3.1) were chosen to illustrate that the magnitude of mis-coordination can be substantial. We chose 4 possible values for the indirect cost coefficient $\gamma$: 0.0005, 0.001, 0.0015, and 0.002.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI’s Transportation Cost ($\beta$)</td>
<td>5</td>
</tr>
<tr>
<td>DI’s Holding Cost ($h$)</td>
<td>20</td>
</tr>
<tr>
<td>Fed’s Fit-Sorting Cost ($\lambda$)</td>
<td>2</td>
</tr>
<tr>
<td>Indirect Cost Exponent ($p$)</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4 Values for Parameters

Results:
For a DI $i$, we first evaluate the optimal fit-sorting fraction, say $\psi_{iF}^*$, under Fed’s mechanism. Next, for each DI, we evaluate the optimal decision, say $\phi_{iD}^*$, decided by social planner.

For a DI $i$, let $SC(\psi_{iF}^*)$ denote the cost, evaluated corresponding to the fit-sorting fraction $\psi_{iF}^*$. Let $SC(\phi_{iD}^*)$ be the cost corresponding to the fit-sorting fraction $\phi_{iD}^*$. Over the 50 DIs, we compute the optimal societal cost $\sum_i SC(\phi_{iD}^*)$ and compare it with the societal cost $\sum_i SC(\psi_{iF}^*)$ incurred using $\psi_{iF}^*$ as follows:

$$\% \text{ Gap} = \frac{\sum_i SC(\psi_{iF}^*) - \sum_i SC(\phi_{iD}^*)}{\sum_i SC(\phi_{iD}^*)} \times 100$$

Figure 5 illustrates the percentage gap of social cost under the Fed’s Mechanism from the optimal social cost. It is clear that the social cost under the Fed’s mechanism is significantly higher than the social optimum for most of the instances. For instance, when $\alpha = 5$ and $\gamma = 0.002$, the social cost under the Fed’s mechanism is approximately 91% higher than the social optimum. Our results indicate a reasonable amount of mis-coordination even for smaller values of $\gamma$. For example, for $\gamma = 0.0005$ and $\alpha = 5$, the social cost under Fed’s mechanism is about 8.01% higher than the social optimum.

At this point, a natural question that arises is whether a non-linear recirculation fee can provide
a coordinating mechanism. The impact of cross-shipping includes the increased fit-sorting cost for the Fed as well the indirect system cost. The nonlinear nature of this “damage”, as discussed in Sections 1 and 3, suggests a strictly convex and increasing form for the recirculation fee $e(y_i)$, where $y_i$ is the cross-shipped amount for DI $i$. However, it can be easily shown (for brevity, we avoid providing a detailed proof) that such a nonlinear fee will also be unable to induce socially optimum behavior from a DI. Thus, we conclude that the implementation of a cross-shipping fee (linear or non-linear) does not result in a coordinating mechanism.

The basic reason behind the inability of the Fed’s mechanism to achieve coordination is that the Fed’s recirculation fee targets fit-sorting expense indirectly by focusing on the tradeoff associated with reducing the cross-shipped amount. This issue, therefore, motivates the need for a coordinating mechanism that can directly concentrate on the fundamentally different tradeoff associated with reducing deposits of used cash to the Fed – the main source of the negative externality imposed on society. We develop such a mechanism in the next section.

5. A Coordinating Mechanism Using Adjusted Fit-Sorting Fees

Our new fee structure is a cross-shipping adjusted fit-sorting fee (hereafter referred to as an adjusted fit-sorting fee) and is based on the amount of (non-fit-sorted) used cash a DI deposits to the Fed.
We start by examining the intuition behind such a fee and then formally define the mechanism. The proof of the coordination is presented in Section 5.1, followed by two important properties of the mechanism in Section 5.2. Finally, we briefly discuss the impact of the custodial inventory program in Section 5.3.

Recall from Section 1.2 that the primary consequence of a reduction in the practice of cross-shipping is the reduced volume of fit-sorting by the Fed (and, therefore, a corresponding reduction in the Fed’s fit-sorting expenses). In other words, the Fed aims to target the “events” of cross-shipping with the intention of lessening its fit-sorting expense (and, consequently, the societal cost of providing cash to the public). This suggests that instead of such an indirect attack on fit-sorting expense, one can consider a fit-sorting fee that directly targets the Fed’s volume of fit-sorting. This volume, in turn, is defined by the (non-fit-sorted) used cash deposited to the Fed by the DIs. Thus, a possible approach is to charge a DI based on the amount of used cash it deposits to the Fed. Note, however, that levying a fee on a DI based only on the amount of used cash it deposits ignores the capability of that DI to recirculate, and is thus likely to be viewed as being unfair (this issue is discussed in detail in Section 5.2). We, therefore, define the adjusted fit-sorting fee in the following manner: A DI will be charged for deposits of non-fit-sorted cash to the Fed only if it also makes positive withdrawals from the Fed. Before examining the implications of this definition, we state our mechanism precisely.

The Definition of the Mechanism: Let \( K_i(x_i, \rho_i) = \lambda x_i (1 - \frac{\rho_i}{g})^+ + \gamma (x_i (1 - \frac{\rho_i}{g})^+)^p \). Suppose a DI fit-sorts a \( \psi_i \)-fraction of its deposits (i.e., \( x_i \)) and sends the remaining \( (1 - \psi_i) \)-fraction to the Fed. Then, the non-fit-sorted used-cash bundles deposited to the Fed by DI \( i \) is \( w_i = (1 - \psi_i)x_i \). The adjusted fit-sorting fee charged to the DI is \( \Omega(w_i) = -K_i(x_i, \rho_i) + \lambda w_i + \gamma w_i^p \). Note that \( \Omega(0) = -\lambda x_i (1 - \frac{\rho_i}{g})^+ - \gamma (x_i (1 - \frac{\rho_i}{g})^+)^p \) and \( \Omega(x_i (1 - \frac{\rho_i}{g})^+)^p = 0 \) for all \( i \in I \).

We first show that this is a coordinating mechanism and then examine its auxiliary properties.

5.1. Proof of Coordination

Under the mechanism defined above, the objective of a DI is:

\[
\min_{\psi_i} \left( \Omega((1-\psi_i)x_i) + \alpha \psi_i x_i + \beta \{(1-g\psi_i)x_i + |\rho_i x_i - g\psi_i x_i|\} + \mu \min \{\rho_i, x_i, g\psi_i x_i\} \right)
\]
Note that each Type 1 DI will always make a non-negative withdrawal from the Fed as it can’t generate enough fit cash (by fit-sorting) to meet its demand. On the other hand, a Type 2 DI, say i, is capable of generating more fit cash than it needs by fit sorting more than a $\frac{\phi_i}{g}$-fraction of its deposits. Our mechanism incentivizes such an action by providing a reward that is proportional to the volume of the extra fit cash generated. Accordingly, we consider the two ranges of $\psi_i$ separately:

(a) $\psi_i \leq \frac{\phi_i}{g}$ and (b) $\psi_i \geq \frac{\phi_i}{g}$. Note that the objective function for a DI is identical to that of the social planner under both the cases, except for the first term $(-K_i(x_i, \rho_i))$ which is independent of $\psi_i$. Thus, the relative comparison of the objectives under the two cases (and, hence, the optimum) is the same as that for the social planner’s problem in Section 3.4 (Propositions 1 and 2).

We formally state this result as follows:

**Corollary 3.** The adjusted fit-sorting fee $\Omega$ defined as above induces the DIs to self-select the social optimum.

To implement the fee structure above, it is reasonable to expect that, over time, the Fed would be able to arrive at fairly accurate estimates of the indirect cost parameters $\gamma$ and $p$. However, it is not necessary to have a precise knowledge of these parameters. Section EC.2 describes a sensitivity analysis to study the impact of imprecision and shows that our mechanism is robust under significant uncertainty in these parameters.

Note that the adjusted fit-sorting fee $\Omega$ is DI-specific: the computation of $K_i$ for a DI requires knowledge of its weekly used-cash deposits $x_i$ and its demand-to-deposits ratio $\rho_i$. The Fed can use a simple consistency check for these two values based on a DI’s transactions. A DI performs three transactions with the Fed: (i) a $(1 - \phi_i)x_i$ deposit of used cash, (ii) a $(1 - g)\phi_i x_i$ deposit of unfit cash, and (iii) either a $(\rho_i x_i - g\phi_i x_i)$ demand of fit cash or a $(g\phi_i x_i - \rho_i x_i)$ deposit of fit cash. The amount of each of these three transactions results in a system of three equalities which has a unique solution for $x_i$ and $\rho_i$. Thus, the Fed can quickly confirm if a DI’s reported values of $x_i$ and $\rho_i$ are consistent with its transactions.

A larger question, however, is whether the values of $x_i$ and $\rho_i$ can be misreported by a DI. Such a possibility can be safely ignored because deposits and demand are transparent to all the stakeholders of these transactions, including the Fed. Furthermore, records of these transactions are carefully maintained for several years after they have been conducted. Consequently, a DI cannot
misreport these numbers for they can be audited by the Fed at any time. Given that collusion is unrealistic in this domain, another possibility is for the Fed to legally allow a group of DIs to reallocate (amongst themselves) their deposits and demand if such a reallocation can benefit these DIs. We will investigate this concept of redistribution in the following subsection.

Before proceeding further, note that the model we analyze is deterministic: we assume that the fit cash fraction (i.e., $g$) in used-cash deposits is fixed, the ratio $\rho_i$ of weekly demand to weekly deposit for a DI $i$ is a constant and that the DI knows its weekly deposits $x_i$ of used cash. A natural question arises: how critical is to consider uncertainty in the model? In Section EC.3, we report on a comprehensive numerical study conducted to analyze the impact of uncertainty on the optimal decisions. Here, we briefly describe the experiment and the results. Let $\bar{x}$ (resp., $\bar{y}$) and $\bar{g}$ denote the mean weekly deposit (resp., demand), for a DI, and mean fit-cash fraction, respectively. To account for uncertainty, we generated symmetric intervals, centered around the respective means. For each triplet $(\bar{x}, \bar{y}, \bar{g})$, we considered 5%, 10%, 15%, 20%, and 25% variation around the mean. The parameters were varied simultaneously in every instance. For example, a 5% variation corresponds to the interval $[0.975\bar{x}, 1.025\bar{x}]$ for $\bar{x}$, interval $[0.975\bar{y}, 1.025\bar{y}]$ for $\bar{y}$ and interval $[0.975\bar{g}, 1.025\bar{g}]$ for $\bar{g}$. Our results show that, while uncertainty does have an impact, the magnitude of this impact is small even under a considerable amount of uncertainty. Therefore, using the mean values of the parameters (weekly deposit, demand, and $g$), provides a very good approximation for the societal cost. For instance, even under a considerable amount (25%) of uncertainty, the percentage gap between the optimal cost and that obtained using mean values of these parameters is negligible (ranging between 0.0005% to 0.5%). We also study the performance of our coordination mechanism under uncertainty and show that it achieves the desired properties.

5.2. Coordination Mechanism: Fairness and Redistribution
The primary aim of our coordinating mechanism is to ensure an effective recirculation of cash by DIs. However, a DI’s capability to recirculate cash depends on the comparison between the deposits (of used cash) it receives and the demand (of fit cash) it faces. Clearly, the extent of a DI’s recirculation should be evaluated relative to its ability. Thus, the issue of fairness of the mechanism becomes important. Our mechanism avoids penalizing DIs that recirculate their fair share of cash and rewards DIs that recirculate more than this amount. We elaborate these properties below:
(i) The DIs that do not make any used cash deposits to the Fed are not charged any fee. For instance, consider a Type 1 DI that fit-sorts all of its used-cash deposits. Clearly, such a DI will not make any used-cash deposits to the Fed (and hence will not indulge in cross-shipping), but may be required to withdraw additional fit cash from the Fed. No fee is charged to such a DI (i.e., \( \Omega(0) = 0 \) for a Type 1 DI).

(ii) The DIs that do not withdraw any fit cash from the Fed are not charged any fee. For example, consider a Type 2 DI that fit-sorts a \( \rho \) -fraction of its used-cash deposits that is sufficient to meet its demand of fit cash. Such a DI deposits the remaining used cash to the Fed, but does not withdraw any fit cash from the Fed (and hence will not indulge in cross-shipping). Again, no fee is charged to such a DI (i.e., \( \Omega((1 - \rho) x) = 0 \) for a Type 2 DI).

(iii) Type 2 DIs can fit-sort more than their individual requirement (of fit cash) and deposit the additional fit-sorted amount to the Fed. Our mechanism appropriately rewards such DIs for the additional fit cash generated. Thus, if a Type 2 DI fit-sorts more than a \( \rho \) -fraction of its deposits, its used-cash deposit (to the Fed) is \( w < (1 - \rho) x \). In this case, \( \lambda w + \gamma w^p < K(x, \rho) \), which implies \( \Omega(w) < 0 \) (i.e., a reward).

The term \(-K_i(x_i, \rho_i)\) of the adjusted fit-sorting fee \( \Omega(w_i) \) helps ensure the above properties; we, therefore, refer to it as the fairness adjustment. Figure 6 shows the variation in the adjusted fit-sorting fee for a DI with respect to its used-cash deposit \( x_i \) and demand-deposit ratio \( \rho_i \) for the optimum fraction \( \psi^*_i \). Note that a positive value for the adjusted fit-sorting fee represents a penalty whereas a negative value represents a reward. Figure 7 represents the variation in the fairness adjustment \(-K_i(x_i, \rho_i)\) for a DI with respect to \( x_i \) and \( \rho_i \). This adjustment takes a value of zero for a DI with \( \rho_i \geq g \); this can be seen from the lines parallel to the x-axis in the figure. Below these lines, the absolute value of the fairness adjustment increases with an increase in \( x_i \) and a decrease in \( \rho_i \). Note that for \( \rho_i < g \), the rate of decrease in the fee (and also the rate of increase in the absolute value of the fairness adjustment) increases with an increase in \( x_i \).

Note that in our mechanism, cross-shipping remains a central criterion for fairness, in the sense that a DI that does not indulge in cross-shipping should not be charged any fee. Furthermore, if the social planner benefits from having a DI fit-sort more than the DI’s need for fit cash, then the mechanism should provide an appropriate reward to incentivize such DIs. Thus, our mechanism
incorporates two key features: (i) achieving coordination by levying a fee on used-cash deposits to the Fed and (ii) achieving fairness through appropriate choices of the fairness adjustments \( K_i(x_i, \rho_i), i \in I \). Without the fairness adjustment, a fit-sorting fee alone is capable of achieving the first feature but not the second.

We now return to the notion of redistribution mentioned earlier. Consider a subset of DIs indexed by \( J \subseteq I \). We envision the possibility of the Fed allowing these DI to reallocate, amongst themselves, their deposits and demand to benefit from our mechanism. Let \( x_j \) (resp., \( \rho_j \)) and \( x'_j \) (resp., \( \rho'_j \)) denote, respectively, the deposit (demand-to-deposit ratio) before and after such a redistribution. To preserve the total deposits and demand over the DIs in \( J \), two obvious constraints on the redistribution\(^5\) are:

\[
\sum_{j \in J} x'_j = \sum_{j \in J} x_j; \quad \sum_{j \in J} \rho'_j x'_j = \sum_{j \in J} \rho_j x_j.
\]

After redistribution, a DI \( j \in J \) continues to communicate with the Fed as an individual entity with its “virtual” parameters \( x'_j, \rho'_j \). In other words, redistribution is an internal reallocation between the DIs in \( J \) and does not affect the structure of their communication with the Fed. We first observe that the total adjusted fit-sorting fee incurred by the DIs in \( J \) can be smaller after the redistribution as compared to that before the redistribution. In such a case, however, the societal cost may not reduce. To illustrate, consider two DIs \( i = 1, 2 \). Let the triplets \((x_i, \rho_i, \phi'_i)\)

\(^5\) Our purpose in discussing redistribution is to explore the potential of an interesting concept. It may be the case that the constraints of their respective business environments may force the DIs to avoid such a possibility.
and \( (x_i', \rho_i', \phi_i^*) \); \( i = 1, 2 \), denote, respectively, their deposit, demand-to-deposit ratio, and optimum fit-sorting fraction, before and after the redistribution. Table 5 illustrates two cases: in Case 1, the total cost of the two DIs as well as the societal cost reduce following the redistribution; in Case 2, the total cost of the two DIs reduces but the societal cost does not.

### Table 5: Scenarios Before and After Redistribution

<table>
<thead>
<tr>
<th>Case</th>
<th>Before Redistribution</th>
<th>After Redistribution</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( (x_1, \rho_1, \phi_1^* = 1 - \frac{\epsilon}{x_1}, (x_2, \rho_2, \phi_2^* = 1 - \frac{\rho_1 x_1 + \rho_2 x_2}{x_2}) )</td>
<td>( (x_1' = \epsilon, \rho_1' = \frac{\phi_1^<em>}{\epsilon}, \phi_1^</em> = 0) ) ( (x_2' = x_1 + x_2 - \epsilon, \rho_2' = 0, \phi_2^* = 1 - \frac{\phi_1^<em>}{\phi_2^</em>}) )</td>
<td>Social Planner (Fed) and DIs both benefit; Figures 8(a) and 8(b).</td>
</tr>
<tr>
<td>2</td>
<td>( (x_1, \rho_1, \phi_1^* = 0) ) ( (x_2, \rho_2, \phi_2^* = 0) )</td>
<td>( (x_1' = \epsilon, \rho_1' = \frac{\phi_1^<em>}{\epsilon}, \phi_1^</em> = 0) ) ( (x_2' = x_1 + x_2 - \epsilon, \rho_2' = 0, \phi_2^* = 0) )</td>
<td>DIs benefit but not the Social Planner (Fed); Figures 9(a) and 9(b).</td>
</tr>
</tbody>
</table>

Clearly, an answer to the question of whether or not redistribution can benefit the DIs and/or the society depends on the parameters of the DIs before the redistribution. For the two cases in Table 5, it is straightforward to compare the objective functions before and after redistribution. Corresponding to Case 1 in Table 5, Figures 8(a) and 8(b) depict, respectively, the variations in the cost of the DIs and the societal cost after redistribution with respect to \( x_1' \) and \( \rho_1' \) (Note that by fixing \( x_2' \) and \( \rho_2' \), we fix \( x_1' \) and \( \rho_1' \) as well since \( x_1' + x_2' = x_1 + x_2 \) and \( \rho_1' x_1' + \rho_2' x_2' = \rho_1 x_1 + \rho_2 x_2 \)). Observe that there indeed exist redistribution solutions that benefit the DIs as well as the society. Figures 9(a) and 9(b) demonstrate points of redistribution (corresponding to Case 2 in Table 5)
Figure 9 Redistribution Improves DIs’ Cost but May Worsen Societal Cost: (a) DIs’ Cost and (b) Societal Cost for Redistributed Demand-Deposits Ratio $\rho_2$ and Deposit $x'_2$

where the DIs benefit but societal cost worsens. In this redistribution, since $\phi_i^* = 0$ irrespective of the values of $\rho_i$ for $i = 1, 2$, the societal cost does not vary with $\rho_2$ in Figure 9(b). A natural question that follows is whether the social planner (i.e., the Fed) can prevent redistribution that increases societal cost. We refer to such a redistribution as welfare-preserving and argue that it is straightforward to incorporate in our mechanism. As before, suppose a subset of DIs indexed by $J \subseteq I$ expresses an interest (to the Fed) in redistribution. As mentioned earlier, the Fed is aware of the true (i.e., before the redistribution) parameters $x_j, \rho_j; j \in J$ and, therefore, the true total societal cost $T_J$ imposed by the DIs in $J$. Let $x'_j, \rho'_j; j \in J$ denote the redistributed parameters.

After redistribution, let $w'_j$ be the used cash deposited by DI $j, j \in J$. Then, the redistribution is welfare-preserving if the total contribution of the DIs in $J$ to the societal cost is at most $T_J$. That is,

$$\sum_{j \in J} f(w'_j) \leq T_J,$$

where $f(w'_j) = \lambda w'_j + \gamma w'_j^p + \alpha (x'_j - w'_j) + \beta \{x'_j - g(x'_j - w'_j)\} + h \min\{\rho'_j x'_j, g(x'_j - w'_j)\}$. To strictly enforce welfare-preserving redistribution on the DIs in $J$, the Fed can enforce the following “customized” criterion.

$$\Omega(w'_j) = \begin{cases} -K_j + \lambda w'_j + \gamma w'_j^p & \text{if } \sum_{j \in J} f(w'_j) \leq T_J \\ \infty & \text{otherwise.} \end{cases}$$

Thus, the Fed allows redistribution, but only if it is welfare-preserving.

5.3. Impact of the Custodial Inventory Program

Recall from Section 1.2 the concept of custodial inventory. The Fed introduced its pilot custodial inventory program in May, 2006, with the intention of reducing the cost of holding currency for
DIs and incentivizing them to recirculate cash:

“The Federal Reserve Banks have created a Custodial Inventory program that will offset the opportunity costs associated with holding additional currency in vaults long enough to facilitate its recirculation.” (Federal Reserve 2006b)

One reason for DIs to avoid fit-sorting and hasten their deposits of non-fit-sorted used cash to the Fed is that deposits with the Fed immediately become earning assets. Naturally, DIs would like to avoid holding “idle” cash in their own vaults and incur the corresponding opportunity costs. The custodial inventory program is aimed at resolving this problem. Custodial inventory is cash owned by the Fed but is located in a DIs facility and segregated from the DIs own vault. The idea is to enable DIs to deposit cash to custodial inventory (i.e., to the Fed) as soon as possible and withdraw it as and when required. However, as a necessary condition, a DI must fit-sort the cash before depositing it into custodial inventory. Thus, custodial inventory allows DIs to reduce (and theoretically eliminate) cross-shipping while mitigating the opportunity cost \( (h) \) they would incur by holding cash long enough to recirculate it (Federal Reserve 2006a).

Thus, a DI can avoid incurring opportunity costs \( (h) \) by using custodial inventory. While the holding cost of custodial inventory, say \( r \), is significantly less than \( h \), it is non-trivial due to the following reason. Custodial inventory needs to be maintained under the Fed’s regulations. These regulations include the requirement of additional qualified personnel, additional counting activities, regular supervision and surveillance, and maintaining the custodial inventory vault and related records on a daily basis (to keep them ready for an audit by the Fed at any time) (Federal Reserve 2008a). Thus, the holding cost, \( r \), of custodial inventory is a service fee charged by a third-party logistics provider for holding cash in custodial inventory at its processing center. More precisely, consider a DI, say \( i \), that fit sorts a \( \phi_i \)-fraction of its used cash and generates \( \min\{\rho_i x_i, g\phi_i x_i\} \) fit cash (see Section 3.3). Then, instead of maintaining fit-cash inventory in its own vaults, the DI will prefer to transfer it to custodial inventory. Note that custodial inventory reduces the cost of the DIs as well as societal cost and, therefore, induces a systemic change.

Our analysis in Sections 3.4, 4, and 5, remains largely unchanged in the presence of custodial inventory. The optimum fraction \( \phi^* \) for the social planner (resp., \( \psi^* \) for DIs) will have the same form as in Section 3.4 (resp., Section 5.1) with the only change being the replacement of the holding
cost $h$ by $r$ in the expression of $c^*$ (Section 3.4); for example, the new value is $c^* = \left(\frac{\alpha - 2\beta g + rg - \lambda}{\phi} \right)^{\frac{1}{r}}$
when $\alpha + rg > 2\beta g + \lambda$. Since $r < h$, the value of $c^*$ reduces under custodial inventory; consequently, $\phi^*$ and $\psi^*$ increase. Thus, the presence of custodial inventory incentivizes DIs to fit sort more than they would have under the adjusted fit-sorting fee alone.

To summarize, the sole impact of custodial inventory is to reduce the cost of holding fit-sorted currency at a DI. As shown in Section 4, the Fed’s mechanism is not guaranteed to coordinate for any non-zero value of the holding cost. Therefore, custodial inventory does not have the capability to convert a non-coordinating mechanism into a coordinating one. It does not eliminate the cause of mis-coordination. However, it mitigates the impact of mis-coordination by reducing both the societal cost as well as those of the DIs.

6. Discussion and Conclusions

The main focus of the Fed’s new cash recirculation policy is to encourage private-sector behavioral changes that would lower the overall societal costs of cash processing and distribution. The primary components of this policy are (i) a recirculation fee on cross-shipped cash and (ii) the custodial inventory program. The Fed intends to use the recirculation fee as a mechanism for inducing DIs to behave in a manner so that the societal cost of providing currency to the public (which is the Fed’s eventual goal) is minimized.

We argue that the mechanism of imposing a recirculation fee on cross-shipped currency will, in general, not induce DIs to behave in a socially optimal manner. Under the new cash-recirculation policy that came into effect from July 2007, the Fed has chosen to impose a constant, per-bundle cross-shipping fee. We show that neither such a linear scheme nor a more sophisticated nonlinear scheme is sufficient to guarantee a socially optimal behavior from DIs. We then derive a fundamentally new coordinating mechanism in which a fee is charged on the used cash deposited to the Fed but is adjusted by the presence of cross-shipping. Specifically, our fairness adjustment (i) avoids penalizing a DI that fit-sorts sufficient cash to meets its demand of fit cash and (ii) rewards a DI for fit-sorting more than its demand of fit cash. We also discuss a concept of welfare-preserving redistribution wherein the Fed allows a group of DIs to reallocate (amongst themselves) their deposits and demand if such a possibility does not increase societal cost. Finally, we analyze the impact of incorporating the custodial inventory program.
6.1. Implications for Implementation

We now discuss a few implementation-related advantages of our mechanism.

- **Flexibility of Individual or Group Transactions**: Our coordination mechanism is DI-specific. That is, the adjusted fit-sorting fee for a DI is a function of two characteristics of that DI: the volume of its (weekly) used-cash deposits and its demand-to-deposits ratio. As mentioned earlier, the transparency of financial transactions to the Fed implies that it has the knowledge of these two parameters for each DI. The DIs, therefore, will not misreport these two values. Furthermore, via the concept of redistribution (Section 5.2), the Fed can allow a group of DIs to reallocate (amongst themselves) their deposits and demand if such a possibility improves social welfare. Thus, a DI can choose to transact with the Fed either individually or as part of a group.

- **Participation Fee, Fairness, and Stability**: If the Fed wants to operate as a zero deficit/income center, then it can impose a participation fee (on each DI) that depends on the mean (over all DIs in the population) deposit of used-cash and mean demand of fit-cash faced by a DI. Once computed, such a participation fee need not change until there is a change in these parameters of the population. Thus, the mechanism is stable and imposes little administrative burden on the Fed. An additional advantage of the mechanism is its fairness (Section 5.2). While there may be other mechanisms that guarantee coordination, to facilitate enthusiastic participation by DIs it is important that they perceive the adjusted fit-sorting fee as fair.

- **A Redistribution Frontier**: Consider a set, say $S$, of DIs interested in exploiting redistribution opportunities. Once the DIs in $S$ identify themselves, the Fed can provide the information required by the DIs to formulate the non-linear optimization problem that computes a welfare-preserving redistribution. However, if required, redistribution can also be effectively implemented by the Fed. Since there may be several distinct welfare-preserving redistribution solutions, the Fed may not be able to assess one that is best suited for the DIs. Instead, the DIs (in $S$) themselves are in an ideal position to estimate the local effort required to effectively implement redistribution. As such, instead of recommending a single redistribution, the Fed may offer the DIs in $S$ a collection of such solutions. Conceptually, this collection can be viewed as part of the frontier of bi-criteria solutions, with the attractiveness of a solution to the Fed and the DI being the two criteria. The DIs can then examine this collection and choose a redistribution that suits them the best.
6.2. Directions for Future Work

Ours is the first attempt to analyze the Fed’s cash recirculation policy as a coordination mechanism. While our analysis incorporates the fundamental features of the Fed’s policy, it may be beneficial to examine any additional details that impact daily operations at the DIs. Consider, for instance, the Fed’s custodial inventory program. While the objective of this program is to offset the opportunity costs (incurred by the DIs) associated with holding additional currency in vaults long enough to facilitate its recirculation, the Fed has imposed a few requirements with an eye on practical implementation. For example, a DI is required to hold in its own vault at least an amount equalling its average daily requirement of fit-cash before transferring any additional fit-cash to custodial inventory. Furthermore, the amount of fit cash transferred to custodial inventory may not exceed four days of a DI’s average (fit-cash) requirements. While such requirements do not influence the fundamental impact of custodial inventories, it might be interesting to analyze their effect on the day-to-day schedule of deposits and withdrawals from custodial inventory.

For our purpose in this paper, a deterministic model with mean values of the parameters (demand, deposits and fit-cash fraction) was well-suited. Nevertheless, a stochastic analysis where these values come from a known distribution would be both relevant and challenging. It may also be useful to further investigate some of the novel ideas of our analysis. For example, one possibility for implementing welfare-preserving redistribution (Section 5.2) is for the Fed to actively suggest attractive opportunities to groups of DIs that are interested in exploiting redistribution. In such a case, the effective computation of a sufficiently diverse set of welfare-preserving redistribution solutions is an interesting and challenging issue that the Fed may need to address.

References


http://www.delarue.com/Display.aspx?MasterId=35ac2a3d-b638-4e71-a1c7-ee9beaa3eb8a&NavigationId=750.


Federal Reserve. 2008a. Custodial Inventory Program.


This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.
Proofs of Propositions and Sensitivity Analysis

EC.1. Proofs of Propositions 1 and 2.

Before proceeding to the proofs of Propositions 1 and 2, we discuss three intermediate results: Lemmas EC.1-EC.3. We start with discussing the two cases mentioned in Section 3.4.

• Case 1 \((\phi_i \leq \rho_i g)\): This case is relevant to both type of DIs. Note that for a Type 1 DI, \(\rho_i g > 1\). Hence, we optimize \(\Upsilon_i(\phi_i)\) for \(0 \leq \phi_i \leq \min\{1, \frac{\rho_i g}{\phi_i}\}\). We have \(\Upsilon_i(\phi_i) = \lambda(1 - \phi_i)x_i + \gamma((1 - \phi_i)x_i)^p + \alpha\phi_i x_i + \beta\{1 - g\phi_i\}x_i + (\rho_i x_i - g\phi_i x_i) + hg\phi_i x_i\). Note that \(\Upsilon_i(\phi_i)\) is differentiable in \(\phi_i\). Let \(\phi^*_i\) denote the optimal fit-sorting fraction and let \(\Upsilon_i(\phi^*_i)\) denote the corresponding cost.

In Lemma EC.1, we first identify a condition under which the social optimum is trivial to obtain.

**Lemma EC.1.** If \((\alpha + \rho g \leq 2\beta g + \lambda)\), then under the constraint \(\phi_i \leq \frac{\rho_i g}{\phi_i}\), the social planner’s optimum fit-sorting fraction for each DI \(i \in I\) is \(\phi^*_i = \min\{1, \frac{\rho_i g}{\phi_i}\}\).

Before moving further, it is instructive to interpret the condition in the statement of the lemma.

**Remark EC.1.** If the social planner wants a DI to fit-sort a bundle of cash, the DI incurs the following costs: (a) a fit-sorting cost of \(\alpha\), (b) a holding cost of \(hg\) for holding a \(g\)-fraction of fit-sorted cash that the DI retains, and (c) a transportation cost of \(\beta(1 - g)\) for sending the unfit cash to the Fed. Thus, the total incurred cost by the system is \(\alpha + \beta(1 - g) + hg\). Note that there is no indirect system cost in this case. The social planner could have avoided these costs by allowing the DI to (i) deposit the bundle to the Fed and incur a transportation cost of \(\beta\) units, (ii) withdraw \(g\) units of fit-cash from the Fed and incur an additional transportation cost of \(\beta g\) units. Since, in this case, the Fed fit-sorts the bundle, the total system cost is \(\beta(1 + g) + \lambda + \gamma\). Thus, if \((\alpha + \beta(1 - g) + hg) \leq (\beta(1 + g) + \lambda)\), the social planner will always want the DI to fit-sort the bundle. This simplification results in a trivial solution to the social planner’s problem.

**Proof:** The social planner’s objective under the constraint \(\phi_i \leq \frac{\rho_i g}{\phi_i}\) is \(\min \Upsilon_i(\phi_i)\), where

\[
\Upsilon_i(\phi_i) = [\lambda(1 - \phi_i)x_i + \gamma((1 - \phi_i)x_i)^p + \alpha\phi_i x_i + \beta\{1 - g\phi_i\}x_i + (\rho_i x_i - g\phi_i x_i)] + hg\phi_i x_i
\]

Rewriting, we have

\[
\Upsilon_i(\phi_i) = \gamma((1 - \phi_i)x_i)^p + (\alpha - 2\beta g + hg - \lambda)\phi_i x_i + \lambda + \beta(1 + \rho_i)\]
Since $\alpha + hg \leq 2\beta g + \lambda$, the second term is non-positive. The first term decreases with an increase in $\phi_{i1}$ and third term is independent of $\phi_{i1}$. Thus, the optimal value of $\phi_{i1}$ is the upper bound $\min\{1, \frac{g}{g}\}$. Q.E.D.

In lemma EC.2, we provide the social planner’s solution when $\alpha + hg > 2\beta g + \lambda$.

**Lemma EC.2.** Let $(\alpha + hg > 2\beta g + \lambda)$. Then, under the constraint $\phi_i \leq \frac{g}{g}$, the optimum solution for the social planner’s problem is:

$$\phi_{i1}^* = \begin{cases} \min\{1 - \frac{c^*}{x_i}, \frac{g}{g}\} & \text{if } c^* < x_i \\ 0 & \text{otherwise.} \end{cases}$$

where $c^* = \left(\frac{\alpha - 2\beta g + hg - \lambda}{p^2}\right)^{\frac{1}{p-1}}$.

**Proof:** Note that $\Upsilon_1(\phi_{i1})$ is strictly convex in the domain $[0, \min\{1, \frac{g}{g}\})$ as

$$\frac{d^2\Upsilon_1(\phi_{i1})}{d(\phi_{i1})^2} = \gamma p(p-1)((1 - \phi_{i1})x_i)^{p-2}x_i^2 > 0$$

Therefore, the first order condition $\frac{d\Upsilon_1(\phi_{i1})}{d\phi_{i1}} = 0$ is necessary and sufficient for the unique minimum. The first order condition is:

$$-\lambda x_i - \gamma p((1 - \phi_{i1})x_i)^{p-1}x_i + (\alpha - 2\beta g + hg)x_i = 0 \quad (EC.1)$$

Solving (EC.1), we have:

$$\phi_{i1} = 1 - \left(\frac{\alpha - 2\beta g + hg - \lambda}{p^2}\right)^{\frac{1}{p-1}}$$

Let $c^* = \left(\frac{\alpha - 2\beta g + hg - \lambda}{p^2}\right)^{\frac{1}{p-1}}$. Then,

$$\phi_{i1} = 1 - \frac{c^*}{x_i} \quad (EC.2)$$

Now, whether or not $\phi_{i1}$ lies in the interval $[0, \min\{1, \frac{g}{g}\}]$ depends upon the relationship between $c^*$ and $x_i$. We specify the interior and the boundary solutions under the following two cases:

**Case a:** $c^* < x_i$

* **Interior Minimum:** Condition $\alpha + hg > 2\beta g + \lambda$ ensures that $c^* \geq 0$. If $1 - \frac{c^*}{x_i} < \min\{1, \frac{g}{g}\}$, then $\phi_{i1}^* = 1 - \frac{c^*}{x_i}$ is the unique interior minimum.
* Boundary Solution $\phi_{i1}^* = \frac{\alpha_i}{y}$: For Type 2 DI $i$, $\min\{1, \frac{\alpha_i}{y}\} = \frac{\alpha_i}{y}$. If $\frac{\alpha_i}{y} \leq 1 - \frac{c^*}{x_i} < 1$ for such a DI, then $\phi_{i1}^* = \frac{\alpha_i}{y}$ is the optimum fraction. To see this, it suffices to observe that $\Upsilon_i(\phi_{i1})$ is decreasing throughout the interval $[0, \frac{\alpha_i}{y}]$. Consider the first order derivative of $\Upsilon_i(\phi_{i1})$ at $\phi_{i1} = 0$,

$$
\frac{d\Upsilon_i(\phi_{i1})}{d\phi_{i1}} |_{\phi_{i1}=0} = -\gamma px_i^{p-1} x_i + (\alpha - 2 \beta g + h g - \lambda) x_i
$$

Since $c^* < x_i$, $\frac{d\Upsilon_i(\phi_{i1})}{d\phi_{i1}} |_{\phi_{i1}=0} < 0$. Thus, $\Upsilon_i(\phi_{i1})$ is decreasing at $\phi_{i1} = 0$. The strict convexity of $\Upsilon_i(\phi_{i1})$ and $1 - \frac{c^*}{x_i} \geq \frac{\alpha_i}{y}$ together imply the required result.

Case b : $c^* \geq x_i$

* Boundary Solution $\phi_{i1}^* = 0$: In this case, $\frac{d\Upsilon_i(\phi_{i1})}{d\phi_{i1}} |_{\phi_{i1}=1} > 0$. As before, the convexity of $\Upsilon_i(\phi_{i1})$ and $1 - \frac{c^*}{x_i} \leq 0$ imply that the optimum fraction is $\phi_{i1}^* = 0$.

Thus, the optimum fraction

$$
\phi_{i1} = \begin{cases} 
\min\{1 - \frac{c^*}{x_i}, \frac{\alpha_i}{y}\} & \text{if } c^* < x_i \\
0 & \text{otherwise.}
\end{cases}
$$

where $c^* = \frac{(\alpha - 2 \beta g + h g - \lambda)}{\rho g}$. Q.E.D.

- Case 2 ($\phi_i \geq \frac{\alpha_i}{y}$): Only Type 2 DIs are capable of fit-sorting more than a $\frac{\alpha_i}{y}$-fraction of their used-cash deposits. For Type 1 DIs, $\frac{\alpha_i}{y} \geq 1$ and hence, trivially, $\phi_i \leq \frac{\alpha_i}{y}$. Thus, our discussion in this case is relevant only for Type 2 DIs. We have $\Upsilon_2(\phi_{i2}) = \lambda (1 - \phi_{i2}) x_i + \gamma (1 - \phi_{i2}) x_i^p + \alpha \phi_{i2} x_i + \beta (1 - g \phi_{i2}) x_i + (g \phi_{i2} x_i - \rho_i x_i)$ + $h \rho_i x_i$. Simplifying, $\Upsilon_2(\phi_{i2}) = \lambda (1 - \phi_{i2}) x_i + \gamma (1 - \phi_{i2}) x_i^p + \alpha \phi_{i2} x_i + \beta (1 - \rho_i) x_i + h \rho_i x_i$. Note that $\frac{\alpha_i}{y} \leq \phi_{i2} \leq 1$ and $\Upsilon_2(\phi_{i2})$ is differentiable in $\phi_{i2}$. Let $\phi_{i2}^*$ denote the optimal fit-sorting fraction and let $\Upsilon_2(\phi_{i2}^*)$ denote the corresponding cost.

**Lemma EC.3.** Under the constraint $\phi_i \geq \frac{\alpha_i}{y}$, the optimum solution for the social planner’s problem is:

$$
\phi_{i2}^* = \begin{cases} 
1 & \text{if } \alpha \leq \lambda \\
\max\{1 - \frac{d^*}{x_i}, \frac{\alpha_i}{y}\} & \text{otherwise.}
\end{cases}
$$

where $d^* = \frac{(\alpha - \lambda)}{\rho g}$. Q.E.D.

**Proof:** The objective of the social planner under the constraint $\phi_i \geq \frac{\alpha_i}{y}$ is $\min_{\phi_{i2}} \Upsilon_2(\phi_{i2})$, where

$$
\Upsilon_2(\phi_{i2}) = \lambda (1 - \phi_{i2}) x_i + \gamma (1 - \phi_{i2}) x_i^p + \alpha \phi_{i2} x_i + \beta (1 - \rho_i) x_i + h \rho_i x_i
$$

```
• If $\alpha \leq \lambda$, then rewriting $\Upsilon_2(\phi_{i2})$, we have

$$
\Upsilon_2(\phi_{i2}) = [\gamma((1 - \phi_{i2})x_i)^p + (\alpha - \lambda)\phi_{i2}x_i + \{\lambda + \beta(1 - \rho_i) + h\rho_i\}x_i]
$$

Since $\alpha \leq \lambda$, the second term is non-positive. The first term decreases with an increase in $\phi_{i2}$ and third term is independent of $\phi_{i2}$. Thus, the optimal value of $\phi_{i2}$ is the upper bound 1.

• If $\alpha > \lambda$, then note that $\Upsilon_2(\phi_{i2})$ is strictly convex in the domain $[\frac{\rho_i}{g}, 1)$ as

$$
d^2\Upsilon_2(\phi_{i2}) \frac{d\phi_{i2}}{d\phi_{i2}} = \gamma p(p - 1)((1 - \phi_{i2})x_i)^{p-2}x_i^2 > 0.
$$

Therefore, the first-order condition $\frac{d\Upsilon_2(\phi_{i2})}{d\phi_{i2}} = 0$ is necessary and sufficient for the unique minimum.

We have

$$
d\Upsilon_2(\phi_{i2}) \frac{d\phi_{i2}}{d\phi_{i2}} = -\lambda x_i - \gamma p((1 - \phi_{i2})x_i)^{p-1}x_i + \alpha x_i = 0 \tag{EC.3}
$$

Solving (EC.3), we have:

$$
\phi_{i2} = 1 - \frac{(\alpha - \lambda)\frac{1}{p}x_i}{x_i}
$$

Let $d^* = (\frac{\alpha - \lambda}{p}x_i)^{\frac{1}{p-1}}$. Then,

$$
\phi_{i2} = 1 - \frac{d^*}{x_i} \tag{EC.4}
$$

The interior and the boundary solutions are as follows:

— **Interior Minimum**: Condition $\alpha > \lambda$ ensures that $d^* \geq 0$. If $1 - \frac{d^*}{x_i} \geq \frac{\rho_i}{g}$, then $\phi_{i2}^* = 1 - \frac{d^*}{x_i}$ is the unique interior minimum.

— **Boundary Solution** $\phi_{i2}^* = \frac{\rho_i}{g}$: If $1 - \frac{d^*}{x_i} < \frac{\rho_i}{g}$ for such a DI, then $\phi_{i2}^* = \frac{\rho_i}{g}$ is the optimum fraction. To see this, it suffices to observe that $\Upsilon_2(\phi_{i2})$ is increasing throughout the interval $[\frac{\rho_i}{g}, 1]$.

Consider the first order derivative of $\Upsilon_2(\phi_{i2})$ at $\phi_{i2} = \frac{\rho_i}{g}$,

$$
d\Upsilon_2(\phi_{i2}) \bigg|_{\phi_{i2} = \frac{\rho_i}{g}} = -\gamma p(x_i^{p-1}(1 - \frac{\rho_i}{g})x_i + (\alpha - \lambda)x_i)
$$

Since $1 - \frac{d^*}{x_i} < \frac{\rho_i}{g}$, $d\Upsilon_2(\phi_{i2}) \bigg|_{\phi_{i2} = \frac{\rho_i}{g}} > 0$. Thus, $\Upsilon_2(\phi_{i2})$ is increasing at $\phi_{i2} = \frac{\rho_i}{g}$. The strict convexity of $\Upsilon_2(\phi_{i2})$ and $1 - \frac{d^*}{x_i} < \frac{\rho_i}{g}$ together imply the required result.
Thus, the optimum fraction
$$\phi^*_i = \begin{cases} 
1 & \text{if } \alpha \leq \lambda \\
\max\{1 - \frac{d^*}{x_i}, \frac{\rho_i}{g}\} & \text{otherwise.}
\end{cases}$$
where $d^* = (\frac{\alpha - \lambda}{\beta^2})^{\frac{1}{\beta-1}}$. Q.E.D.

**Proof of Proposition 1:** Note that for a Type 1 DI, only constraint $\phi_i \leq \frac{\rho_i}{g}$ is relevant as $\phi_i \leq 1 \leq \frac{\rho_i}{g}$. Now, consider the following two cases:

- If $\alpha + hg \leq 2\beta g + \lambda$, then from Lemma EC.1, we have $\phi^*_i = 1$ for a Type 1 DI $i \in I$. Thus, $\phi^*_i = 1$.
- If $\alpha + hg > 2\beta g + \lambda$, then from Lemma EC.2, we have the social planner’s optimum solution for a Type 1 DI $i \in I$ as follows:
$$\phi^*_i = \begin{cases} 
1 - \frac{c^*}{x_i} & \text{if } c^* < x_i \\
0 & \text{otherwise}
\end{cases}$$
$$\text{or, } \phi^*_i = \max\{1 - \frac{c^*}{x_i}, 0\}$$
as for a Type 1 DI, we have $1 - \frac{c^*}{x_i} < 1 \leq \frac{\rho_i}{g}$. Q.E.D.

**Proof of Proposition 2:** For Type 2 DIs, both the constraints $\phi_i \leq \frac{\rho_i}{g}$ and $\phi_i \geq \frac{\rho_i}{g}$ are relevant. Thus, to obtain the social planner’s optimum, we compare the optimum objective values $\Upsilon_1(\phi^*_i)$ and $\Upsilon_2(\phi^*_i)$ corresponding to these two constraints. We divide our comparison in the following cases:

- If $(\alpha + hg \leq 2\beta g + \lambda)$ and $(\alpha \leq \lambda)$, then from Lemma EC.1 (resp., EC.3), we have $\phi^*_i = \frac{\rho_i}{g}$ (resp., $\phi^*_i = 1$). The social planner’s objective values corresponding to these solutions are:
$$\Upsilon_1(\phi^*_i) = \lambda(1 - \frac{\rho_i}{g})x_i + \gamma(1 - \frac{\rho_i}{g})x_i^p + \alpha \frac{\rho_i}{g}x_i + \beta(1 - \rho_i)x_i + h\rho_i x_i$$
$$= (\alpha - \lambda)\frac{\rho_i}{g}x_i + \lambda x_i + \gamma(1 - \frac{\rho_i}{g})x_i^p + \beta(1 - \rho_i)x_i + h\rho_i x_i \quad \text{(EC.5)}$$
$$\Upsilon_2(\phi^*_i) = \alpha x_i + \beta(1 - \rho_i)x_i + h\rho_i x_i$$
$$= (\alpha - \lambda)x_i + \lambda x_i + \beta(1 - \rho_i)x_i + h\rho_i x_i \quad \text{(EC.6)}$$

It follows from EC.5 and EC.6 that $\Upsilon_2(\phi^*_i) < \Upsilon_1(\phi^*_i)$ as $\frac{\rho_i}{g} \leq 1$ and $\alpha \leq \lambda$. Thus, the social planner’s optimum is
$$\phi^*_i = \phi^*_i = 1.$$
If \((\alpha + h g \leq 2\beta g + \lambda)\) and \((\alpha > \lambda)\), then from Lemma EC.1 (resp., EC.3), we have \(\phi_{i2}^* = \frac{\phi_i}{g}\) (resp., \(\phi_{i2}^* = \max\{1 - \frac{\phi_i}{x_i}, \frac{\phi_i}{g}\}\)). Note that the objective values \(Y_1(\phi_{i1})\) and \(Y_2(\phi_{i2})\) coincide at the boundary value \(\phi_{i1} = \phi_{i2} = \frac{\phi_i}{g}\). Then, if \(\max\{1 - \frac{\phi_i}{x_i}, \frac{\phi_i}{g}\} = \frac{\phi_i}{g}\), then \(\phi_{i2}^* = \frac{\phi_i}{g}\) and \(Y_1(\frac{\phi_i}{g}) = Y_2(\frac{\phi_i}{g})\). Similarly, if \(\max\{1 - \frac{\phi_i}{x_i}, \frac{\phi_i}{g}\} = 1 - \frac{\phi_i}{x_i}\), then \(\phi_{i2} = 1 - \frac{\phi_i}{x_i}\) and \(Y_2(1 - \frac{\phi_i}{x_i}) \leq Y_1(\frac{\phi_i}{g}) = Y_2(\frac{\phi_i}{g})\). Thus, the social optimum is

\[
\phi_{i2}^* = \max\{1 - \frac{\phi_i}{x_i}, \frac{\phi_i}{g}\}.
\]

- If \((\alpha + h g > 2\beta g + \lambda)\) and \((\alpha \leq \lambda)\), then from Lemmas EC.2 and EC.3, we have

\[
\phi_{i1}^* = \begin{cases} 
\min\{1 - \frac{\phi_i}{x_i}, \frac{\phi_i}{g}\} & \text{if } c^* < x_i \\
0 & \text{otherwise}
\end{cases}
\]

and \(\phi_{i2}^* = 1\). We consider the following subcases:

- If \(c^* < x_i\) and

  - if \((1 - \frac{\phi_i}{g})x_i \geq c^*\), then \(\phi_{i1}^* = \frac{\phi_i}{g}\). Using an argument similar to that for the case when \((\alpha + h g \leq 2\beta g + \lambda)\) and \((\alpha \leq \lambda)\), the social planner’s optimum is

\[
\phi_{i2}^* = 1 - \frac{\phi_i}{x_i}
\]

and \(\phi_{i2}^* = 1\) with

\[
Y_2(\phi_{i2}^*) = \alpha x_i + \beta(1 - \rho_i)x_i + h\rho_i x_i.
\]

By comparing the two objective values, we observe that \(Y_2(\phi_{i2}^*) < Y_1(\phi_{i1}^*)\) if \((1 - \frac{\phi_i}{g})x_i > \frac{\gamma(1 - \rho_i)x_i}{\lambda h - 2\beta g}\).

Thus, the social planner’s optimum is:

\[
\phi_{i2}^* = \begin{cases} 
1 & \text{if } (1 - \frac{\phi_i}{g})x_i > \frac{\gamma(1 - \rho_i)x_i}{\lambda h - 2\beta g} \\
1 - \frac{\phi_i}{x_i} & \text{otherwise}.
\end{cases}
\]

From the above two cases, we have

\[
\phi_{i2}^* = \begin{cases} 
1 & \text{if } (1 - \frac{\phi_i}{g})x_i \geq \min\{c^*, \frac{\gamma(1 - \rho_i)x_i}{\lambda h - 2\beta g}\} \\
1 - \frac{\phi_i}{x_i} & \text{otherwise}.
\end{cases}
\]

- If \(c^* \geq x_i\), we have \(\phi_{i1}^* = 0\) with

\[
Y_1(\phi_{i1}^*) = \lambda x_i + \gamma x_i^* + \beta(1 + \rho_i)x_i.
\]
and \( \phi_{i2}^* = 1 \) with

\[
Y_2(\phi_{i2}^*) = \alpha x_i + \beta (1 - \rho_i) x_i + h \rho_i x_i
\]

Note that \( Y_2(\phi_{i2}^*) < Y_1(\phi_{i1}^*) \) if \( (1 - \frac{\rho_i}{g}) x_i > \frac{\gamma (p_x e^{c'(p-1)} - x_i^g)}{h g - 2 \beta g} \). Thus, the social planner’s optimum is:

\[
\phi_i^* = \begin{cases} 
1 & \text{if } (1 - \frac{\rho_i}{g}) x_i > \frac{\gamma (p_x e^{c'(p-1)} - x_i^g)}{h g - 2 \beta g} \\
0 & \text{otherwise.}
\end{cases}
\]

- **If** \( (\alpha + h g > 2 \beta g + \lambda) \) and \( (\alpha > \lambda) \), then from Lemmas EC.2 and EC.3, we have \( \phi_{i1}^* = \begin{cases} 
\min\{1 - \frac{c^*}{x_i}, \frac{\rho_i}{g}\} & \text{if } c^* < x_i \\
0 & \text{otherwise}
\end{cases} \) and \( \phi_{i2}^* = \max\{1 - \frac{d^*}{x_i}, \frac{\rho_i}{g}\} \). Now consider the following sub-cases:

  . **if** \( (1 - \frac{\rho_i}{g}) x_i \geq c^* \) and

    - **if** \( (1 - \frac{\rho_i}{g}) x_i \leq d^* \), then \( \phi_{i1}^* = \phi_{i2}^* = \frac{\rho_i}{g} \). Thus, \( \phi_i^* = \frac{\rho_i}{g} \).

    - **if** \( (1 - \frac{\rho_i}{g}) x_i > d^* \), then \( \phi_{i1}^* = \frac{\rho_i}{g} \) and \( \phi_{i2}^* = 1 - \frac{d^*}{x_i} \). Note that \( Y_1(\phi_{i1}) \) and \( Y_2(\phi_{i2}) \) coincide at the boundary value \( \phi_{i1} = \phi_{i2} = \frac{\rho_i}{g} \). Moreover, \( \phi_{i2}^* = 1 - \frac{d^*}{x_i} \) implies that \( Y_2(\phi_{i2}^*) < Y_1(\frac{\rho_i}{g}) \). Thus, we have

    \[
    \phi_i^* = \phi_{i2}^* = 1 - \frac{d^*}{x_i}
    \]

From the above two cases, the social planner’s optimum is:

\[
\phi_i^* = \max\{1 - \frac{d^*}{x_i}, \frac{\rho_i}{g}\}
\]

. **if** \( (1 - \frac{\rho_i}{g}) x_i < c^* \) and \( (1 - \frac{\rho_i}{g}) x_i \leq d^* \), then \( \phi_{i1}^* = \max\{1 - \frac{c^*}{x_i}, 0\} \) and \( \phi_{i2}^* = \frac{\rho_i}{g} \). Note that \( Y_1(\phi_{i1}^*) < Y_1(\frac{\rho_i}{g}) = Y_2(\frac{\rho_i}{g}) \). Hence, the social planner’s optimum is:

\[
\phi_i^* = \max\{1 - \frac{c^*}{x_i}, 0\}
\]

. **if** \( d^* < (1 - \frac{\rho_i}{g}) x_i < c^* \) and

  - **if** \( c^* < x_i \), then \( \phi_{i1}^* = 1 - \frac{c^*}{x_i} \) and \( \phi_{i2}^* = 1 - \frac{d^*}{x_i} \). Thus, we have

    \[
    Y_1(\phi_{i1}^*) = \lambda c^* + \gamma c'p + \alpha (x_i - c^*) + \beta \{ (x_i - g(x_i - c^*)) + (\rho_i x_i - g(x_i - c^*)) \} + h g(x_i - c^*)
    \]

    \[
    Y_2(\phi_{i2}^*) = \lambda d^* + \gamma d'p + \alpha (x_i - d^*) + \beta (1 - \rho_i) x_i + h \rho_i x_i
    \]
By comparing the two objective values, we obtain the social planner’s optimum as follows:

\[
\phi^*_i = \begin{cases} 
1 - \frac{d^*}{x_i} & \text{if } (1 - \frac{\rho_i}{g})x_i > \frac{\gamma(p-1)(c^* p - d^* p)}{h g - 2 b g} \\
1 - \frac{c^*}{x_i} & \text{otherwise}.
\end{cases}
\]

If \( c^* \geq x_i \), then \( \phi^*_{i1} = 0 \) and \( \phi^*_{i2} = 1 - \frac{d^*}{x_i} \). Thus, we have

\[
\Upsilon_1(\phi^*_{i1}) = \lambda x_i + \gamma x_i^p + \beta (1 + \rho_i)x_i \\
\Upsilon_2(\phi^*_{i2}) = \lambda d^* + \gamma d^p + \alpha (x_i - d^*) + \beta (1 - \rho_i)x_i + h \rho_i x_i
\]

Note that \( \Upsilon_2(\phi^*_{i2}) < \Upsilon_1(\phi^*_{i1}) \) if \( (1 - \frac{\rho_i}{g})x_i > \frac{\gamma(p-1)(c^* p - d^* p)}{h g - 2 b g} \). Thus, the social’s planner’s optimum is:

\[
\phi^*_i = \begin{cases} 
1 - \frac{d^*}{x_i} & \text{if } (1 - \frac{\rho_i}{g})x_i > \frac{\gamma(p-1)(c^* p - d^* p)}{h g - 2 b g} \\
0 & \text{otherwise}.
\end{cases}
\]

Q.E.D.

**EC.2. Sensitivity Analysis for Indirect Cost Parameters**

To perform the sensitivity analysis for \( p \) and \( \gamma \), we conducted a numerical experiment where errors of up to \( \pm 10\% \) were introduced in measuring these parameters. We then compared the optimal social cost (which is same as the social cost under our mechanism) calculated with the true value of a parameter against the social cost calculated using an incorrect value of the parameter. Our experiments show that despite a fairly large measurement error in \( p \) (upto \( \pm 10\% \)) the social cost only increases by a maximum of 5.48%. Similarly, the social cost only increases by a maximum of 0.005% under a similar variation in the coefficient \( \gamma \). To further test the robustness of our proposed coordination mechanism, we evaluated its performance in situations where errors in both parameters were simultaneously introduced. The maximum increase in the social cost with simultaneous errors is only 3.48%. Thus, we conclude that our proposed mechanism can be expected to work very well even if only crude estimates of \( \gamma \) and \( p \) are available.

**EC.3. Sensitivity Analysis for Uncertainty**

Our discussion with Brink’s indicated that, for a particular DI, the total weekly deposits of used-cash (resp., the total weekly demand for fit cash) varies uniformly within a small interval around a
mean value. Also, for a DI, the ratio \( \rho \) of the weekly demand to weekly deposits is (approximately) a constant. To see this, it is important to note that, in our study, the demand and deposits of a DI represent aggregated amounts over a large number of its branches. The Fed has mandated that, within a Federal Reserve zone (there are a total of 12 Fed zones in the U.S.), a DI and all its branches must transact with the Fed as a single entity (Federal Reserve 2003). Thus, all the branches of a DI in a single zone (typically more than 100) communicate (i.e., deposit/withdraw cash) with the Fed as a single entity. It is, therefore, not surprising that the weekly demand-to-deposit ratio for a particular DI does not change much across weeks. We, therefore, assume that the weekly deposits \( x \) and the weekly demand \( y = \rho x \) are known and constant. Brink’s also indicated that typically 75% of used-cash is fit. Therefore, we assume that the fit-cash fraction \( g \) is a constant. However, it is reasonable to assume that there is some uncertainty in these parameters, which may impact the optimal decisions. We now describe the sensitivity analysis conducted to assess the magnitude of this impact.

The purpose of our numerical study is to study the sensitivity of the optimal societal cost as well as the coordination mechanism with respect to variation in the parameters \( x \), \( y \), and \( g \). We first describe the setting to generate the test bed.

**Parameter Settings:** As stated above, we assume that the realizations of \( x \), \( y \), and \( g \), are uniformly distributed around their respective means. Let \( \bar{x} \) (resp., \( \bar{y} \)) and \( \bar{g} \) denote the mean weekly deposit (resp., demand), for a DI, and mean fit-cash fraction, respectively. As explained in Section 4.3, we estimate the mean weekly demand (in number of bundles) of fit cash of a DI to be between 2,500 and 7,500 bundles. For experimental purposes, we characterize the mean weekly demand of a DI as either low, medium, or high. A DI with low (resp., medium, high) mean weekly demand corresponds to \( \bar{y} = 2,500 \) (resp., 5,000, 7,500) bundles.

For a DI, let \( \hat{\rho} = \frac{\bar{y}}{\bar{x}} \) be ratio of the mean demand to mean deposits. Based on the characteristic of the ratio \( \hat{\rho} \) mentioned in Section 4.3, we consider three values of \( \hat{\rho} \): low (0.25), medium (0.75), and high (1.5). Using this, the mean weekly deposits for a DI is set at \( \bar{x} = \frac{\bar{y}}{\hat{\rho}} \). As mentioned earlier, typically about 75% of used-cash is fit. Therefore, we set \( \bar{g} = 0.75 \).
Based on these values, as shown in Table EC.1, we have nine settings for the triplet of means \((\bar{x}, \bar{y}, \bar{g})\). For a comprehensive testing, we generated symmetric intervals, centered around the respective means, for the deposits, demands, and fit-cash fraction. For each triplet, we considered 5%, 10%, 15%, 20%, and 25% variation around the mean. For example, a 5% variation for \(\bar{x}\) corresponds to the interval \([0.975\bar{x}, 1.025\bar{x}]\). We generated 1000 random realizations for each of the parameters \(x, y\) and \(g\), from their respective intervals for each of the nine settings in Table EC.1, for each value of the percentage variation (for a total of \(1000 \times 9 \times 5 = 45000\) instances).

As described in Section 4.3, the holding cost per bundle per week \((h)\) was set to $20. After comprehensive testing, we present two parameter settings of the cost parameters (Table EC.2): the setting under Case 1 illustrates the typical impact of uncertainty while the setting under Case 2 is shown specifically to illustrate the worst-case impact we observed in our experiments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Case 1</th>
<th>Case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI’s Transportation Cost (\beta)</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>DI’s Holding Cost ((h))</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>DI’s Fit-Sorting Cost (\alpha)</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Fed’s Fit-Sorting Cost (\lambda)</td>
<td>2</td>
<td>1.5</td>
</tr>
<tr>
<td>Indirect Cost Coefficient (\gamma)</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Indirect Cost Exponent (p)</td>
<td>2</td>
<td>1.7</td>
</tr>
</tbody>
</table>

The goal of the sensitivity analysis is to answer the following question: How far will the social planner be from the optimal cost (i.e. one corresponding to the decisions based on the realized values
of a DI’s parameters) when decisions on fit-sorting are taken based on the DI’s mean parameters \((\bar{x}, \bar{y}, \bar{g})\)? Thus, we first evaluate the optimal fit-sorting fraction, say \(\phi_D^*\), by considering the mean values of the parameters. Next, for each randomly-generated instance, by assuming full knowledge, i.e., assuming that the Fed knows the DIs’ realized values of the parameters, we evaluate the optimal decision \(\phi_V^*\).

For an instance, let \(SC(\phi_D^*)\) denote the cost, evaluated using the true (i.e., realized) values of the parameters, corresponding to the fit-sorting fraction \(\phi_D^*\). Let \(SC(\phi_V^*)\) be the cost corresponding to the fit-sorting fraction \(\phi_V^*\). Over the 1000 instances corresponding to each row of Table EC.1 and each level of variation (i.e., 5%, 10%, 15%, 20%, and 25%), we compute the average optimal societal cost \(\overline{SC}(\phi_V^*)\) and compare it with the average societal cost \(\overline{SC}(\phi_D^*)\) incurred using \(\phi_D^*\) as follows:

\[
\% \text{ Gap} = \frac{\overline{SC}(\phi_D^*) - \overline{SC}(\phi_V^*)}{\overline{SC}(\phi_V^*)} \times 100
\]

Below, we summarize our findings.

**Case 1:** This case represents the typical impact of uncertainty on the optimal societal cost. Figure EC.1 illustrates this impact for Triplets 1, 2, and 3 of Table EC.1; the results for the other triplets are similar. The cost parameters are fixed as indicated in Table EC.2 under Case 1. As can be seen from the figure, even under a considerable amount (25%) of uncertainty, the percentage gap between the optimal cost and that obtained using \(\phi_D^*\) is negligible.

**Case 2:** Figure EC.2 illustrates the worst-case impact we observed in our experiments. In this case, the percentage gap between the optimal cost and that obtained using \(\phi_D^*\) is negligible when \(\bar{\rho}\) is high (1.5) or low (0.25). For the instances corresponding to \(\bar{\rho} = 0.75\), the average gap is around 1.5% (with the worst case of about 8% when the variation around the mean is 25%). In this case too, the results were similar for the other triplets of Table EC.1.

We conclude that using the fit-sorting fraction \(\phi_D^*\), obtained using the mean values of the parameters, provides a very good approximation for the societal cost.

Next, we study the behavior of our coordination mechanism if a DI uses the mean values \((\bar{x}, \bar{y}, \bar{g})\)
Figure EC.1  Impact of the Uncertainty on the Social Optimal Cost for Triplets 1, 2 and 3 under Case 1.

Figure EC.2  Impact of the Uncertainty on the Social Optimal Cost for Triplets 1, 2 and 3 under Case 2.
for its decision on fit-sorting. We start with an understanding of the structure of the mechanism defined in the Section 5.

**Coordination Mechanism:**

Recall that our coordination mechanism is based on *adjusted fit-sorting fees*. The fee is defined (see Section 5) as $\Omega(w) = -k(x, \rho) + \lambda w + \gamma w^p$, where $w$ is the used cash deposited to the Fed and $k(x, \rho) = \lambda x(1 - \frac{\rho}{g})^+ + \gamma x^p(1 - \frac{\rho}{g})^+$. Note that this fee is based on the *realized* values of the parameters $x, \rho,$ and $g$. In other words, a DI can make its optimal fit-sorting decision, $\psi^*_D$, by using the mean values of the parameters. The actual fee, however, is charged on their realized values. Clearly, the decision (of fit-sorting a $\psi^*_D$ fraction of deposits) may be more or less than the optimal fraction (i.e., the optimum fraction for the problem with the realized values of the parameters). Thus, the actual fee charged to a DI may be more or less than that corresponding to $\psi^*_D$. We use the setting in Table EC.1 and a variation of 25%. Consider a DI with mean values of the parameters corresponding to a row of Table EC.1. We show numerically that, over the long term (over several weeks, with one random instance generated each week from the corresponding intervals to specify the realized values of the parameters in that week), the average fee using $\psi^*_D$ converges to the average fee the DI would have paid if it had a priori knowledge of the realized values. This is depicted in the Figures EC.3 for Triplet 1, i.e., $(10000, 2500, 0.75)$ over 400 weeks under Case 1 of Table EC.2. The results are similar for the other settings in Table EC.1 and Table EC.2. Thus, our mechanism, when based on the fit-sorting fraction $\psi^*_D$ computed using the mean values of the parameters, has the desired convergence property.

**Remark EC.2.** The structure of our fee is such that the Fed does not have to change it every week even in the presence of uncertainty. As mentioned above, a DI can compute an “anticipated fee” by using an optimal fit-sorting decision, $\psi^*_D$, corresponding to the mean values of the parameters. The actual fee, is charged on the realized values of the parameters.

**Remark EC.3.** Since uncertainty has negligible impact on the optimal societal cost and that decisions based on mean values of the parameters are sufficiently accurate. Hence, for describing
Figure EC.3  The Convergence of the Average Fee using Decisions Based on Mean Values of the Parameters under Case 1.

our test bed in the experiment of Section 4.3, we assume a deterministic setting for the deposit, demand, and the fit-cash fraction.