Economic Quarterly—Volume 98, Number 3—Third Quarter 2012—Pages 159–183

Debit Card Interchange Fee Regulation: Some Assessments and Considerations

Zhu Wang

In the summer of 2011, the Federal Reserve Board of Governors issued a final rule governing debit card interchange fees. This regulation, named Regulation II (Debit Card Interchange Fees and Routing), was required by the Durbin Amendment to the Dodd-Frank Act. The regulation, which went into effect on October 1, 2011, limits the maximum permissible interchange fee that a covered issuer can collect from merchants for a debit card transaction.

The Durbin Amendment and the resulting regulation were created to resolve the long-time conflicts between card issuers and merchants regarding payment card interchange fees. The interchange fee is the amount that a merchant has to pay the cardholder's bank (the so-called issuer) through the merchant acquiring bank (the so-called acquirer) when a card payment is processed. Merchants have criticized that card networks (such as Visa and MasterCard) and their issuing banks have used market power to set excessively high interchange fees, which drive up merchants' costs of accepting card payments. Card networks and issuers disagree, countering that interchange fees have been properly set to serve the needs of all parties in the card system, including funding better consumer reward programs that could also benefit merchants.

By capping debit card interchange fees, the regulation has generated significant impact on the U.S. payments industry since its implementation. The most visible impact is the drop of multibillion-dollar

I thank Kartik Athreya, Borys Grochulski, Sam Marshall, and Ned Prescott for helpful comments, and John Muth for excellent research assistance. The views expressed herein are solely those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: zhu.wang@rich.frb.org.

annual revenues for card issuers in terms of the interchange fees that they collect from merchants. Meanwhile, the regulation has yielded other intended and unintended consequences. In this article, we review the regulation's impact from both positive and normative perspectives. We first look into the empirical evidence of the regulation's first-year effects on different players in the debit card market, namely issuers, merchants, and consumers. We then provide a simple two-sided market model, based on the work of Rochet and Tirole (2011), to assess the regulation's implications on payments efficiency. The model sheds light on important policy questions, for example, whether the debit card market performs inefficiently without regulation and whether the Durbin regulation can improve market outcome. Finally, we extend the model to explain the regulation's unintended consequence on small-ticket merchants and discuss an alternative regulatory approach.

The article is organized as follows. Section 1 provides the background of payment card markets and the interchange fee regulation. Section 2 reviews the empirical evidence on the regulation's impact on different players in the debit card market. Section 3 lays out a simple model of the payment card market and discusses the regulation's implication on payments efficiency. We then extend the model to address the regulation's unintended consequence on small-ticket merchants. Finally, Section 4 provides concluding remarks.

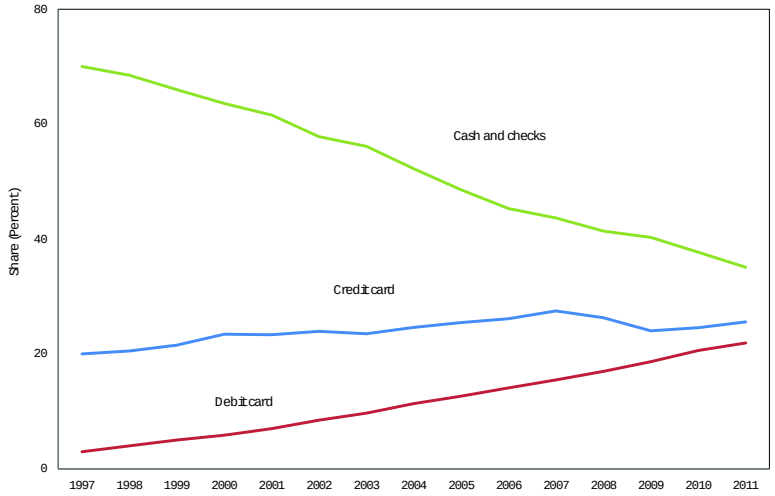
1. INDUSTRY BACKGROUND

As payments migrate from paper to electronic forms, credit and debit cards have become an increasingly important part of the U.S. payments system. Recent data show that the payment share of credit and debit cards in personal consumption expenditures rose from 23 percent in 1997 to 48 percent in 2011, while the share of cash and checks dropped from 70 percent to 35 percent (Figure 1).[1] In 2011, debit cards were used in 49 billion transactions for a total value of $1.8 trillion, and credit cards were used in 26 billion transactions for a total value of $2.1 trillion.

Along with this development has come controversy. Merchants are critical of the fees that they pay to accept cards. These fees are often referred to as the "merchant discounts," which are composed mainly of interchange fees paid by merchants to card issuing banks through merchant acquiring banks. Merchants believe that the card networks

[1] The data are drawn from various issues of the *Nilson Report*. Payment shares not shown in Figure 1 include the automated clearing house and some other miscellaneous types.

Figure 1 Payment Shares of U.S. Personal Consumption Expenditures



and issuing banks have wielded their market power to set excessively high interchange fees. The card networks and issuers counter that these interchange fees are necessary for covering issuers' costs as well as providing rewards to cardholders, which may also benefit merchants by making consumers more willing to use the cards.

Market Overview

To understand the interchange fee controversy, some familiarity with the payment card markets is helpful. Credit and debit cards are two of the most popular general-purpose payment cards in the United States.[2] Credit cards typically provide credit or float to cardholders, while debit cards directly draw from the cardholder's bank account right after each transaction. Debit card payments are authorized either by the cardholder's signature or by a personal identification number (PIN). The

[2] Pre-paid cards are another type of general-purpose card, but their market size is much smaller compared with credit and debit cards. In 2011, the transaction value of pre-paid cards accounted for 2 percent of U.S. personal consumption expenditures (Data source: *Nilson Report*).

former is called signature debit and the latter is called PIN debit. In terms of transaction volume, signature debit accounts for 60 percent of debit transactions, while PIN debit accounts for 40 percent.

Visa and MasterCard are the two major credit card networks in the United States. They provide card services through member financial institutions and account for 85 percent of the U.S. consumer credit card market.[3] Visa and MasterCard are also the primary providers of debit card services. The two networks split the signature debit market, with Visa holding 75 percent of the market share and MasterCard holding 25 percent.[4] In contrast, PIN debit transactions are routed over the PIN debit networks. Currently, there are 14 PIN debit networks in the United States. Interlink, Star, Pulse, and NYCE are the top four networks, together holding 90 percent of the PIN debit market. The largest PIN network, Interlink, is operated by Visa.

Visa, MasterCard, and PIN debit networks are commonly referred to as four-party schemes because four parties are involved in each transaction in addition to the network whose brand appears on the card. These parties include: (1) the cardholder who makes the purchase; (2) the merchant who makes the sale and accepts the card payment; (3) the financial institution that issues the card and makes the payment on behalf of the cardholder (the so-called issuer); and (4) the financial institution that collects the payment on behalf of the merchant (the so-called acquirer).

In a four-party card scheme, interchange fees are collectively set by the card network on behalf of their member issuers. For a simple example of how interchange functions, imagine a consumer making a $50 purchase with a payment card. For that $50 item, the merchant would get approximately $49. The remaining $1, known as the merchant discount, gets divided up. About $0.80 would go to the card issuing bank as the interchange fee, and $0.20 would go to the merchant acquiring bank (the retailer's account provider). Interchange fees serve as a key element of the four-party scheme business model and generate significant revenues for card issuers. In 2009, U.S. card issuers made approximately $48 billion revenue in interchange fees, with debit interchange revenues being $17 billion and credit interchange revenues being $31 billion.[5]

---

[3] American Express and Discover are the other two credit card networks holding the remaining market shares. They handle most card issuing and merchant acquiring by themselves, and are called "three-party" systems. For a "three-party" system, interchange fees are internal transfers.

[4] Discover has recently entered the signature debit market, but its market share is small.

[5] See Levitin (2010).

Figure 2 Interchange Fees for a $50 Transaction

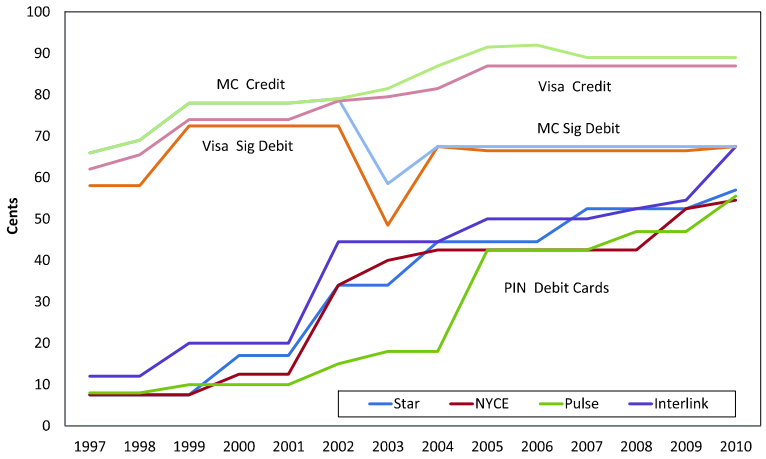


Figure 2 plots the interchange fee for a $50 non-supermarket transaction for Visa and MasterCard credit cards, signature debit cards, as well as the top four PIN debit cards in the United States.[6] As the figure shows, credit and PIN debit interchange fees have been rising since the late 1990s, while signature debit interchange fees came down in 2003 before rising again soon after.[7] Over the years, the gap of interchange fees between PIN debit and signature debit has also narrowed substantially.

Interchange Battles

Merchants criticize the interchange fees for being excessively high. They point out that the high and rising interchange fees deviate from cost basis and are in sharp contrast to the falling card processing and fraud costs during the same period.[8] In recent years, merchant groups

[6] Data source: *American Banker* (various issues).

[7] The temporary drop of signature debit interchange fees was due to the settlement of the Wal-Mart case, which allowed merchants who accept Visa or MasterCard credit cards to not have to accept their signature debit cards.

[8] Payment cards is primarily an information-processing industry. As the information technology progresses, the relative prices of computers, communications, and

launched a series of litigation against what they claim is anticompetitive behavior by the card networks and their issuers. Some of the lawsuits have been aimed directly at interchange fees, including both credit and debit cards. For example, a group of class-action suits filed by merchants against Visa and MasterCard in 2005 alleged that the networks violated antitrust laws by engaging in price fixing. As a result, Visa and MasterCard recently agreed to a $7.25 billion settlement with U.S. retailers, which could be the largest antitrust settlement in U.S. history.[9] Other merchant lawsuits have focused not on interchange fees *per se*, but on alleged anticompetitive practices. A prime example is the lawsuit filed by Wal-Mart and other merchants in 1997 against the networks' honor-all-cards rule, which required a merchant accepting a network's credit cards to also accept its signature debit cards. The Wal-Mart case was settled in 2003. As a result, Visa and MasterCard agreed to unbundle credit cards and signature debit cards, and also temporarily lowered their interchange fees on signature debit cards (Figure 2).

The interchange fee controversy has also attracted great attention from policymakers, who are concerned that interchange fees inflate the cost of card acceptance without leading to proven efficiency.[10] In the two years leading up to the passage of the Durbin Amendment, three separate bills restricting interchange fees were introduced in Congress: a House version of the Credit Card Fair Fee Act of 2009, a Senate version of the same act, and the Credit Card Interchange Fees Act of 2009.[11] Before any of these bills could be brought to a vote, the Dodd-Frank Act was passed and signed into law in July 2010. A provision of the Dodd-Frank Act, known as the Durbin Amendment, mandates a regulation aimed at debit card interchange fees and increasing competition in the payment processing industry.

---

software have been declining rapidly, which should have driven down the card processing costs. Meanwhile, industry statistics show that card fraud rates also have been declining steadily. For the U.S. credit card industry as a whole, the net fraud losses as a percent of total transaction volume has dropped from roughly 16 basis points in 1992 to about 7 basis points in 2009. Data source: *Nilson Report* (various issues).

[9] Visa, MasterCard, and their major issuers reached the settlement agreement with merchants in July 2012. The settlement is currently pending final court approval.

[10] Worldwide, more than 20 countries and areas have started regulating or investigating interchange fees. Primary examples include Australia, Canada, the European Union, France, Spain, and the United Kingdom (Bradford and Hayashi 2008).

[11] None of the bills called for direct regulation of interchange fees, and all three applied to interchange fees for both credit and debit cards (Hung 2009).

Durbin Amendment and Regulation

The Durbin Amendment of the Dodd-Frank Act directs the Federal Reserve Board to regulate debit card interchange fees "reasonable and proportional to the cost incurred by the issuer with respect to the transaction." The Federal Reserve Board subsequently issued the final rule on debit cards in July 2011, effective on October 1, 2011.

The Federal Reserve Board ruling establishes a cap on the debit interchange fees that financial institutions with more than $10 billion in assets can charge to merchants through merchant acquirers. The permissible fees were set based on the Federal Reserve Board's evaluation of issuers' costs associated with debit card processing, clearance, and settlement. The resulting interchange cap is composed of the following: a base fee of 21 cents per transaction to cover the issuer's processing costs, a five basis point adjustment to cover potential fraud losses, and an additional 1 cent per transaction to cover fraud prevention costs if the issuer is eligible. This cap applies to both signature and PIN debit transactions.

In addition, the regulation sets rules that prohibit certain restraints imposed by card networks on merchants. First, networks can no longer prohibit merchants from offering customers discounts for using debit cards versus credit cards. This gives merchants a way to steer consumers toward using less expensive payment means.[12] Second, issuers must put at least two unaffiliated networks on each debit card and are prohibited from inhibiting a merchant's ability to direct the routing of debit card transactions. This gives merchants more freedom for routing debit transactions through less costly networks. Third, networks can no longer forbid merchants from setting minimum values for credit card payments. Going forward, merchants are allowed to establish such minimum values as long as the minimum does not exceed $10.

2. EMPIRICAL IMPACT

A direct impact of the debit card regulation is the redistribution of interchange revenues from issuers to merchants. According to a Federal Reserve study, the average debit card transaction in 2009 was approximately $40. Post regulation, the maximum interchange fee applicable to a typical debit card transaction is capped at 24 cents (21 cents + ($40 × .05%) + 1 cent), which is about half of its pre-regulation

[12] Since the passage of the Cash Discount Act in 1981, merchants have been allowed to offer their customers discounts for paying with cash or checks. However, the card networks have continued to prohibit merchants from offering customers discounts for using one type of card rather than another.

industry average level. As a result, issuers were expected to lose multi-billion dollar annual revenues in terms of the interchange fees that they collect from merchants. In this section, we look into the empirical evidence of the regulation's first-year effects on different players in the debit card market.

Impact on Issuers

The regulation reduces debit card interchange fees by about half and also introduces more competition by abolishing certain network restrictions. As a result, issuers face a big drop in their interchange revenues. Meanwhile, the regulation allows small issuers to be exempt from the interchange fee cap—those with less than $10 billion in assets.[13]

To assess the regulation's impact on covered and exempt issuers, we conduct a study on a subsample of card issuers, which includes all the commercial banks that report their interchange revenues in the quarterly Call Report. Our sample includes 7,049 commercial banks between the first quarter of 2009 and the third quarter of 2012. Among those, we identify 102 covered issuers and 6,969 exempt issuers. The status of exemption is based on whether the bank asset value exceeds the $10 billion threshold as of prior year end.[14]

We first compare the interchange revenues of all covered and exempt banks right before and right after the regulation, as shown in Figure 3 with solid lines. Covered banks had a substantial loss of interchange revenues during the period. Between the third quarter and fourth quarter of 2011 (i.e., the immediate quarter before and after the regulation took effect), covered banks' interchange revenues dropped by $2.1 billion (or 29 percent), equivalent to an $8.5 billion drop annually. In contrast, exempt banks' quarterly interchange revenues did not fall during the same period, instead rising by $11.8 million (or 2 percent).

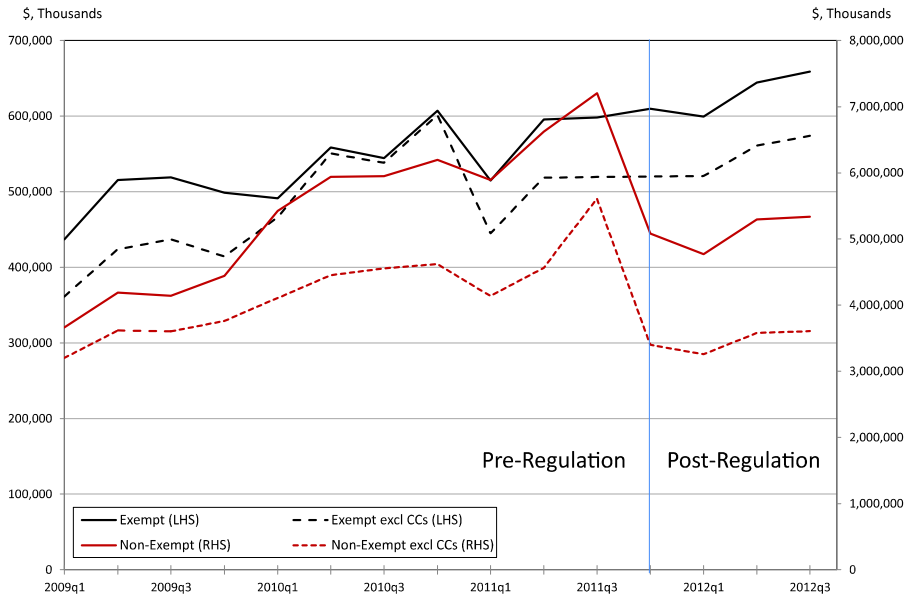
We also compare the interchange revenues one year before and one year after the regulation to control for potential seasonality. The result is similar: Covered banks' annual interchange revenues dropped by $5.4 billion (or 21 percent), while exempt banks' annual interchange revenues increased by $198 million (or 9 percent).

For an alternative check, we construct counterfactual interchange revenues for one year after the regulation (the fourth quarter of 2011

[13] This exemption is applied at the holding company level, to ensure that large issuers cannot evade the regulations by establishing subsidiaries under the size limit.

[14] Note that a bank's exemption status may change as its asset size changes, so the sum of non-exempt banks and exempt banks may exceed the total number of banks in the sample.

Figure 3 Aggregate Interchange Fee Revenues

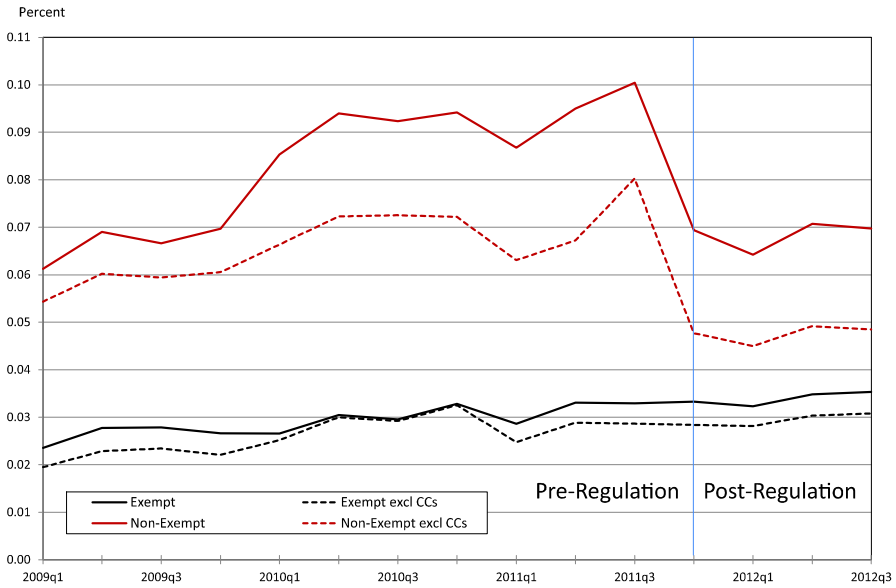


through the third quarter of 2012), assuming that the regulation did not take effect and the annual interchange revenues kept a constant growth rate since two years ago. The finding shows that the annual interchange revenues for covered banks dropped by $10.4 billion (or 34 percent) compared with the counterfactual. In contrast, exempt banks' interchange revenues only dropped by $47 million (or 2 percent).

A limitation of the Call Report data is that they do not separate interchange revenues between debit and credit cards. Therefore, when we conduct the above exercises, we implicitly assume that the changes in interchange revenues were primarily driven by the debit card transactions (but not credit card transactions). In order to focus more on debit interchange fees, we then re-ran the above exercises by excluding mono-lined credit card banks.[15] The pattern, shown in Figure 3 with dashed lines, turns out to be similar. In terms of actual interchange revenues one year before and after the regulation, covered banks'

[15] Mono-lined credit card banks are defined as commercial banks with a minimum of 50 percent of assets in consumer lending and 90 percent of consumer lending in the form of revolving credit. See the "Report to the Congress on the Profitability of Credit Card Operations of Depository Institutions," Board of Governors of the Federal Reserve System, 2011.

Figure 4 Aggregate Interchange Revenues as a Percent of Deposits



annual interchange revenues dropped by $5.1 billion (or 27 percent), while exempt banks' annual interchange revenues increased by $90.9 million (or 4 percent). In terms of the counterfactual comparison, covered banks' annual interchange revenues dropped by $7.4 billion (or 35 percent), while exempt banks' annual interchange revenues only dropped by $31.1 million (or 1 percent).

We also replicated the above exercises by comparing the interchange-revenue-to-bank-deposits ratio. By focusing on the ratio to deposits, we may control for the potential effect of changing bank sizes on interchange revenues. Again, as shown in Figure 4, the results are very similar.

Overall, the empirical evidence suggests that the debit regulation has largely achieved its objective of reducing the interchange revenues for large issuers, while exempt small issuers so far have been well protected.[16]

[16] Further monitoring is needed to evaluate the regulation's long-run impact on issuers. There are three concerns that the exempt small issuers might be adversely affected by the regulation. First, networks may voluntarily lower the interchange rates for small issuers to level the playing field between large and small issuers. Second,

Impact on Merchants

Merchants as a whole have greatly benefited from the reduced interchange fees under the regulation. Presumably, the loss of issuers' interchange revenues would be the gain of the merchants. However, the distribution of the gain appears uneven among merchants. In fact, the regulation has yielded an unintended consequence: Interchange fees rose for small-ticket merchants.

Prior to the regulation, Visa, MasterCard, and most PIN networks offered discounted debit interchange fees to small-ticket transactions as a way to encourage card acceptance by merchants specializing in those transactions. For example, Visa and MasterCard used to set the small-ticket signature debit interchange rate at 1.55 percent of the transaction value plus 4 cents for sales of $15 and below. As a result, a debit card would only charge a 7 cents interchange fee for a $2 sale or 11 cents for a $5 sale. However, in response to the regulation, card networks eliminated the small-ticket discounts, and all transactions (except those on cards issued by exempt issuers) have to pay the maximum cap amount set by the regulation (i.e., 21 cents plus 0.05 percent of the transaction value).[17] For merchants selling small-ticket items, this means that the cost of accepting the same debit card doubled or even tripled after the regulation.

The rising interchange fee on small-ticket sales could affect a large number of transactions. According to the 2010 Federal Reserve Payments Study, in 2009 debit cards were used for 4.9 billion transactions below $5, and 10.8 billion transactions between $5–$15. The former accounts for 8.3 percent of all payment card transactions (including credit, debit, and prepaid cards), and the latter accounts for 18.3 percent. Since merchants may have different compositions of transaction sizes, they could be affected differently by the changes of interchange fees.[18] However, merchants who specialize in small-ticket transactions would be most adversely affected.[19]

---

merchants may offer preferential treatment to cards issued by large issuers that carry lower interchange rates. Third, the regulation requires each debit card be connected to at least two unaffiliated networks and merchants have the freedom to choose the lower-cost routing. This provision took effect after April 2012 and small issuers are not exempt from it.

[17] E.g., in the case of signature debit, any sales below $11 now face a higher interchange rate.

[18] Shy (2012) used the data from the Boston Fed's 2010 and 2011 Diary of Consumer Payment Choice to identify the types of merchants who are likely to pay higher and lower interchange fees under the debit regulation.

[19] E.g., Visa classifies merchant sectors specializing in small-ticket sales, which include local commuter transport, taxicabs and limousine, fast food restaurants, coffee shops, parking lots and garages, motion picture theaters, video rental stores, cashless

In response, many small-ticket merchants have tried to offset their higher rates by raising prices, encouraging customers to pay with alternative payment means, or dropping card payments altogether.[20] In the meantime, a lawsuit was filed in November 2011 in federal court by three of the retail industry's largest trade associations and two retail companies against the Federal Reserve's debit interchange regulation. The lawsuit alleges that the Fed has set the interchange cap too high by including costs that were barred by the law, and "forcing small businesses to pay three times as much to the big banks on small purchases was clearly not the intent of the law and is further evidence that the Fed got it wrong."[21]

The unintended consequence on small-ticket merchants calls for a further examination on the regulation, which we will provide in Section 3.

Impact on Consumers

The regulation's impact on consumers is less clear. On the one hand, merchants argue that with a lower interchange fee, they would be able to offer lower retail prices to consumers. On the other hand, issuers argue that they will have to reduce card rewards and raise banking service fees to consumers in order to make up for the lost interchange revenues.

At this point, little empirical evidence has been reported on the change of merchant prices due to the debit interchange regulation. After all, even if the reduced interchange fees have resulted in lower retail prices, the magnitude would be quite small so it is not easy to measure. Meanwhile, several studies report that consumers now face higher banking and card service fees. A recent Pulse debit issuer study shows that 50 percent of regulated debit card issuers with a reward program ended their programs in 2011, and another 18 percent planned to do so in 2012.[22] The Bankrate's 2012 Checking Survey shows that the average monthly fee of noninterest checking accounts rose by 25 percent

---

vending machines and kiosks, bus lines, tolls and bridge fees, news dealers, laundries, dry cleaners, quick copy, car wash and service stations, etc.

[20] See "Debit-Fee Cap Has Nasty Side Effect," *Wall Street Journal*, December 8, 2011.

[21] Source: "Merchants' Lawsuit Says Fed Failed to Follow Law on Swipe Fee Reform," *Business Wire*, November 22, 2011.

[22] The 2012 Debit Issuer Study, commissioned by Pulse, is based on research with 57 banks and credit unions that collectively represent approximately 87 million debit cards and 47,000 ATMs. Research was conducted in April and May of 2012, and the data provided by issuers is for 2011. The sample is nationally representative, with issuers segmented into "regulated" ($\geq$ \$10 billion in assets) and "exempt" ($<$ \$10 billion in assets) to report on the impact of the interchange provision of Regulation II.

compared with last year, and the minimum balance for free-checking services rose by 23 percent.[23] According to the report, the rising bank fees are largely due to banks' response to recent regulations including the debit interchange cap. In addition, several major banks including Bank of America, Wells Fargo, and Chase attempted to charge a monthly debit card fee to their customers in response to the interchange regulation, but they eventually backed down due to customer outrage.[24]

3. THEORETICAL CONSIDERATIONS

The debit card regulation was created to reduce the interchange fee by capping the fee at the card issuers' marginal cost. To understand the welfare implications of the regulation, we turn to a theoretical analysis in this section.

First, we lay out a simple model based on the work of Rochet and Tirole (2011). The model conceptualizes payment cards as a two-sided market, that is, two end-user groups (i.e., merchants and consumers) who jointly use the card services.[25] The interchange fee serves as a transfer between merchants and consumers to balance their joint demand for using cards. Under the assumption of homogenous merchants, the model shows that (1) market-determined interchange fees tend to exceed the socially optimal level, so reducing interchange fees may improve the payments efficiency; (2) however, capping interchange fees based on issuers' marginal cost does not necessarily restore the social optimum; and (3) the theory suggests an interchange fee regulation based on the merchant transaction benefit of accepting cards.

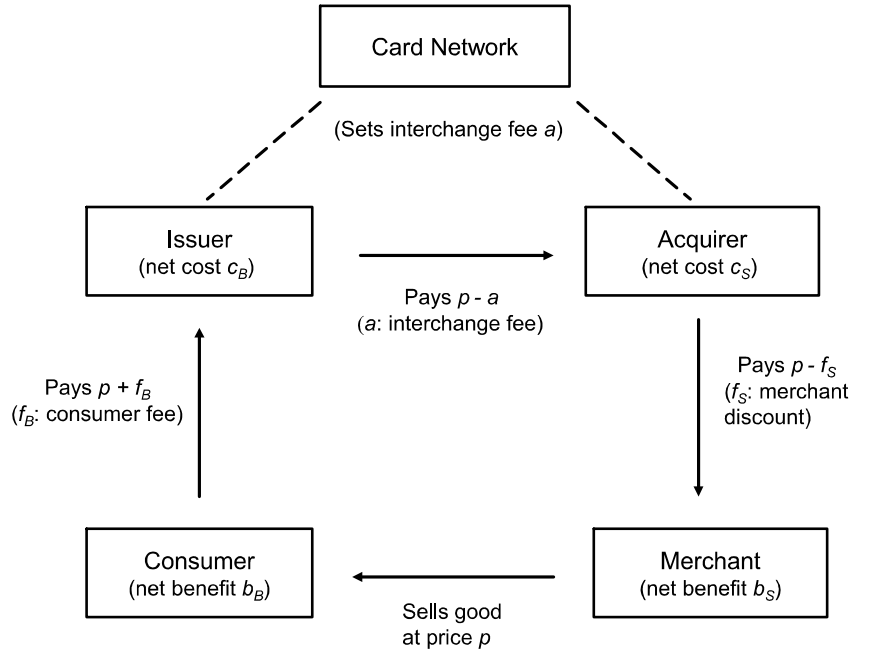
While the simple two-sided market model sheds light on key policy issues related to the interchange fee regulation, it does not address the regulation's unintended consequence on small-ticket merchants. To fill the gap, we then introduce an extension of the model by considering card demand externalities across heterogenous merchant sectors, based on the work of Wang (forthcoming). The findings suggest that an

[23] Bankrate surveyed banks in the top 25 U.S. cities to find the average fees associated with checking accounts in their annual Checking Account Survey, which was conducted in July and August 2012.

[24] Source: "Banks Adding Debit Card Fees," *The New York Times*, September 29, 2011.

[25] In recent years, a sizeable body of literature, called "two-sided market theory," has been developed to evaluate payment card market competition and pricing issues. For instance, Baxter (1983), Rochet and Tirole (2002, 2006, 2011), Schmalensee (2002), Wright (2003, 2004, 2012), Armstrong (2006), Rysman (2007, 2009), Prager et al. (2009), Wang (2010, forthcoming), Weyl (2010), Shy and Wang (2011), and McAndrews and Wang (2012).

Figure 5 A Payment Card System



alternative regulation, capping the weighted average interchange fee, instead of the maximum interchange fee, may restore the social optimum and avoid the unintended consequence on small-ticket merchants.

A Simple Model

We first lay out a model with homogenous merchants, which is a simplified version of Rochet and Tirole (2011). The model considers a payment card system that is composed of five types of players: consumers, merchants, acquirers, issuers, and the card network, as illustrated in Figure 5.

Consumers

There is a continuum of consumers who purchase goods from competitive merchants selling a homogenous good. Consumers have inelastic demand and each buy one unit of the good. Consumers need to decide which store to patronize. They know the stores' price and card acceptance policy before making the choice. Once in the store they then

select a payment method (a card or an alternative payment method such as cash), provided that the retailer indeed offers a choice among payment means. We assume price coherence such that retailers find it too costly to charge different prices for purchases made by different payment means.[26] Whenever a transaction between a consumer (buyer) and a retailer (seller) is settled by card, the buyer pays a fee f_B to her card issuing bank (issuer) and the seller pays a merchant discount f_S to her merchant acquiring bank (acquirer). We allow f_B to be negative, in which case the cardholder receives a card reward. There are no annual fees and all consumers have a card.

The consumer's convenience benefit of paying by card relative to using cash is a random variable b_B drawn from a cumulative distribution function H on the support $[\underline{b}_B, \bar{b}_B]$, which has a monotonic increasing hazard rate.[27] Cardholders are assumed to only observe the realization of b_B once in the store.[28] Because the net benefit of paying by card is equal to the difference $b_B - f_B$, a card payment is optimal for the consumer whenever $b_B \geq f_B$. The proportion of card payments at a store that accepts cards is denoted $D(f_B)$:

$$D(f_B) = \Pr(b_B \geq f_B) = 1 - H(f_B). \tag{1}$$

Let $v(f_B)$ denote the average net cardholder benefit per card payment:

$$\begin{aligned} v(f_B) &= E[b_B - f_B | b_B \geq f_B] \\ &= \frac{\int_{f_B}^{\bar{b}_B} (b_B - f_B) dH(b_B)}{1 - H(f_B)} > 0. \end{aligned} \tag{2}$$

The monotonic hazard rate of H implies that $v(f_B)$ decreases in f_B.

[26] Price coherence is the key feature that defines a two-sided market. Rochet and Tirole (2006) show that the two-sided market pricing structure (e.g., interchange fees) would become irrelevant without the price coherence condition. In reality, price coherence may result either from network rules or state regulation, or from high transaction costs for merchants to price discriminate based on payment means. In the United States, while merchants are allowed to offer their customers discounts for paying with cash or checks, few merchants choose to do so. On the other hand, card network rules and some state laws explicitly prohibit surcharging on payment cards.

[27] The hazard rate is assumed increasing to guarantee concavity of the optimization problem.

[28] This is a standard assumption introduced by Wright (2004) and used in the subsequent literature, which simplifies the analysis of retailers' acceptance of cards without changing the equilibrium outcome. Alternatively, Rochet and Tirole (2002) assume cardholders differ systematically in the benefit that they derive from card payments. However, as shown in Rochet and Tirole (2011), these two alternative assumptions deliver broadly convergent results.

Merchants

Merchants derive the convenience benefit b_S of accepting payment cards (relative to handling cash). By accepting cards under the price coherence assumption, a merchant is able to offer each of its card-holding customers an additional expected surplus of $D(f_B)v(f_B)$, but faces an additional expected net cost of $D(f_B)(f_S - b_S)$ per cardholder. Denote c as the cost of the good. Competitive merchants then set a retail price equal to marginal cost, namely

$$p = c + D(f_B)(f_S - b_S) \tag{3}$$

if they accept cards, or $p = c$ if they reject cards. Consumers choose the stores that accept cards if and only if their increased surplus $D(f_B)v(f_B)$ exceeds the price increase $D(f_B)(f_S - b_S)$. Therefore, all merchants accept cards if and only if

$$f_S \leq b_S + v(f_B). \tag{4}$$

Rochet and Tirole (2011) show that (4) also holds for a variety of other merchant competition setups, including monopoly and Hotelling-Lerner-Salop differentiated products competition with any number of retailers. Wright (2010) shows the same condition holds for Cournot competition.

Acquirers

We assume acquirers incur per-transaction cost c_S and are perfectly competitive. Thus, given an interchange fee a, they charge a merchant discount f_S such that

$$f_S = a + c_S. \tag{5}$$

Because acquirers are competitive, they play no role in our analysis except passing through the interchange charge to merchants.

Issuers

Issuers are assumed to have market power.[29] We consider a symmetric oligopolistic equilibrium at which all issuers charge the same

[29] This is a standard assumption in the literature. As pointed out in Rochet and Tirole (2002), the issuer market power may be due to marketing strategies, search costs, reputation, or the nature of the card. Note that were the issuing side perfectly competitive, issuers and card networks would have no preference over the interchange fee, and so the latter would be indeterminate.

consumer fee f_B, which can be negative if the cardholder receives a reward. Issuers incur a per-transaction cost c_B and receive an interchange payment of a for a card transaction. At equilibrium, the net per-transaction cost for issuers is $c_B - a$. For simplicity, we consider that issuers set a constant markup φ.[30] Hence, the consumer fee f_B is determined as

$$f_B = \varphi + c_B - a. \tag{6}$$

Network

We consider a monopoly network, which sets the interchange fee a to maximize the total profit of issuers from card transactions, namely,

$$\Pi = \varphi D(f_B) = \varphi[1 - H(f_B)] \ .$$

Alternatively, we could consider a regulator who instead sets the interchange fee to maximize social welfare or user surplus.

Timing

The timing of events is as follows.

1. The card network (or the regulator) sets the interchange fee a.
2. Issuers and acquirers set fees f_B and f_S. Merchants then decide whether to accept cards and set retail prices.
3. Consumers observe the retail prices and whether cards are accepted, and choose a store. Once in the store, the consumer receives her draw of b_B and decides which payment method to use.

Model Characterization

We first consider the market equilibrium under a monopoly network. Given the model setup, the network solves the following problem:

$$\max_a \varphi[1 - H(f_B)] \tag{7}$$

$$s.t. \quad f_B = \varphi + c_B - a, \tag{8}$$

[30] This is a simplifying assumption, and the findings of the model hold if we instead consider an endogenous issuer markup. See Wang (forthcoming).

$$a \leq b_S + v(f_B) - c_S, \tag{9}$$

where the condition (9) is derived from equations (4) and (5).

Since the issuers' profit (7) is maximized by setting the highest possible merchant fee at which merchants still accept cards, the condition (9) holds with an equality. Therefore, the profit-maximizing interchange fee is determined as

$$a^m = b_S + v(f_B^m) - c_S, \tag{10}$$

where f_B^m solves

$$b_S + v(f_B^m) - c_S = \varphi + c_B - f_B^m.$$

Here the superscript m denotes market-determined rates.

This simple model helps illustrate the impact of an interchange cap regulation as we found (or expect to find) in empirical evidence. According to the model, if a regulation pushes down the interchange fee to a level a^r, where $a^r < a^m$, we would have the results as follows.

Result 1 *If a regulation pushes down the interchange fee below the market-determined rate such that $a^r < a^m$, the model implies that (i) consumer card fee increases; (ii) merchant retail price decreases; (iii) card usage falls; and (iv) issuers' profit declines.*[31]

Proof. (i) Conditions (8) and (9) imply that consumer card fee f_B increases as the interchange fee a decreases; (ii) according to (3), merchant retail price p depends on $D(f_B)$ and f_S, both increasing in a; (iii) card usage $1 - H(f_B)$ decreases in f_B; and (iv) issuer profits $\varphi[1 - H(f_B)]$ decrease in f_B. ■

We now turn to the welfare discussion. We first consider that the card network is run by a regulator who maximizes social welfare. Social welfare is generated if consumers use cards for payment at retailers whenever consumer and merchant joint transaction benefits exceed the joint cost of doing so, namely $b_S + b_B > c_B + c_S$. It can be shown that social welfare is the sum of issuers' profit, consumer surplus, and merchants' profit. Accordingly, the regulator solves the problem

$$\max_{f_B} \int_{f_B}^{\bar{b}_B} (b_S + b_B - c_B - c_S) dH(b_B). \tag{11}$$

The first-order condition with regard to f_B requires that

$$f_B^w = c_B + c_S - b_S,$$

[31] In theory, an interchange fee cap can be set too low so that the card market shuts down. For example, for a distribution H with a finite support, consumer fee f_B can become so high that $1 - H(f_B) = 0$.

which implies that the welfare-maximizing interchange fee is

$$a^w = b_S - c_S + \varphi. \tag{12}$$

Here the superscript w denotes welfare-maximizing rates.

Comparing (10) and (12), we have the following findings.

Result 2 *(i) When $\varphi < v(f_B^m)$, the market-determined interchange fee a^m is higher than the welfare-maximizing interchange fee a^w; (ii) when $\varphi \geq v(f_B^m)$, the market-determined interchange fee a^m coincides with the welfare-maximizing interchange fee a^w.*

Proof. (i) Equations (10) and (12) suggest that $a^w = a^m - v(f_B^m) + \varphi$. Therefore, $a^m > a^w$ when $\varphi < v(f_B^m)$. (ii) When $\varphi \geq v(f_B^m)$, we have $a^w \geq a^m$. Because a^m is the highest interchange fee that merchants can accept, a^m then coincides with the welfare-maximizing interchange fee a^w. ■

Similarly, we can consider the card network run by a regulator who maximizes user surplus. Note that user surplus is the sum of consumer surplus and merchants' profit (but not issuers' profit). In the case of competitive merchants, merchants earn zero profit so user surplus equals consumer surplus. Accordingly, the regulator solves the following problem:

$$\max_{f_B} \int_{f_B}^{\bar{b}_B} (b_S + b_B - f_B - f_S) dH(b_B). \tag{13}$$

Recall (5) and (6), which imply that $f_B + f_S = c_B + c_S + \varphi$. Maximizing the user surplus (13) then requires

$$f_B^u = c_B + c_S + \varphi - b_S, \tag{14}$$

which implies that the user-surplus-maximizing interchange fee is

$$a^u = b_S - c_S. \tag{15}$$

Here, the superscript u denotes user-surplus-maximizing rates.

Comparing (10), (12), and (15), we have the following findings.

Result 3 *(i) The interchange fee a^u maximizing the user surplus is lower than the welfare-maximizing interchange fee a^w; (ii) a^u is also lower than the market-determined interchange fee a^m.*

Proof. (i) Equations (12) and (15) suggest that $a^u = a^w - \varphi$, so $a^u < a^w$. (ii) Equations (10) and (15) suggest that $a^u = a^m - v(f_B^m)$, so $a^u < a^m$. ■

Results 2 and 3 show that the market-determined interchange fee tends to be too high, based on the criterion of either social welfare maximization or user surplus maximization. The reason is that under price coherence, merchants internalize consumers' expected card usage benefits when they decide whether to accept cards and set retail prices. This allows the card network to charge too high an interchange fee and too low a consumer fee. As a result, cards get used even when consumer and merchant joint card usage costs exceed their joint transaction benefits. Therefore, regulating down the interchange fee may potentially improve payments efficiency.

However, (12) and (15) also clarify that the socially optimal interchange fee is not determined by the issuer cost, c_B, but rather by the merchant transaction benefit of accepting cards, b_S. Particularly, (15) suggests that a regulator may consider setting the merchant discount $f_S = b_S$, at which the resulting interchange fee maximizes the user surplus. This is the criterion proposed by Rochet and Tirole (2011), which they call the "merchant avoided-cost test."[32]

Small-Ticket Effect

Our analysis so far does not explain the regulation's unintended consequence on small-ticket merchants. This is largely because we have only assumed homogenous merchants in the model. However, even if in a model with multiple (heterogenous) merchant sectors, as long as those merchant sectors are independent from one another in terms of card acceptance and usage, it is still a puzzle to think why card networks would abandon the interchange differentiation in response to a cap regulation. In other words, if it was profitable for a card network to charge a lower fee to small-ticket merchants in the absence of regulation, why would the card network want to change the practice because of a non-binding cap? To address this issue, Wang (forthcoming) extends the model of Rochet and Tirole (2011) by considering card demand externalities across merchant sectors.

In the setup of Wang (forthcoming), there are multiple merchant sectors (e.g., large-ticket merchants and small-ticket merchants). Different merchant sectors are charged different interchange fees due to their (observable) heterogenous benefits of card acceptance and usage. In addition, consumers' benefits of using cards in a merchant sector are

[32] Focusing on user surplus is legitimate if card issuer profits are not considered or weighed much less by competition authorities. The criterion proposed by Rochet and Tirole (2011) is adopted by the European Commission and renamed the "merchant indifference test," while some other countries, including the United States and Australia, adopt the issuer cost-based cap regulation.

positively affected by their card usage in other sectors, which is called the "ubiquity externalities."[33] Based on this setup, Wang (forthcoming) again finds that market-determined interchange fees tend to exceed the socially optimal level. The reason is similar to before: Under price coherence, consumers are provided with excessive incentives to use cards. In addition, Wang (forthcoming) offers the following new findings.

Result 4 *(i) Card demand externalities across merchant sectors explain why card networks eliminate the interchange fee discount to small-ticket merchants in response to the interchange cap regulation; (ii) the social planner who maximizes social welfare would set a discounted interchange fee for small-ticket merchants; (iii) capping the weighted average interchange fee, instead of the maximum interchange fee, may restore the social optimum and avoid the unintended consequence on small-ticket merchants.*

Wang (forthcoming) offers a formal derivation of the above results. Here we provide an intuitive discussion. First, the "ubiquity" externalities may explain card networks' pricing response to the cap regulation: Before the regulation, card networks offer a discounted interchange fee (i.e., a subsidy) to small-ticket merchants because their card acceptance boosts consumers' card usage for large-ticket purchases from which card issuers can collect higher interchange fees. After the regulation, however, the interchange fees on large-ticket purchases are capped. As a result, card issuers profit less from this kind of externality so card networks discontinued the discount.

Second, despite privately determined interchange fees tending to exceed the socially optimal level, the social planner who maximizes social welfare would behave similar to the private network by setting differentiated interchange fees, i.e., charging a high interchange fee to large-ticket merchants but a low interchange fee to small-ticket merchants. Essentially, both the social planner and the private network treat the small-ticket transactions as a loss leader. By subsidizing small-ticket transactions, they internalize the positive externalities of card usage between the small-ticket and large-ticket sectors.

Third, it is possible to design a cap regulation that may restore the social optimum and avoid the unintended consequence on small-ticket merchants. Conceptually, this can be done by imposing a cap on the weighted average interchange fee instead of the maximum interchange

[33] Ubiquity has always been a top selling point for brand cards. This is clearly shown in card networks' advertising campaigns, such as Visa's "It is *everywhere* you want to be," and MasterCard's "There are some things money can't buy. For *everything* else, there's MasterCard."

fee. This alternative regulation provides card networks with incentives to continue charging differentiated interchange fees to different merchant sectors. Note that, under such a cap, a card network can either set an uniform interchange fee limited by the cap, or they could set an above-cap (respectively, below-cap) interchange fee to large-ticket (respectively, small-ticket) transactions as long as the weighted average fee does not exceed the cap. When the cap and weights are appropriately chosen, profit-driven card networks are induced to set differentiated interchange fees at the socially optimal level.

4. CONCLUSION

The recent debit card regulation introduced by the Durbin Amendment to the Dodd-Frank Act has generated significant impact on the U.S. payments industry. In this article, we provide a review of the first-year experience of the regulation.

We first investigate the regulation's empirical impact on different players in the debit card market. We find that the regulation has substantially reduced interchange revenues of large issuers who are covered by the regulation, while small issuers who are exempt have been shielded well so far. We also find that merchants are affected unevenly by the regulation. While merchants as a whole have benefited from the reduced interchange rates, merchants specializing in small-ticket transactions have been adversely affected.

We then provide a theoretical framework to assess the regulation's implications on payments efficiency. We show that market-determined interchange fees tend to be too high compared with the social optimum, so regulating down interchange fees could be welfare enhancing. However, the regulation based on issuer cost is less consistent with theoretical foundation. Rather, policymakers may consider capping interchange fees based on the merchant transaction benefit of accepting cards. Moreover, we discuss that capping the weighted average interchange fee, instead of the maximum interchange fee, may avoid the unintended consequence on small-ticket merchants.

Many issues remain to be addressed for improving the efficiency of the U.S. card payments system. First of all, in order to assess the pricing and performance of payment card markets, policymakers need a good measurement of the costs and benefits of different payment means. These include both private costs and benefits as well as social costs and benefits. Second, policymakers may want to consider policy options other than interchange fee regulation. For instance, in theory, if merchants can set different retail prices conditioning on payment means (e.g., surcharging card usage), the interchange fee becomes less of an

issue. However, those policy options may also have their own limitations, so some cautions need to be taken.[34] Finally and more broadly, we need a better understanding of the functioning of payment card markets, especially the complicated issues regarding the two-sided market nature, the network externalities, and the cooperation and competition between payment platforms.

REFERENCES

Armstrong, Mark. 2006. "Competition in Two-Sided Markets." *RAND Journal of Economics* 37 (Autumn): 668–91.

Baxter, William. 1983. "Bank Interchange of Transactional Paper: Legal Perspectives." *Journal of Law and Economics* 26: 541–88.

Bradford, Terri, and Fumiko Hayashi. 2008. "Developments in Interchange Fees in the United States and Abroad." Federal Reserve Bank of Kansas City *Payments System Research Briefings* (May).

Hayashi, Fumiko. 2012. "Discounts and Surcharges: Implications for Consumer Payment Choice." Federal Reserve Bank of Kansas City *Payments System Research Briefings* (June).

Hung, Christian. 2009. "An Update of Interchange Legislation in the United States." Federal Reserve Bank of Kansas City *Payments System Research Briefings*.

Levitin, Adam. 2010. "Interchange Regulation: Implications for Credit Unions." Filene Research Institute Research Brief 224 (November).

McAndrews, James, and Zhu Wang. 2012. "The Economics of Two-Sided Payment Card Markets: Pricing, Adoption and Usage." Federal Reserve Bank of Richmond Working Paper 12-06 (September).

[34] For example, in countries where card surcharging is allowed, few merchants choose to do so. Moreover, for some merchants who are indeed surcharging, they are found surcharging more than card acceptance costs or imposing surcharging in nontransparent ways. See Hayashi (2012).

Prager, Robin A., Mark D. Manuszak, Elizabeth K. Kizer, and Ron Borzekowski. 2009. "Interchange Fees and Payment Card Networks: Economics, Industry Developments, and Policy Issues." *Finance and Economics Discussion Series* 2009-23. Washington: Board of Governors of the Federal Reserve System (May).

Rochet, Jean-Charles, and Jean Tirole. 2002. "Cooperation among Competitors: Some Economics of Payment Card Associations." *RAND Journal of Economics* 33 (Winter): 549–70.

Rochet, Jean-Charles, and Jean Tirole. 2006. "Two-Sided Markets: A Progress Report." *RAND Journal of Economics* 35 (Autumn): 645–67.

Rochet, Jean-Charles, and Jean Tirole. 2011. "Must-Take Cards: Merchant Discounts and Avoided Costs." *Journal of the European Economic Association* 9 (3): 462–95.

Rysman, Marc. 2007. "An Empirical Analysis of Payment Card Usage." *Journal of Industrial Economics* 55 (1): 1–36.

Rysman, Marc. 2009. "The Economics of Two-Sided Markets." *Journal of Economic Perspectives* 23 (Summer): 125–43.

Schmalensee, Richard. 2002. "Payment Systems and Interchange Fees." *Journal of Industrial Economics* 50 (June): 103–22.

Shy, Oz. 2012. "Who Gains and Who Loses from the 2011 Debit Card Interchange Fee Reform?" Federal Reserve Bank of Boston Public Policy Discussion Paper 12-6 (June).

Shy, Oz, and Zhu Wang. 2011. "Why Do Payment Card Networks Charge Proportional Fees?" *American Economic Review* 101 (June): 1,575–90.

Wang, Zhu. 2010. "Market Structure and Payment Card Pricing: What Drives the Interchange?" *International Journal of Industrial Organization* 28 (January): 86–98.

Wang, Zhu. Forthcoming. "Demand Externalities and Price Cap Regulation: Learning from A Two-Sided Market." Federal Reserve Bank of Richmond Working Paper.

Weyl, Glen. 2010. "A Price Theory of Multi-Sided Platforms." *American Economic Review* 100 (September): 1,642–72.

Wright, Julian. 2003. "Optimal Card Payment Systems." *European Economic Review* 47 (August): 587–612.

Wright, Julian. 2004. "Determinants of Optimal Interchange Fees in Payment Systems." *Journal of Industrial Economics* 52 (March): 1–26.

Economic Quarterly—Volume 98, Number 3—Third Quarter 2012—Pages 185–207

Housing Services Price Inflation

Marianna Kudlyak

The cost of housing services constitutes more than 30 percent of the cost of the consumer basket used to measure the consumer price index (hereafter, CPI), a major indicator of inflation in the consumer prices produced by the Bureau of Labor Statistics (BLS). Thus, understanding housing services price inflation is important for understanding the aggregate fluctuations of prices in the economy.

In this article, we provide an explanation of how inflation of the price of housing services is measured by the BLS and describe alternative approaches. We then describe the contribution of inflation of the price of housing services to inflation in the CPI during the Great Recession and its aftermath.[1] Finally, we examine new data series that provide additional information about the rental market for housing services and use this information to evaluate the direction of the pressure on housing services price inflation (hereafter, housing services inflation).

Between 2005 and 2007, housing services inflation, as measured by the CPI, was rising, while house price inflation exhibited a steep decline. Such periods, i.e., when the CPI measure of housing services inflation diverges particularly far from house price inflation, often reignite the debate about whether the CPI adequately reflects the cost of housing services.

This debate fails to recognize that the CPI program measures the price of the services provided by housing and not the price of the asset (i.e., house) itself. If the household buys the housing services in

The author is grateful to Andreas Hornstein, Robert Hetzel, Zhu Wang, and Jonathan Tompkins for their generous comments and suggestions. Steven Sabol provided excellent research assistance. The views expressed here are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: marianna.kudlyak@rich.frb.org.

[1] In the analysis, we use data up through the second quarter of 2012.

the market, i.e., rents an apartment, then the rental price is the price of the services. If the household owns the housing unit that provides housing services, then the price of the flow of housing services that the household receives must be imputed because the price is not observed. Given that a majority of U.S. households own their housing, the imputation procedure is one of the main issues associated with calculating the CPI. The measure of the hypothetical rent paid by homeowners is the major component of the CPI and is called the owner's equivalent rent (OER).

This article argues that the changes in the price of housing services should not necessarily move with the changes of house prices. In particular, currently, the BLS calculates the owner's equivalent rent using a rental-equivalence approach, in which only data on rental prices are collected. Under this approach, the house prices are reflected in the CPI to the extent that they are reflected in the current rent in the ongoing rent contracts. An alternative imputation mechanism for the owner's equivalent rent is the user cost approach. The user cost approach is arguably more attractive conceptually because it explicitly treats a house as an asset. The user cost approach shows directly that the cost of housing services depends not only on the contemporaneous house prices but also on their expected change. Despite being conceptually more attractive, the user cost approach has proven hard to implement in practice.

Currently, the monthly CPI housing services inflation is measured by a repeat-rent index, which represents the monthly average of the change in the rental price of rental units over the last six months. Recently, new data on the rental housing market, which reflect month-to-month changes, have become available. Examining the series that describe month-to-month changes can help gauge the direction of changes of the CPI housing services inflation index in upcoming months. We examine the behavior of the new series on residential rents, rental vacancies, and rent concessions. The developments in the rental housing market suggest that since 2010 there has been increasing upward pressure on housing services inflation.

The remainder of the article is organized as follows. The next section describes the measurement of housing services price inflation. Section 2 summarizes the recent behavior of housing services price inflation as measured by the BLS. Section 3 examines new additional series that describe the rental housing market. Section 4 concludes.

1. ACCOUNTING FOR HOUSING SERVICES PRICE INFLATION

Current Accounting for Housing in the CPI

The CPI is a cost of living index, that is, the cost of generating a certain level of consumption for a certain time period, usually a month. The construction of the CPI views housing units as capital goods rather than as consumption items. The relevant consumption item for the CPI is shelter—the service that the housing unit provides. The CPI Shelter constitutes the major part of the CPI.

The CPI Shelter represents a weighted average of the four component indexes: (1) rent of primary residence (CPI Rent), (2) owners' equivalent rent of primary residence (CPI OER), (3) lodging away from home, and (4) tenants' and household insurance. Residential rents and OER data are collected from the CPI Housing Survey. The other two components, lodging away from home and tenants' and household insurance, are obtained from the CPI Commodities and Services Survey.

The CPI program calculates the price of the housing services of the owner-occupied housing using the rental equivalence approach. Under this approach, the cost of the shelter services provided by owner-occupied housing is the implicit rent (i.e., the amount the owner would pay for rent or would earn from renting his home in a competitive market) that is imputed from the actual rental prices collected from renters. The BLS employs the re-weighting method to the rental equivalence approach of calculating the hypothetical rents paid by homeowners. Under this method, the owners' equivalent of rent is calculated by re-weighting the rent sample to represent owner-occupied units.

Essentially, the CPI Rent and the CPI OER are the repeat-rent indexes, the information for which is collected from rental units. The idea behind the index is to obtain the price change between period t and period $t+1$ for the same rental unit, and then aggregate these price changes. The rent information in period t and in $t+1$ is collected from the same unit to ensure that recoded change in rent is because of inflation rather than the quality difference between t and $t+1$. The quality difference is an issue because it is conceivable that in the case with housing, rental or owner-occupied, there are large unmeasured differences in the quality. Each rental unit is surveyed every six months. Thus, the CPI Rent and the CPI OER define the month-to-month change in the price of housing services as the average monthly price change over the last half year. The Appendix contains details on (1)

how the data on rental prices are collected, and (2) how the data are used to construct the CPI Rent and the CPI OER.[2]

For cost efficiency, each rental unit is surveyed every six months. The CPI Rent is a weighted average of the change in the same-unit rents where the weights reflect the quality distribution of rental units. The CPI OER is a weighted average of the same rent changes (minus the cost of utilities if they are included in the rent) where the weights reflect the OER characteristics in the sample. The CPI Rent and the CPI OER define the month-to-month change in the price of housing services as the average monthly price change over the last half year.

A few additional notes are in order. First, for segments that contain largely owner-occupied housing, the CPI program selects rental units from the nearby segments. Second, for the vacant rental units, the estimated current rent is its previous rent times the average rent change of newly occupied units. Third, some rental units represent only rental units (for example, rental units under rent control), while other rental units represent only owner-occupied units. The CPI program's handling of the rental units under rent control and the differences between economic and pure rent contribute to the differences between OER and Rent indexes.

As described above, the existing CPI approach to accounting for owner-occupied housing services simply re-weights the rent sample to represent owner-occupied units. Prior to 1999, the BLS employed the matching method to the rental equivalence approach (Diewert and Nakamura 2009). Under this method, information is collected from both renter and owner samples. Then, the owner's unit is matched with a renter's unit with similar characteristics (i.e., location, structure type, age, number of rooms, type of air conditioning, and other attributes). The change in implicit rent is derived from the change in the pure rents of its matched set of renters. However, this method requires large cost associated with collecting data from both renters and owners and is no longer used.

We can identify two main problems associated with the current accounting for housing in the CPI. First, most rental contracts are long-term, and rents are sticky in the ongoing contracts. There is also considerable evidence that the rents are sticky not only within the contracts but also within the entire tenure of a renter with a particular

[2] In this section we largely follow the BLS description of the measurement of CPI inflation (see Bureau of Labor Statistics [2007, 2009)]). See Diewert and Nakamura (2009); Diewert, Nakamura, and Nakamura (2009); and Crone, Nakamura, and Voith (2010) for a description of the current measurement approach. Wolman (2011) provides an alternative inflation measure that uses a different aggregation procedure for the existing CPI components.

landlord (for example, Genesove [2003]). Thus, houses cannot likely be rented at the same price as the rental units in ongoing rent contracts. Consequently, the rents in newly signed leases, which reflect the contemporaneous house prices and rental vacancies, might better reflect the implicit rent of owner-occupied housing. Second, rental housing might not be that close a substitute for owner-occupied housing.[3] An alternative approach to calculating the rental price of owner-occupied housing, the user cost approach, explicitly recognizes that a house is a capital good and addresses some of these concerns. We discuss the user cost approach next.

User Cost Approach

The user cost approach to owner-occupied housing treats the services provided by owner-occupied dwellings differently from the services provided by rental dwellings. The user cost of housing services can be thought of as a cost to a household of purchasing a house at the beginning of the period, living in it during the period, and then selling it at the end of the period at the prevailing market price.

Kudlyak (2009) uses a similar approach to measure the firm's labor cost. Since employment relationships often last for more than one period, wage usually does not represent the period's labor cost but rather it is an installment payment on an employment contract. Kudlyak empirically constructs the user cost of labor, which is the difference between the present discounted value of wages to be paid to a worker hired in the current period and the expected present discounted value of wages to be paid to a worker hired the next period. Importantly, she finds that the user cost of labor is much more procyclical than the average wage or the wage of newly hired workers in the economy because of the effect the economic conditions at the time of hiring have on future wages within the employment relationship.

To introduce the user cost, let V_t^v denote the purchase price of a v-year durable in year t, u_t^v denote the end-of-period value of the period t services provided by this durable, O_t^v denote the operating expenses, and r_t denote the nominal interest rate. Assuming, in equilibrium, the purchase price of a durable equals the expected present discounted value of its net benefits yields the following expression for the expected user cost of housing services in period t, $E_t u_t^v$,

$$E_t u_t^v = r_t V_t^v + E_t O_t^v - (E_t V_{t+1}^{v+1} - V_t^v). \tag{1}$$

[3] Prescott (1997) provides a good description of the problems associated with defining real consumption from owner-occupied housing and medical insurance.

Equation (1) states that the expected user cost in period t equals the foregone interest rate payments, $r_t V_t^v$, the expected operating costs (maintenance plus property taxes), $E_t O_t^v$, and the expected change in the house price, $E_t V_{t+1}^{v+1} - V_t^v$, where the superscript on V_t^v takes into account depreciation. In a frictionless equilibrium with risk-neutral landlords and no transaction costs, the user cost of housing equals the rent.

An early theoretical application of the user cost approach to the measurement of the price of services of owner-occupied housing is found in Dougherty and Van Order (1982), and recent estimates of the user cost are provided by Garner and Verbrugge (2007) and Verbrugge (2008). Verbrugge (2008) calculates a one-year user cost as follows:

$$E_t u_t = P_t \left(r_t + \gamma - E_t \pi_t \right), \tag{2}$$

where P_t is the price of the house; r_t is the nominal interest rate; γ is the sum of depreciation, maintenance and repair, insurance, and property taxes (all assumed constant); and π_t is the four-quarter constant-quality home price appreciation between year t and year $t+1$.

Rewriting equation (2) shows that the change in the user cost is a function of the change in the house prices and the change in the second term, $(r_t + \gamma - E_t \pi_t)$, i.e.,

$$d \ln E_t u_t = d \ln P_t + d \ln \left(r_t + \gamma - E_t \pi_t \right). \tag{3}$$

The change in the second term, $(r_t + \gamma - E_t \pi_t)$, is governed by the movements in $(r_t - E_t \pi_t)$, which can be thought of as the real interest rate, and is less volatile the larger is the fixed cost, γ. Thus, unless expected house price changes move in sync with nominal interest rates, i.e., $d \ln \left(r_t + \gamma - E_t \pi_t \right) = 0$, the user cost, $d \ln E_t u_t$, is more volatile than house prices, $d \ln P_t$.

To calculate the user cost, Verbrugge (2008) obtains information on the current market value of the house from the Consumer Expenditure Survey. Then, he estimates the expected price change, $E_t \pi_t$, using four-quarters-ahead forecasts from the regional house price indexes. Because the period under study is characterized by a substantial house price appreciation, the second term in equation (2), $\left(r_t + \gamma - E_t \pi_t \right)$, can be negative. Thus, whenever the estimated $E_t \pi_t$ delivers negative $E_t u_t$, Verbrugge sets $E_t u_t$ to 0.

Garner and Verbrugge (2007, Figure 1) show Verbrugge's user cost series (logarithm of the levels) and the two rental series, the official CPI Rent Index, and the series constructed by Verbrugge (2008) that tracks only rental units comparable to those used in the house price indexes (i.e., detached properties) from 1980–2005. Their figure shows that there is little evidence that the user costs and rents are equivalent measures. In fact, the user costs do not exhibit a positive trend

observed in rents. After 1997, the rent series are higher than the user cost series; this suggests that owning is cheaper than renting and can explain the increase in the homeownership rates during that period. However, it also suggests the presence of non-exploited arbitrage or large transaction costs of converting owner units into rentals.

The fact that house prices were rising steadily over the period up to 2005 while the user cost shows no such trend suggests that the movements in the user cost were dominated by the movements in the second term in equation (2). As Garner and Verbrugge (2007) note, expected house price appreciation is responsible for user cost not tracking the rise in house prices. Importantly, Verbrugge (2008) notes that if instead of the forecast house price changes, $\widehat{E_t\pi_t}$, the expected CPI inflation is used, then the user cost measure is much closer to the rent index measure. Poole, Ptacek, and Verbrugge (2005) revisit the user cost approach to examine whether the user cost can reflect the rapidly rising house prices in 2005. They conclude that the user cost approach would not mirror the increase in house prices.

The literature lists the following factors that can explain possible divergence of the user costs and rents: (i) rent stickiness during the tenant's tenure with the landlord, even beyond one-year rent contracts; (ii) the thinnest of the rental market for luxury homes; and (iii) the differential tax treatments. For example, Diaz and Luengo-Prado (2008) show that a rental equivalence approach, as compared to a user cost approach, overestimates the cost of shelter services provided by owner-occupied housing because owner-occupied housing services are not taxed and mortgage interest payments are deductible.

The Bureau of Economic Analysis and the BLS attempted to develop the user cost approach in the 1980s. However, these attempts were abandoned because the researchers concluded that it was impossible to estimate the user cost without directly or indirectly using the rent information (Gillingham [1980]; see a discussion in Diewert and Nakamura [2009]). Summarizing, despite the fact that the user cost approach is (arguably) conceptually more attractive for the measurement of the price of the flow of services provided by an asset, the approach has proved hard to implement in practice.

One way to modify the expression for the user cost is to recognize that the owners usually have a mortgage on the house and distinguish between the return on equity and the mortgage interest rate in equation (1). Early implementations of the mortgage payments in the price of the housing services provided by owner-occupied housing are studied by Kearl (1979) and Gillingham (1980).

Diewert and Nakamura (2009) incorporate debt into an alternative approach that explicitly takes into account the financing of the house

purchase, which they refer to as the opportunity cost approach. They seek to compare the implications for homeowner wealth of selling the property at the beginning of a period with an alternative of planning to keep the house for m more years and then either renting or occupying for the coming year. The opportunity cost is defined as the greater of the rental opportunity cost (which is an implicit rent) and the "financial opportunity cost." Thus, there is never an issue of running into a negative financial opportunity cost.

Diewert and Nakamura specify the financial user cost of owning a home in period t as follows (abstracting from depreciation):

$$E_t u_t = r_t^D D_t + r_t \left(V_t - D_t\right) + E_t O_t^v - \left(E_t \overline{V_{t+1}} - V_t\right), \tag{4}$$

where D_t is a debt owned on the house, i.e., $V_t - D_t$ is the value of equity in the house, which is assumed to be nonnegative; $\overline{V_{t+1}}$ is the value of the home at the beginning of period $t+1$ plus the expected average appreciation of the home value over the number of years before the owner plans to sell; and r_t^D is the nominal interest on the debt owned. Note that if $r_t^D = r_t$, i.e., if the homeowners who have mortgages on their homes are charged an interest rate on their debt that equals the rate of return on their financial investments, then equation (4) reduces to the usual expression for the user cost (equation [1]) (except for the details on the definition of the $E_t\overline{V_{t+1}}$ term). Examining equation (4) shows that for a homeowner with low-cost borrowing, i.e., $r_t^D < r_t$, the user cost of owning is lower than that for a homeowner with high-cost borrowing, i.e., $r_t^D > r_t$. The financial opportunity cost component of Diewert and Nakamura can be thought of as the user cost approach with debt. To our knowledge, this version of the user cost has not been implemented empirically.

Diewert and Nakamura (2009) provide an insightful review of alternative approaches to the accounting for housing in a consumer price index. In particular, they describe an acquisitions approach and a payment approach. Under the acquisitions approach, the entire cost of a purchase of the house is charged to the period. The objective of the approach is to measure the average change in the price of a product irrespective of whether the product is fully used in the period or fully paid in the period. However, only the goods that the household sector purchases from other sectors are included. Thus, the housing-related expenditures that enter a CPI are mostly expenditures on new dwellings, while the secondhand dwellings and land are excluded. The payments approach only measures actual cash outflows associated with the owner-occupied housing: cost of repairs, maintenance, house insurance, local authority charges, and mortgage interest.

Table 1 CPI-U: City-Average Expenditure Category Relative Importance

Expenditure Category and Items	Expenditure Share, March 2012
Food and Beverages	15.11
Housing	40.59
Shelter	31.26
Rent of Primary Residence	6.49
Lodging Away from Home	0.81
Owners' Equivalent Rent of Residences	23.66
Owners' Equivalent Rent of Primary Residence	22.29
Tenants' and Household Insurance	0.34
Fuels and Utilities	5.26
Household Energy	4.10
Water and Sewer and Trash Collection Services	1.16
Household Furnishings and Operations	4.07
Apparel	3.61
Transportation	17.58
Medical Care	7.05
Recreation	6.01
Education and Communication	6.71
Other Goods and Services	3.34

Notes: Category "Other Goods and Services" includes tobacco, smoking products, and personal care.

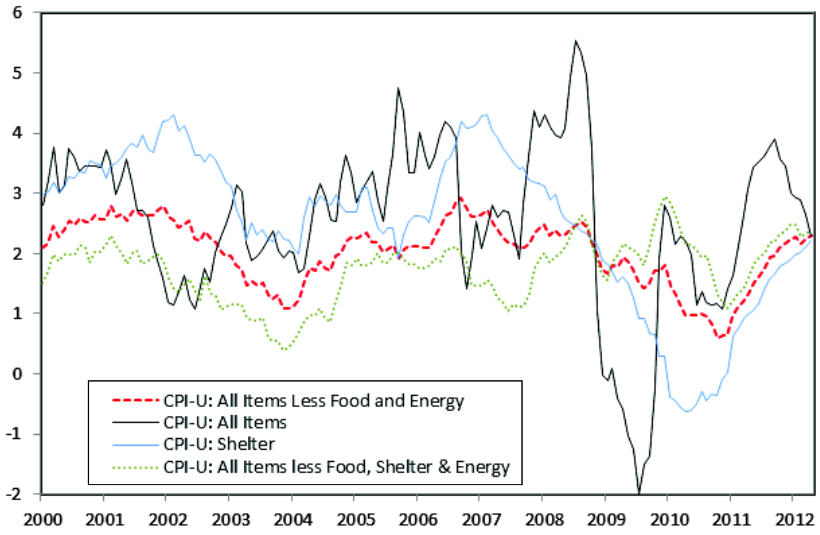
Source: BLS

2. HOUSING SERVICES PRICE INFLATION

CPI Measures of Housing Services Price Inflation and CPI Inflation

Shelter, the service that housing units provide to consumers, constitutes the major part of the consumer market basket, which is used to construct the consumer price index. Table 1 shows that in 2012 households allocated 31.3 percent of their consumption expenditures to shelter. The expenditure shares are the weights by which different component price indexes are aggregated. The CPI Shelter represents a weighted average of the four component indexes: (1) rent of primary residence (6.49 percent of the CPI); (2) owners' equivalent rent of residences (23.66 percent of the CPI, including the owners' equivalent rent of primary residence, which constitutes 22.29 percent of the CPI); (3)

Figure 1 CPI and CPI Shelter Inflation, Percent Year-Over-Year



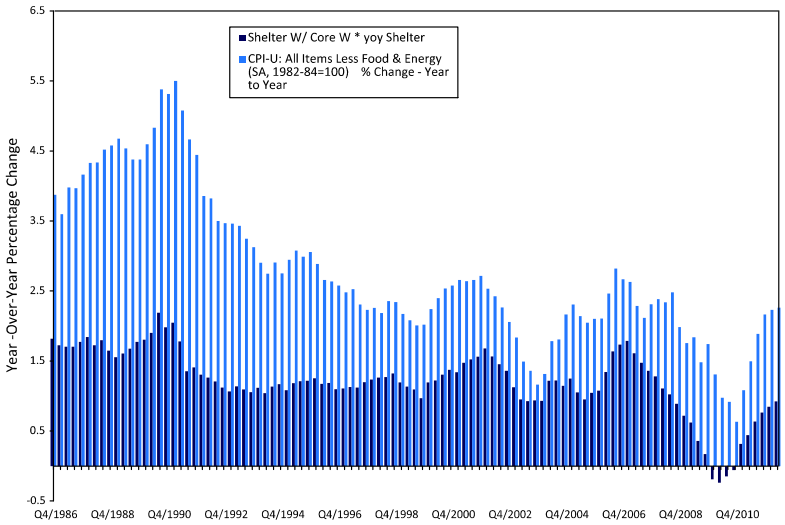
lodging away from home (0.81 percent of the CPI); and (4) tenants' and household insurance (0.34 percent of the CPI).[4]

The expenditure shares are estimated from the data reported by sampled households in the Consumer Expenditure Interview Survey, which includes both renters and homeowners, and is updated approximately every two years. Shelter is part of a larger category, housing, which also includes fuels and utilities and household furnishings and operations. "Housing" constitutes approximately 41 percent of the CPI.

From its recent peak, the first quarter of 2007, to its recent trough, the fourth quarter of 2010, CPI Shelter inflation declined from 4.3 percent to −0.44 percent (monthly, year-over-year). In April 2012, CPI Shelter inflation stood at 2.23 percent. Figure 1 shows inflation in the CPI All Items; the CPI All Items Less Food and Energy; the CPI Less Food, Energy, and Shelter; and the CPI Shelter. During 2001–2008, CPI Shelter inflation was always higher than CPI All Items Less

[4] At the beginning of 2010, the BLS moved the expenditure weight of second homes from "lodging away from home" to a new item, "owners' equivalent rent of residences," which includes secondary and primary residences, and did not revise prior data. The new series "owners' equivalent rent of residences" contain data for second homes only starting in January 2010. The series "lodging away from home" contains data on second homes up to December 2009.

Figure 2 Contribution of CPI Shelter Inflation to Core CPI Inflation, Year-Over-Year



Source: Author's calculations using BLS data.

Food and Energy Inflation (hereafter, core CPI inflation). However, from the fourth quarter of 2008 up until the first quarter of 2012, the situation is reversed: Core CPI inflation exceeds CPI Shelter inflation.

Figure 2 shows the contribution of CPI Shelter inflation to core CPI inflation calculated as a product of the CPI Shelter weight in the core CPI and its year-over-year inflation rate. The figure shows that CPI Shelter inflation contributed 1.38 percent out of 2.63 percent of core CPI inflation in the first quarter of 2007. The contribution proceeded to decline until it became negative in 2010. The contribution of CPI Shelter to core CPI inflation has been steadily increasing since then.

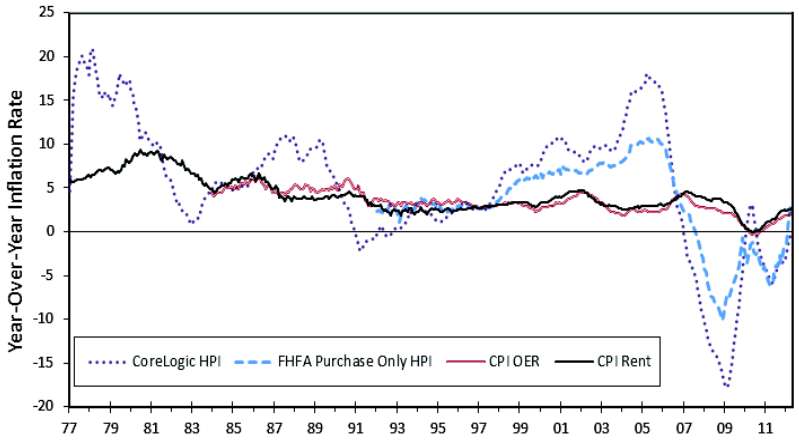
Table 2 shows the change in consumer price index inflation by major expenditure category during the Great Recession, from December 2007 to June 2009, and its aftermath, from June 2009 to April 2012.

Table 2 Change of Inflation During the Great Recession and its Aftermath, by Major Expenditure Category, Percent

Expenditure Category and Items	December 2007–June 2009	June 2009–April 2012
CPI-U: All Items	1.56	6.73
CPI-U: All Items Less Shelter	1.22	8.65
CPI-U: All Items Less Food, Shelter, and Energy	3.23	5.60
Food and Beverages	5.27	7.04
Housing	2.10	2.58
Shelter	2.25	2.78
Rent of Primary Residence	4.38	3.82
Lodging Away from Home	−8.28	7.83
Owners' Equivalent Rent of Residences	2.99	2.68
Owners' Equivalent Rent of Primary Residence	2.99	2.68
Fuels and Utilities	1.00	6.07
Household Energy	−0.74	3.53
Water and Sewer and Trash Collection Services	9.18	16.13
Household Furnishings and Operations	2.26	−2.59
Apparel	0.76	4.49
Transportation	−6.52	20.70
Medical Care	4.53	9.46
Recreation	2.29	−0.07
Education and Communication	4.97	4.73
Other Goods and Services	9.73	5.22

Notes: Author's calculations using BLS data.

Figure 3 CPI Rent Inflation, CPI OER Inflation, and House Price Inflation, Year-Over-Year



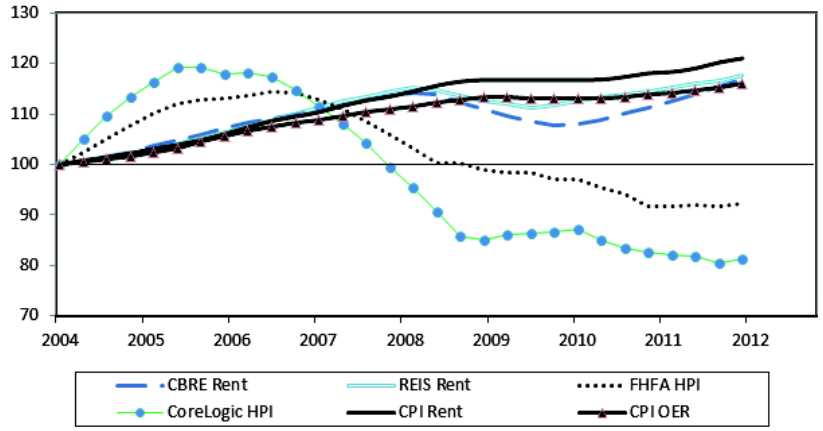
CPI Measures of Housing Services Price Inflation and House Prices

As can be seen from Table 1, the main components of the CPI Shelter are the CPI Rent of Primary Residence (CPI Rent) and the CPI Owners' Equivalent Rent of Primary Residence (CPI OER). Figure 3 shows CPI Rent inflation and CPI OER inflation along with inflation in house prices as measured by the Core Logic house price index and the Federal Housing Finance Agency Purchase Only Index (see Figure 4).

Figure 3 shows that house price inflation fluctuates significantly more than CPI Rent or CPI OER inflation. It is especially evident during 2004–2010. The figure also shows that house price inflation and inflation in the CPI measures of housing often do not move in the same direction. Between 2002 and 2004, house price inflation was rising while inflation in the CPI housing indexes was falling. During 2005–2009, when house price inflation rapidly fell from 15–20 percent in 2005 to −15 to −20 percent in 2009, CPI housing inflation was fluctuating around 4 percent and started decreasing only after 2008.

The periods when the CPI measure of inflation diverges particularly far from house price inflation often reignite a debate about whether the CPI Rent and CPI OER adequately reflect the cost of shelter. As emphasized in Section 1, it is important to recognize that the cost of

Figure 4 CPI Rent, CPI OER, and the Rent and House Price Indexes



Notes: Q1:2004 = 100.

Source: Author's calculations using BLS data.

housing services should not necessarily move with house prices. The CPI program's indexes of housing inflation measure inflation in the prices of housing services rather than inflation in house prices. Given the method that the BLS currently uses to measure the cost of the housings services of owner-occupied units, house prices are reflected in the CPI index to the extent that they are reflected in the current rent in the ongoing rent contracts (via the supply and demand of rental units and the substitution between renting and owning). Alternatively, the user cost approach to measuring the cost of owner-occupied housing shows more directly that the cost of shelter depends both on current house prices and on their expected change.

3. RECENT DEVELOPMENTS IN THE RENTAL HOUSING MARKET

As described above, the current accounting for price of housing services in the CPI almost entirely relies on the data on rental prices from rental units. In addition, the monthly price changes used for calculation of the inflation in the price of housing services is the monthly average of the price change over the last half year. Thus, a direct examination of the

recent developments in the rental market can be useful in gauging the direction of changes of housing services price inflation. Recently, new data series that describe the aggregate rental market became available. In contrast to the CPI housing services price indexes, these series reflect month-to-month changes and, thus, can serve as leading indicators of the changes in rental prices. In this section, we describe the behavior of different indicators of the rental market and the behavior of alternative measures of rent price inflation.

Additional Indicators of the Rental Market

Rent Concessions

One way to gauge the pressure on rent prices is to examine the series of discounts that landlords are willing to extend to renters. Figure 5 shows the difference between the asking rent and the effective rent as a share of asking rent obtained from Reis Inc. The larger the difference, the more concessions a landlord is willing to provide to a renter. The figure shows that the discount is at its lowest level of the last 10 years. It has declined from its peak of 6.3 percent in the second quarter of 2009 to 4.8 percent in the first quarter of 2012. Reis Inc. forecasts a further decline in concessions to 3.23 percent by 2016.

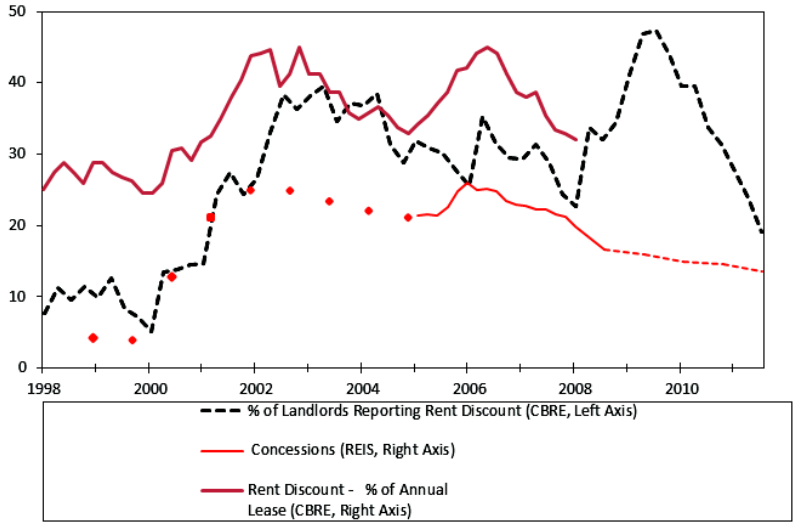
Figure 5 also shows the share of properties offering a discount and the average discount in the annual rent, the series obtained from CB Richard Ellis (hereafter, CBRE). The share of properties offering a discount has declined from approximately 47 percent in the first quarter of 2010 to 19 percent in the first quarter of 2012. The average annual discount has also been declining during this period.

Rental Vacancy Rates and Net Absorption

An alternative way to examine the direction of the pressure on the rent prices is to examine the supply of the properties available for rent. The vacancy rate for renter-occupied housing is defined as the number of vacant units for rent over the stock of vacant and occupied units for rent. Figure 6 shows the vacancy rate series from the Census, CBRE, and Reis Inc. The three series show a decline in the vacancy rates since mid-2009. In particular, Reis data show that the vacancy rate has declined from 8 percent in mid-2009 to 4.9 percent in the first quarter of 2012.

The net absorption, NA_t, as measured by Reis Inc., is the difference between the occupied stock of rental units in the current period, O_t, and in the last period, O_{t-1}, which is the difference between the number of newly signed leases and the number of leases that were terminated

Figure 5 Measures of Rent Discounts



Notes: Concessions are the ratio of the difference between asking rent and effective rent to the asking rent. Markers indicate annual observations.

Source: CBRE and Reis Inc.

and not renewed, $NA_t \equiv O_t - O_{t-1} = NR_t - TR_t$. Figure 7 shows the net absorption as a share of the previous period stock of occupied rental vacancies. As can be seen from the figure, after mid-2008 the net absorption has been positive and increasing since 2011.

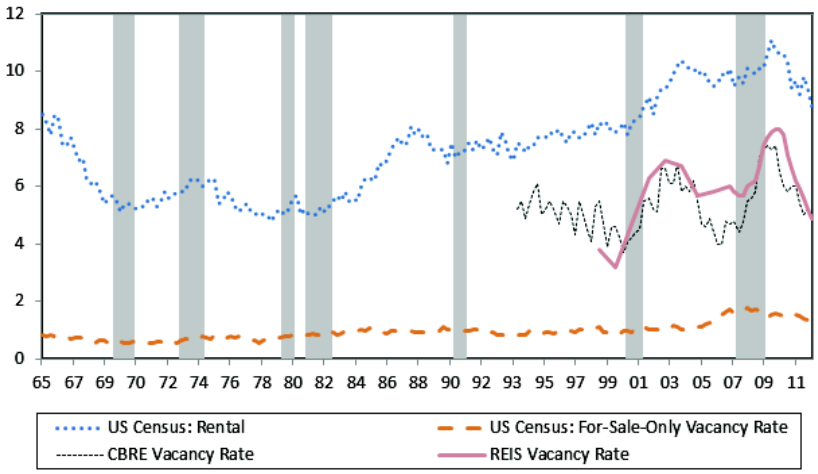
The increase in the net absorptions has been feeding into the recent rapid decline in vacancy rates. To see this, note that the evolution of the number of vacancies, v_t, can be described by the following equation

$$v_t = v_{t-1} + (NCompl_t + NConv_t) - NR_t + TR_t, \tag{5}$$

where $NCompl_t$ is the number of new completions and $NConv_t$ is the number of net conversions into the rental units.

Assuming that the change in the stock of rental properties from $t-1$ to t is negligible as compared to the change in the number of vacancies, equation (5) shows that the decrease in the vacancy rate from $t-1$ to t can be brought by a decrease in net completions, a decrease in net conversions, or by an increase in the net number of newly signed rental contracts, $(NR_t - TR_t)$. Reis Inc. predicts an increase in net completions from 39,400 properties in 2011 to 66,500

Figure 6 Rental Vacancy Rates



Source: Census Bureau, CBRE, and Reis Inc.

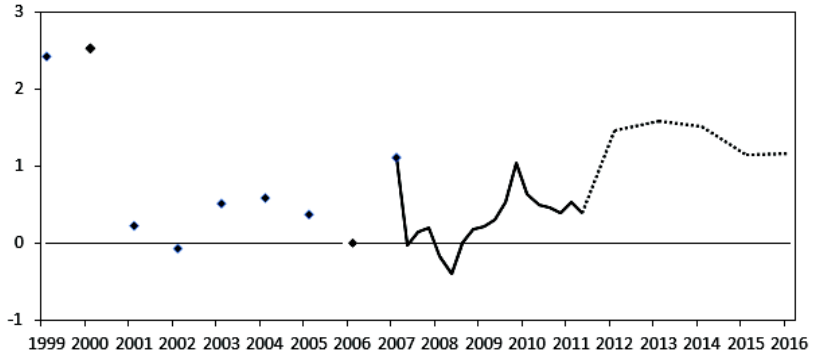
properties in 2012. Given the negligible role of net conversions, the decrease in the vacancy rent is mostly because of an increased demand for rental units.

The series of the rental vacancy rates and the rent concessions suggest that there is an upward pressure on the rent prices.

Alternative Indicators of Rent Price Inflation

There are two alternative rent indexes that measure aggregate rent inflation. The first index is the REIS Rent Index, which is provided by Reis Inc. The second index is the CBRE Rent Index, provided by CBRE. Reis Inc. collects data on the asking rent, Reis Asking rent, and on the effective rent in newly signed leases, Reis Effective rent. The rent data do not include information from the renewed leases and Reis Inc. does not collect information on the rents in ongoing lease contracts. The rent information for the CBRE Rent Index is obtained by asking the managers of the properties about what the rent would be if they were to rent a unit in the current market, regardless of whether the unit is currently occupied or vacant. Thus, the recoded information might be on the rents in ongoing contracts as well as on the

Figure 7 Net Absorptions of Rental Properties



Notes: The figure shows net absorptions as the share of the previous period stock of occupied rental units. Net absorption is the difference between the occupied stock of rental units in the current period and the occupied stock in the previous period. The figure shows annual observations prior to 2007 and quarterly observations thereafter. The dotted line indicates forecast.

Source: Reis Inc.

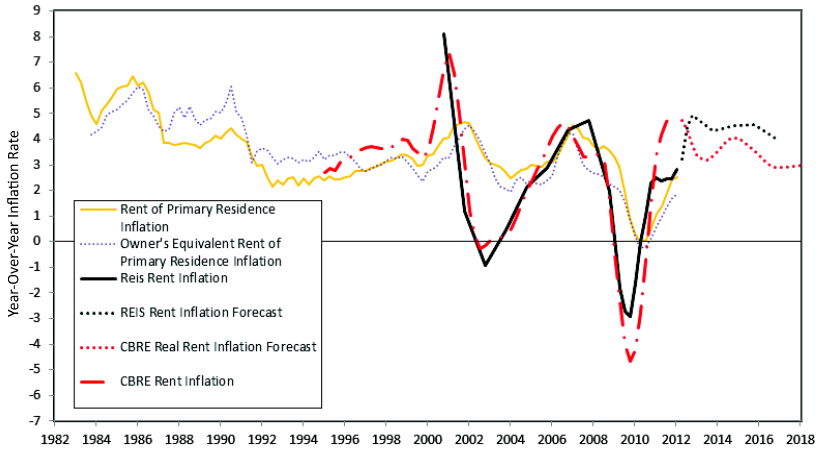
perceived effective new rents. Thus, both indexes contain information about month-to-month changes in rental prices.

The Reis Asking rent and the CBRE Rent Index both provide information on the apartment rents in the multi-housing market, with some differences in the coverage. Data from Reis Inc. cover rental complexes consisting of 40 or more units (except for California metropolitan areas, where complexes of 15 or more units are included). Data for the CBRE Rent Index cover multi-housing properties with five or more units.[5] Housing data from the U.S. Census Bureau has a much wider scope. The Census uses residential properties regardless of rent restrictions and does not have a restriction on the number of rental units. The CPI Rent also includes data on rent-controlled properties.

Figure 8 shows quarterly year-over-year inflation in the CPI Rent, the CPI OER, the Reis Effective rent, and the CBRE Rent. All four inflation series show a decline during the 2001 and 2007–2009 recessions. The figure suggests that Reis Rent Index inflation and CBRE Rent inflation appear to lead the CPI Rent and CPI OER inflation measures.

[5] This information was obtained from CBRE and Reis Inc. representatives in June 2011.

Figure 8 Rent Inflation



Source: BLS, CBRE, and Reis Inc.

A particularly striking feature of Figure 8 is that Reis Rent Index inflation and CBRE Rent inflation experienced a significantly larger drop during 2007–2009 as compared to the CPI inflation measures. Such a discrepancy between Reis Rent Index inflation or CBRE Rent inflation and the CPI housing services inflation can, at least partially, be attributed to the different time reference period of these measures. Recall from Section 1 that the CPI month-to-month housing services price inflation measure essentially represents a monthly average over the past six-month change, while Reis Rent Index inflation and CBRE Rent inflation represent month-to-month changes.

Inflation as measured by the CBRE Rent Index has been increasing from its recent trough of −4.95 percent in the fourth quarter of 2009 to 4.67 percent in the first quarter of 2012. During the same period, inflation as measured by the Reis Rent Index has increased from its trough of −2.92 percent to 2.83 percent in the first quarter of 2012. CPI Rent inflation and CPI OER inflation lagged the other two inflation measures and reached their troughs, at 0 percent and −0.2 percent, respectively, in the second quarter of 2010. CPI Rent inflation stands at 2.5 percent and CPI OER inflation stands at 1.9 percent in the first quarter of 2012.

4. CONCLUSIONS

The CPI is a cost of living index that measures the price of a constant flow of consumption during a period. One of the challenges of accounting for the price of consumption is accounting for the price of housing services. The issue is that a large fraction of the U.S. population owns their housing. The price of housing services for owner-occupied housing is not observed directly and, thus, the price for the hypothetical market transaction involving the housing services of owner-occupied housing must be imputed.

The Bureau of Labor Statistics employs a particular imputation mechanism, the rental equivalence approach, which implies a close substitutability between rental and owner-occupied housing. An alternative, conceptually more attractive approach to accounting for the price of the flow of services provided by an asset (i.e., by a house) is the user cost approach. Despite its conceptual attractiveness, the approach has proven hard to implement in practice.

Currently, the monthly CPI measures of housing services price inflation represent a repeat-rent index, which is calculated as the monthly average of the past six-month change of the rental price of rental units. The newly available data from the rental housing market, which usually reflects month-to-month changes, can be informative about the direction of changes in the CPI measure of housing services inflation. The data on residential rents, rental vacancies, and rent concessions suggest that since 2010 there has been an increasing upward pressure on rent price inflation.

APPENDIX

Below, we describe (1) how the data on rental prices are collected, and (2) how the data are used to construct the CPI Rent and the CPI OER.

The collection of rent information for construction of the CPI Rent and CPI OER is conducted as follows. The CPI program collects price information from 87 urban areas (i.e., index areas). Each of the index areas is divided into six strata, each representative of the area. Within each stratum, the program defines small segments. For each segment, the CPI program collects information on the number of renter- and owner-occupied units, and the average rent of renter units. Based on this information, the program calculates the total spending on shelter for each segment. The total spending on shelter is the sum of (1)

the product of the number of rental units and the average rent in the segment, and (2) the product of the number of owned units and the average owner's equivalent of rent in the segment. The segments in the stratum are selected with the probability proportional to the segment's size, where the size of the segment corresponds to the segment's estimated total spending on shelter. Finally, the CPI program selects a representative sample of renters in each segment.

The rental units in each of the six strata are interviewed every six months on a panel basis. One of the six panels is priced each month and each panel is priced twice per year. Thus, the month-to-month price changes in housing services are calculated using the six-month changes in rents.

From each rental unit in the sample, information on the economic rent and on the pure rent is collected. The economic rent is the contract rent (including the value of certain rent reductions) adjusted by the value of any changes in the services the landlord provides. A change in what renters obtain for their rents is considered to be a quality change, and the value of any quality change is applied to the current economic rent to make it consistent with the previous data. The pure rent is used in calculations of the owners' equivalent of rent. It is the economic rent minus any utilities included in the contract rent. The utilities paid by homeowners are counted outside the CPI Shelter.

To construct the CPI Rent and CPI OER, the CPI program uses the so-called price relatives. The price relative is the ratio of (weighted) prices from the current month to the (weighted) prices in the previous month. Since each housing unit is interviewed every six months, the monthly price relative is the sixth root of the six-month price change. For example, the six-month change in rent for all renter-occupied units in a segment is the ratio of (1) the sum of the current economic rents for each sampled unit within the segment, weighted by the total renter weight for that segment, and (2) the sum of the economic rents charged six months ago for each sampled unit within the segment, weighted by the total renter weight for that segment. The total renter weight in a segment is the product of the segment's weight, the renters' share in the total renter- and owner-occupied spending on shelter in the segment, and the inverse of the probability of a housing unit in the segment to be selected to the sample. The latter corrects for the sampling design. The segment's weight is the inverse of the probability of its being included in the stratum, where the probability is the ratio of the total spending on shelter in the segment to the total spending on shelter in the stratum.

Consider rental unit i in segment s, which is located in pricing area a. Let W_s denote the segment's s weight. Let S_s denote the renters' share in the total renter- and owner-occupied spending on shelter in

segment s. Let p_s denote the probability of a unit in segment s to be selected to the sample. Then, the monthly relative price change for the CPI Rent for area a, $\Delta_{a,rent}^{t-1,t}$, is

$$\Delta_{a,rent}^{t-1,t} = \sqrt[6]{\frac{\sum_{i\in a}\left[\left(W_s\frac{S_s}{p_s}\right)\text{econ rent}_{i,t}\right]}{\sum_{i\in a}\left[\left(W_s\frac{S_s}{p_s}\right)\text{econ rent}_{i,t-6}\right]}}.$$

The monthly relative price change for the OER index for area a, $\Delta_{a,OER}^{t-1,t}$, is

$$\Delta_{a,OER}^{t-1,t} = \sqrt[6]{\frac{\sum_{i\in a}\left[\left(W_s\frac{1-S_s}{p_s}\right)\text{pure rent}_{i,t}\right]}{\sum_{i\in a}\left[\left(W_s\frac{1-S_s}{p_s}\right)\text{pure rent}_{i,t-6}\right]}}.$$

Then, the CPI Rent and the CPI OER for area a are calculated as follows:

$$\begin{aligned} I_{a,rent}^{t} &= I_{a,rent}^{t-1}\Delta_{a,rent}^{t-1,t} \\ I_{a,OER}^{t} &= I_{a,OER}^{t-1}\Delta_{a,OER}^{t-1,t}. \end{aligned}$$

These measures are then used to aggregate the indexes across all CPI index areas.

REFERENCES

Bureau of Labor Statistics. 2007. "The Consumer Price Index." Available at www.bls.gov/opub/hom/pdf/homch17.pdf.

Bureau of Labor Statistics. 2009. "How the CPI Measures Price Changes of Owners' Equivalent Rent of Primary Residence (OER) and Rent of Primary Residence (Rent)." Available at www.bls.gov/cpi/cpifacnewrent.pdf (April).

Crone, Theodore M., Leonard I. Nakamura, and Richard Voith. 2010. "Rents Have Been Rising, Not Falling, in the Postwar Period." *The Review of Economics and Statistics* 92 (August): 628–42.

Diaz, Antonia, and Maria Jose Luengo-Prado. 2008. "On the User Cost and Homeownership." *Review of Economic Dynamics* 11 (July): 584–613.

Diewert, W. Ervin, and Alice O. Nakamura. 2009. "Accounting for Housing in a CPI." The Federal Reserve Bank of Philadelphia Working Paper No. 09-4 (March).

Diewert, W. Ervin, and Alice O. Nakamura, and Leonard I. Nakamura. 2009. "The Housing Bubble and a New Approach to Accounting for Housing in a CPI." *Journal of Housing Economics* 18 (September): 156–71.

Dougherty, Ann, and Robert Van Order. 1982. "Inflation, Housing Costs, and the Consumer Price Index." *American Economic Review* 72 (March): 154–64.

Garner, Thesia I., and Randal Verbrugge. 2007. "The Puzzling Divergence of U.S. Rents and User Costs, 1980–2004: Summary and Extensions." Bureau of Labor Statistics Working Paper No. 409 (October).

Genesove, David. 2003. "The Nominal Rigidity of Apartment Rents." *Review of Economics and Statistics* 85 (November): 844–53.

Gillingham, Robert. 1980. "Estimating the User Cost of Owner-Occupied Housing." *Monthly Labor Review* 103 (2): 31–5.

Kearl, J. R. 1979. "Inflation, Mortgage, and Housing." *Journal of Political Economy* 87 (October): 1,115–38.

Kudlyak, Marianna. 2009. "The Cyclicality of the User Cost of Labor with Search and Matching." Federal Reserve Bank of Richmond Working Paper No. 09-12.

Poole, Robert, Frank Ptacek, and Randal Verbrugge. 2005. "Treatment of Owner-Occupied Housing in the CPI." Bureau of Labor Statistics Working Paper, presented to the Federal Economic Statistics Advisory Committee (December).

Prescott, Edward C. 1997. "On Defining Real Consumption." The Federal Reserve Bank of St. Louis *Review* May: 47–53.

Verbrugge, Randall. 2008. "The Puzzling Divergence of Rents and User Costs, 1980–2004." *Review of Income and Wealth* 54 (December): 671–99.

Wolman, Alexander L. 2011. "K-Core Inflation." Federal Reserve Bank of Richmond *Economic Quarterly* 97 (4): 415–30.

Economic Quarterly—Volume 98, Number 3—Third Quarter 2012—Pages 209–230

When Do Credit Frictions Matter for Business Cycles?

Felipe Schwartzman

The Great Recession took a turn for the worse in October 2008, at the same time as the collapse of Lehman Brothers. Many researchers viewed this confluence of events as evidence that credit frictions—disfunction in credit markets that distort the cost of intertemporal trade—were, if not the ultimate cause, at least a key mechanism that made the recession much deeper and more prolonged than it would otherwise have been.[1] As a consequence, there has been renewed interest in constructing macroeconomic models that are able to capture this idea (see Kiyotaki and Gertler [2011], Quadrini [2011], and Brunnermeier, Eisenbach, and Sannikov [2012] for recent reviews of that literature).

One difficulty that stems from that view is that adding credit frictions to an otherwise standard frictionless general equilibrium business cycle model is not necessarily sufficient to generate dynamics that are quantitatively or even qualitatively compatible with actual business cycles. The reasons behind this difficulty are fairly general, and stem from the fact that most models of credit frictions normally work by distorting the terms of *inter*-temporal tradeoff faced by firms or households. However, as we will see, realistic business cycle dynamics require shocks to affect *intra*-temporal labor supply decisions.

In order for models with credit frictions to generate compelling results, it is necessary to depart from more conventional ways of modelling preferences and technology, for example by adding a capacity

The views expressed here do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: felipe.schwartzman@rich.frb.org.

[1] See Campello, Graham, and Harvey (2010); Ivashina and Scharfstein (2010); and Puri, Rocholl, and Steffen (2011) for empirical articles that establish the link between changes in credit conditions over that period and production decisions.

utilization margin, by introducing working capital requirements, or by allowing for firm-level heterogeneity. Furthermore, credit frictions can play a more realistic role in the presence of additional frictions, such as sticky prices and incomplete contracts. This article is a guide to these modifications, explaining how they work, why they are necessary, and how they can be motivated

The identification of credit frictions with distortions to the intertemporal tradeoffs faced by particular agents allows us to cover a large part of the literature, but not all of it. The last section of this article discusses a couple of recent examples in the literature where credit frictions act instead by affecting agents' risk management opportunities and bargaining position. This line of research is promising exactly because it sidesteps a lot of the difficulties associated with a heavy reliance on intertemporal distortions.

As with any literature review, this is by necessity limited in scope. The focus is on articles that strive to make a quantitative point, rather than only exposing a qualitative mechanism. Furthermore, we do not discuss the vast literature about what exactly gives rise to these frictions, rather, taking as given that they might become more important in certain instances and tracking down what this implies.

We proceed as follows: First, we motivate interpreting credit frictions as a tax on intertemporal trade. This is a simplification that will be useful for the rest of the text, since it will allow us to focus sharply on the impact of credit frictions on the decisions of households and of non-financial firms while abstracting from the precise mechanism that gives rise to those frictions. We then review two ways in which general equilibrium considerations can limit the impact of credit frictions. The first is that changes in the demand for physical capital induced by changes in the intensity of credit frictions only translates into a significantly lower capital stock over a long period of time. Second, changes in intertemporal tradeoffs faced by households often lead them to increase consumption as they reduce labor supply and vice versa. We then discuss extensions and modifications to the baseline model that help mitigate or reverse some of these effects. The last section concludes.

1. CREDIT FRICTIONS AS A TAX ON INTERTEMPORAL TRADE

Credit frictions appear in many forms. They can originate from asymmetric information (as in Bernanke and Gertler [1989]) or from limited commitment problems (as in Kiyotaki and Moore [1997]), and may show up in the data as quantitative limits on borrowing, down

payment or margin requirements, non-linear pricing for debt, and outright exclusion of particular agents from credit markets. All of these forms have one feature in common: The agents directly affected behave as if they were subject to a tax on borrowing, effectively applying to their decisions interest rates that are higher than if they were not subject to the friction. Furthermore, in the same way as a tax on credit, credit frictions impact equilibrium interest rates, which also impact agents who are not subject to the underlying commitment or informational problems.

In that spirit, much of this article will take a reduced form approach to credit frictions, equating variations in the intensity of the friction with variations in the "after tax" interest rate faced by borrowers or the "before tax" rate faced by lenders. The reduced form approach is appropriate given that the purpose of the article is to describe model dynamics rather than to discuss policy. It has the added advantage of putting the focus sharply on the reaction of individual agents to changes in the credit frictions as opposed to the details about how they are determined.

In many instances it will be useful to take a partial equilibrium approach, to take other prices as given when discussing the impact of the change in the after tax interest rate on an agent's decision. This should capture the primary impact of credit frictions in most of the models under review. In some important instances a full comprehension of the mechanism will also require referring to general equilibrium effects. We will address these as needed.

The interpretation of credit frictions as a tax on borrowing is in line with the interpretation given by Chari, Kehoe, and McGrattan (2005) and, in policy circles, is used by the Estimated, Dynamic, Optimization-based Model of the U.S. Economy used by the Federal Reserve Board (see Chung, Kiley, and Laforte [2010]). Chari, Kehoe, and McGrattan (2005) discuss how the canonical models with credit frictions by Bernanke and Gertler (1989), Carlstrom and Fuerst (1997), and Kiyotaki and Moore (1997) can be reinterpreted as models of the determination of a tax on borrowing. In these models, the tax wedge appears as one operating between households, who save, and firms, who borrow. In richer environments with firm or household heterogeneity, the tax wedge can also appear as differences in the interest rate faced by different firms or different households (see Buera and Moll [2012] for a discussion).

When bringing the models to the data, it is important to remember that from the perspective of individual agents, changes in this tax wedge can appear either as a change in the risk premium paid by an agent on her loans, or, given a quantitative limit on debt, as an increase in

the shadow cost of funds. Therefore, the intensity of credit frictions are not well-measured by the riskless interest rate paid by the U.S. Treasury or the overnight interest rate paid in interbank markets, both of which often drop in recessions. Rather, they are best measured by a wide spectrum of credit indicators that are strongly pro-cyclical such as credit flows, the fraction of senior bank managers who report tightening of credit standards, and spreads between bonds with different ratings (see Quadrini [2011] for a discussion of these indicators).

Not all the effects of credit frictions can be easily summarized with a tax on credit relationships. In the last section we examine two cases of recent articles where the most interesting effects of the credit frictions are unrelated to that particular aspect.

2. PROBLEMS FOR MODELS WITH CREDIT FRICTIONS

Take a basic real business cycle model such as King, Plosser, and Rebelo (1988) and add credit frictions to it. Most likely, these frictions are going to either imply counterfactual dynamics or will matter very little for aggregate fluctuations. There are two main problems: The first is that while investment might react strongly to increases in credit frictions, the capital stock is a slow-moving variable. Thus, for credit frictions to matter, they need to have an impact on employment. The second problem is that the typical impact of credit frictions on employment is such that employment and consumption have the wrong co-movement, with consumption booming when employment drops and vice versa. We discuss each of these problems in turn.

Problem 1: Capital Stock is Inelastic in the Short Run

Firms that face higher borrowing costs are likely to reduce their investment. If nothing else, this should be one channel through which tighter credit would lead to lower output. However, given conventional calibration of technology and preferences, tightening credit will only have a modest impact on output through this channel.[2]

Over the short run, the capital stock is inelastic because buildings and equipment in place do not become unproductive overnight for lack of replacement or maintenance. For example, suppose yearly depreciation is 10 percent of capital and steady-state investment is 12 percent of

[2] This is a problem originally pointed out by Kocherlakota (2000).

capital stock (so that capital grows by 2 percent a year). Furthermore, suppose that over the course of a year, investment drops by 18 percent, which is approximately the drop in fixed capital formation over the four quarters starting in Q4:2008 when compared to the previous four quarters. Then, the capital stock drops by about 0.16 percent. With a capital elasticity of output of about $\frac{1}{3}$, this would account for a drop in output of 0.05 percent. While this is a significant deviation for an economy that grows at 2 percent in a normal year, it cannot account for the almost 4 percent drop in gross domestic product that took place in the year after the collapse of Lehman Brothers.

One may wonder whether the almost 20 percent drop in investment is an understatement, given imperfect measurement of intangible capital and the violence of the crisis. As it turns out, this drop in investment is in line with recent quantitative work by Khan and Thomas (2011), who model the impact of a credit shock on investment decisions made by firms. Their model includes two mechanisms that keep investment from falling more substantially. First, there are adjustment costs to capital at the firm level. Second, consistent with empirical work that has found little effect of interest rates on household savings decisions (see Deaton [1992] for a review of that work), in equilibrium, interest rates have to drop substantially to convince households to reduce their savings. The drop in the interest rate, in turn, relaxes the constraints faced by firms, undoing much of their effect on investment. (This effect is emphasized and discussed in detail by Coen-Pirani [2005].)

Problem 2: Co-Movement between Labor Supply and Consumption

Business cycle models with credit frictions will usually change the households' incentives to save. This may occur directly if households are subject to time varying credit frictions, as in Cúrdia and Woodford (2009), Mendoza (2010), and Guerrieri and Lorenzoni (2011), or indirectly if, as in Carlstrom and Fuerst (1997), changes in the intensity of credit frictions applying to firms affect the equilibrium interest rate received by households who lend to these firms. Such an equilibrium adjustment is necessary since credit frictions imply that, for a given interest rate received by lenders, borrowers do not borrow as much as they would otherwise. Fluctuations in the interest rate can assure that equilibrium is maintained.

As discussed above, the empirical evidence suggests that household savings are unlikely to be very elastic to interest rates or, more generically, to incentives for intertemporal substitution. However, to the extent that they are, under conventional assumptions households

will choose to reduce labor supply at the same time that they choose to increase consumption and vice versa. If a household faces a higher borrowing interest rate, it will choose to reduce borrowing both by consuming less and by working longer hours to increase income.

This point is made transparently by Barro and King (1984) in a slightly different context. They investigate the impact of intertemporal prices on household behavior. Their result relies on two assumptions: 1) leisure is a normal good, its demand increasing in household wealth, and 2) utility is time separable, with consumption or leisure in a given time period having no effect on the enjoyment of consumption or leisure in subsequent periods. The first assumption conforms to the long-run evidence that, in spite of substantial increases in wages over time, labor supply does not exhibit a strong secular trend (King, Plosser, and Rebelo 1988). The second assumption is more controversial since models with habits are very common, but should be less controversial the longer it is under consideration.

Suppose we write the intertemporal optimization problem of a household as:

$$\max_{C_t, L_t} \sum_{t=0}^{\infty} \beta^t u\left(C_t, 1 - L_t\right)$$

$$s.t. \quad : \quad \sum_{t=0}^{\infty} \frac{1}{R_{0,t}} \left[C_t - w_t L_t\right] \leq B_0,$$

with u increasing and concave in both arguments. C_t is consumption, L_t is labor supply, $1 - L_t$ is leisure, B_0 is initial wealth, w_t is the rate at which the household can transform labor hours into consumption goods (the wage rate), and $\frac{1}{R_{0,t}}$ is the price of time t consumption relative to time 0 consumption. The one-period interest rate between t and $t+1$ is $R_{t,t+1} = \frac{R_{0,t+1}}{R_{0,t}}$.

We are interested in knowing how consumption and leisure change in response to changes in the interest rate. This can be interpreted either as the equilibrium rate faced by households or, more generally in the case where households are directly affected by credit frictions, as the "after tax" interest rate that captures the incentive impact of those frictions.

We can solve the problem in two steps. First, for a given saving decision $\{S_t\}_{t=0}^{\infty} \equiv \{w_t L_t - C_t\}_{t=0}^{\infty}$, we find how the household optimally chooses consumption and labor supply. This is the solution to the static optimization problem:

$$\begin{aligned} \left\{C\left(S_t, w_t\right), L\left(S_t, w_t\right)\right\} \quad &= \quad \arg\max_{C_t, L_t} u\left(C_t, 1 - L_t\right) \\ & \qquad C_t + S_t \leq w_t L. \end{aligned}$$

Given the solution to the static problem, we then choose a sequence of $\{S_t\}$s to solve the dynamic problem:

$$\begin{aligned} & \max_{e_t} \sum \beta^t u\left(C\left(w_t, S_t\right), 1 - L\left(w_t, S_t\right)\right) \\ s.t. \quad : \quad & \sum_{t=0}^{\infty} \frac{1}{R_{0,t}} S_t \geq -B_0, \end{aligned}$$

where the constraint states that the discounted present value of savings cannot be less than the negative initial assets of the household.

It is easy to see that consumption and leisure choices only depend on the interest rate through its effect on savings, S_t. Thus, in order to understand the impact of changes in credit frictions on consumption and labor supply, we need to understand the impact of a change in S_t in the static problem. We can rewrite the budget constraint of the static problem as:

$$C_t + w_t\left(1 - L_t\right) \leq w_t - S_t.$$

Given the saving decision, S_t, and the wage rate, w_t, the static problem has the same form as an intermediate microeconomics textbook consumer maximization problem over two goods, C_t and $1 - L_t$, with relative price w_t and wealth given by $w_t - S_t$. Since both goods are normal, the optimal response of the household to an increase in S_t is to reduce *both* consumption and leisure. Thus, if a household decides to increase saving, it will both reduce consumption and increase employment. Negative co-movement of consumption and employment is of course at odds with business cycle data.

How big are wealth effects on labor supply likely to be? Baseline calibrations of preferences imply that they ought to be substantial. Over a span of several decades, hours worked have moved relatively little when compared to the manifold increase in wages over that same period. A commonly used functional that captures this fact is[3]

$$u\left(C, L\right) = \frac{\left(C^{1-\kappa}\left(1 - L\right)^{\kappa}\right)^{1-\sigma}}{1 - \sigma}.$$

[3] More generally, King, Plosser, and Rebelo (1988) show that, in order for hours worked to remain constant over time, even as wages increase, the period utility function of households has to be:

$$\begin{aligned} u\left(C, L\right) &= \frac{1}{1-\sigma} C^{1-\sigma} v\left(1 - L\right) \text{ or} \\ u\left(C, L\right) &= \log\left(C\right) + v\left(1 - L\right). \end{aligned}$$

The solution to the period-by-period static problem implies

$$\frac{\kappa}{1-\kappa}\frac{C}{1-L} = w,$$

so that L remains constant if wages and consumption grow at the same rate. A typical calibration chooses κ so that $L = \frac{1}{3}$ since this implies that people work approximately one-third of their available time (eight hours in a day). Applying the implicit function theorem,

$$\begin{aligned}\frac{dC}{C} &= -\frac{L}{1-L}\frac{dL}{L} \\ &= -\frac{1}{2}\frac{dL}{L}.\end{aligned}$$

Thus, a change in the interest rate that leads to a 1 percent drop in consumption will also generate a 2 percent rise in employment.

An instructive example of how this mechanism operates in an equilibrium environment is Chari, Kehoe, and McGrattan (2005). The article studies the effect of a "sudden stop" in foreign capital flows to a small open economy. The shock takes the form of a temporary quantitative limit in net imports from abroad. The economy accommodates to the tightened limit with a reduction in consumption and investment, but an increase in employment. The sudden stop in foreign capital flows leads to an output boom.

Another example is Carlstrom and Fuerst (1997). They study the effect of a shock to entrepreneurial wealth in a closed economy. Given the credit friction, the shock reduces the borrowing capacity of entrepreneurs, forcing them to reduce investment. However, in general equilibrium this can only be accomplished through an increase in household consumption. The price change that supports this consumption boom is a reduction in the interest rate faced by households. In response, households increase leisure and reduce employment. In sum, in response to an entrepreneurial wealth shock, the model generates a recession with a consumption boom.

One important lesson from Barro and King (1984) is that, given their assumptions, for shocks to generate realistic co-movement between consumption and labor supply, they need to have an impact on the wage rate, w_t. In the rest of the article, we will review some of the strategies that the literature has devised to have the wage rate move in response to credit frictions.

Summary

The analysis above also gives some hints as to which mechanisms are likely to generate realistic business cycle fluctuations. These are

typically mechanisms that 1) have an impact on employment and 2) affect real wages. Wages can drop if a shock generates a reduction in labor productivity or if the shock acts as a tax on wages (a "labor wedge"). Such shocks will lead firms to want to hire fewer workers and, in equilibrium, there will be a drop in wages that will induce households to reduce both their consumption and their leisure time.

Numerical studies of general equilibrium stochastic growth models bear out this intuition, implying that productivity shocks and labor wedge shocks account for the bulk of business cycle fluctuations, including the Great Depression (see Chari, Kehoe, and McGrattan [2005] and the various articles collected in Kehoe and Prescott [2002]), but an "interest rate wedge" (i.e., a tax on saving or investment) measured in a similar way cannot account for much.

3. MODIFICATIONS OF THE STANDARD MODEL

We now turn to modifications to the baseline model that help increase the potential impact of credit on the real economy.

Labor Supply

The literature on credit frictions has adopted particular functional forms for the utility function that eliminate or greatly mitigate the wealth effect on labor supply. One popular solution is to postulate a utility function as in Greenwood, Hercowitz, and Huffmann (1988):

$$u\left(C_t, L_t\right) = u\left(C_t - \psi\left(L_t\right)\right),$$

where ψ is increasing and convex. A static optimization problem using this utility function yields the following first-order condition:

$$w_t = \psi'\left(L_t\right),$$

where w_t is the wage rate. Now, labor supply depends only on current wages, regardless of consumption. Thus, with this utility function, a change in credit conditions can lead to a decrease in consumption without an increase in labor supply.

One motivation for using this utility function is that it captures home production (Greenwood, Rogerson, and Wright 1995). In this interpretation, $\psi\left(L_t\right)$ is the loss in goods produced at home that occurs when a household decides to offer its labor in the market. An objection to the Greenwood, Hercowitz, and Huffmann (1988) utility function is that it implies a long-run trend in working hours as wages increase over time. This need not be the case if labor productivity in home production increases at the same rate as labor productivity in

market production. In modelling terms, all this assumption requires is substituting $\psi(L_t)$ for $(1+g)^t \psi(L_t)$, where g is the per-period growth rate in the economy.

An alternative proposed in a different context, but that preserves the long-run properties of the utility function advanced by King, Plosser, and Rebelo (1988) while generating short-term properties more in line with Greenwood, Hercowitz, and Huffmann (1988), is Jaimovich and Rebelo (2009), who put forward a utility function of the form

$$\begin{aligned} u(C_t, X_t, L_t) &= u\left(C_t - \psi L_t^{\theta} X_t\right), \\ X_t &= C_t^{\gamma} X_{t-1}^{1-\gamma}, \end{aligned}$$

where X_t can be interpreted as a habit in consumption. With $\gamma = 1$, the preference is in the class discussed by King, Plosser, and Rebelo (1988), whereas with $\gamma = 0$ it is as proposed by Greenwood, Hercowitz, and Huffmann (1988). In a model with news shock, Schmitt-Grohé and Uribe (2012) estimate γ to be close to zero, implying a utility function very close to the one proposed by Greenwood, Hercowitz, and Huffmann (1988).

Capacity Utilization

While the stock of buildings and machinery cannot change quickly, the utilization of that stock can. However, in a conventional model of capacity utilization, the same credit frictions that lead firms to reduce fixed investment will also lead them to increase capacity utilization. The intuition is similar to the incentive for households to increase labor supply when facing a higher cost of borrowing: If borrowing is more costly, this raises the value of current income relative to future income.

Suppose capital depreciation is an increasing and convex function of capacity utilization u as in Greenwood, Hercowitz, and Huffmann (1988). That is, firms refrain from using their capital at full capacity because higher capital utilization subjects them to more frequent breakdowns in their machinery, thus requiring them to replace damaged capital. Suppose also that firms face convex installation costs to new capital, which imply that they would optimally choose to avoid wide swings in investment. Firms face a marginal one-period interest rate $R_{t,t+1}$ on borrowing and lending so that they use its inverse as its discount rate when evaluating production decisions. As before, we interpret this interest rate as being the "after tax" cost of capital faced by firms, including the various credit frictions that they might be subject to. At $t = 0$, the problem of a firm with convex capital installation

costs is

$$\max_{\{u_t,K_{t+1},L_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \frac{1}{R_{0,t}} \left[A_t F\left(u_t K_t, L_t\right) - w_t L_t - I_t - g\left(I_t\right)\right]$$
$$s.t. \quad : \quad K_{t+1} = I_t + \left(1 - \delta\left(u_t\right)\right) K_t,$$

where, as in the household problem, $R_{0,t}$ is the discount rate applied by the firm between 0 and t, I_t is investment, L_t is labor, K_t is capital, A_t is total factor productivity, w_t is the wage rate, $g\left(I\right)$ is the installation cost,[4] with g increasing and convex, δ is increasing and convex, and F has the usual properties (increasing, concave, differentiable, constant returns to scale). We can solve out L_t and write the problem as[5]

$$\max_{\{u_t,I_t,K_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \frac{1}{R_{0,t}} \left[\pi\left(w_t, A_t\right) u_t K_t - I_t - g\left(I_t\right)\right]$$
$$s.t. \quad : \quad K_{t+1} = I_t + \left(1 - \delta\left(u_t\right)\right) K_t,$$

where $\pi\left(w_t, A_t\right) u_t K_t$ is firm revenue net of the wage bill and $\pi_w\left(w_t, A_t\right) < 0$, $\pi_A\left(w_t, A_t\right) > 0$ since higher wages relative to labor productivity lead the firm to use less labor, thus decreasing the marginal product of capital.

The first-order condition with respect to u_t is (after cancelling out K_t)

$$\pi\left(w_t, A_t\right) = \delta'\left(u_t\right) \chi_t.$$

Since $\delta\left(u_t\right)$ is convex ($\delta''\left(u_t\right) > 0$), we have that capacity utilization u_t increases with productivity, decreases with wages, and decreases with the shadow value of capital at period t, χ_t.

The first-order condition with respect to I_t is

$$1 + g'\left(I_t\right) = \chi_t.$$

Since g is convex, g' is increasing in I_t, and investment increases with the shadow value of capital χ_t. Take the first-order condition with respect to K_{t+1}:

$$\chi_t = \frac{\pi\left(w_{t+1}\right) u_{t+1} + \left(1 - \delta\left(u_{t+1}\right)\right) \chi_{t+1}}{R_{t,t+1}}.$$

---

[4] The functional form for capacity utilization costs is slightly unusual and is adopted for didactic purposes. More common forms are $g\left(\frac{I_t}{K_t}\right)$ and $g\left(\frac{I_t}{I_{t-1}}\right)$, with g increasing and concave. These forms ensure that $\frac{I}{K}$ remains cosntant over a balanced growth path. We adopt the simpler $g\left(I_t\right)$ because this conveys the main intuition without burdening the notation.

[5] For example, if $F\left(u_t K_t, L_t\right) = \left(u_t K_t\right)^{\alpha} L_t^{1-\alpha}$, then $\pi\left(w_t\right) = \alpha\left(u_t K_t\right)^{\alpha} \left(\frac{1-\alpha}{w_t}\right)^{\frac{1-\alpha}{\alpha}}$.

Iterating forward and imposing the transversality condition $\lim_{T\to\infty} \frac{\chi_T}{R_{t,T}} = 0$, we get that

$$\chi_t = \sum_{v=1}^{T-1} \Delta_{t,t+v} \frac{\pi\left(w_{t+v}\right) u_{t+v}}{R_{t,t+v}} = 0.$$

A higher interest rate $R_{t,t+v}$ decreases χ_t, the shadow value of capital in place. This implies that the firm has a lower incentive to invest, but also a higher incentive to utilize capacity more intensively. The reason is that a firm that faces high borrowing costs is less concerned about preserving its production capacity in the future relative to generating current cash flows. Thus, an increase in borrowing costs leads to a production boom.

In their study of the Korean crisis, Gertler, Gilchrist, and Nataluci (2007) propose a modification to the cost of capacity utilization that is able to sidestep this difficulty. Their proposed setup is equivalent to assuming that capacity utilization does not require replacing the capital stock, thus forcing the firm to incur new convex installation costs, but instead leads to an increase in maintenance expenses, which can be paid for without paying installation costs again. Under that assumption, the problem of the firm becomes[6]

$$\begin{aligned} &\max_{\{u_t, K_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \frac{1}{R_{0,t}} \left[\pi\left(w_t\right) u_t K_t - I_t - \delta\left(u_t\right) K_t - g\left(I_t\right)\right], \\ s.t. \quad &: \quad K_{t+1} = I_t + (1-\lambda) K_t, \end{aligned}$$

where, as before, g is increasing and convex and δ is increasing and convex; λ is a scalar capturing the depreciation rate. The first-order condition for capacity utilization becomes

$$\pi\left(w_t\right) u_t = \delta'\left(u_t\right).$$

Utilization does not depend on the price of capital, since maintenance does not have any bearing on future capital stock and, therefore, the firm does not face any intertemporal tradeoff when setting its capacity utilization.

---

[6] Gertler, Gilchrist, and Nataluci (2007) define variables slightly differently, with investment given by the sum of new capital and maintenance costs. Under this redefinition, the problem becomes

$$\begin{aligned} &\max_{\{u_t, K_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \frac{1}{R_{0,t}} \left[\pi\left(w_t\right) u_t K_t - \bar{I}_t - g\left(\bar{I}_t - \delta\left(u_t\right) K_t\right)\right], \\ s.t. \quad &: \quad K_{t+1} = \bar{I}_t + \left(1 - \lambda - \delta\left(u_t\right)\right) K_t, \end{aligned}$$

where $\bar{I}_t \equiv I_t + \delta\left(u_t\right)$.

Working Capital

Macroeconomic models normally use capital as a metaphor for machinery and buildings. However, an important part of corporate finance concerns the management of working capital. This includes all the short-term assets and liabilities that firms need to hold in order to run their business. A large part of working capital is linked to payroll and to other variable inputs. Hence, increases in the cost of working capital could presumably lead to a reduction in the use of those variable inputs and of production. Early studies of banking crises have emphasized the effect of credit shocks on the ability of firms to manage working capital (see, for example, Sprague [1907]). More recently, models of financial crises often feature working capital as an important propagation mechanism (see, for example, Neumeyer and Perri [2005] and Mendoza [2010] for discussions of financial shocks in emerging economies and Perri and Quadrini [2011] and Jermann and Quadrini [2012] for discussions of financial shocks in advanced economies).

There are two motivations for the need to borrow in order to fund payroll and acquisition of materials. The first, and most common one, emphasizes the need to hold cash in order to pay for variable inputs. In its modern macroeconomics form it was pioneered by Christiano and Eichenbaum (1992) and Fuerst (1992). Increases in borrowing costs increase the opportunity cost of holding cash and, thus, of hiring labor and buying materials. This view of working capital also underlies much of the work on emerging market crises, starting with Neumeyer and Perri (2005). One difficulty for the emerging market literature is that many emerging economies experienced periods of very high inflation in which holding any cash whatsoever would be extremely costly. To get around this problem, articles in that literature assume that the opportunity cost of holding cash is proportional not to the nominal interest rate, but to the real interest rate. The implicit assumption is that firms are able to perform their payments with inflation indexed securities that, however, do not pay any real interest rate.

A second approach that does not rely on a need for a special asset to make payments is simply to recognize that there is a time lag between the acquisition and use of inputs and the sale of output, as evidenced by holdings of inventories not only of finished goods, but also of raw materials and work in process (Schwartzman 2010). This time to produce and distribute goods implies that fluctuations in borrowing costs affect the demand for variable inputs in a very similar way to the payment friction channel emphasized in other articles. One advantage of this approach is that it allows for a clean calibration of working capital demand using steady-state inventory/cost ratio as a target. Also, it provides a clear motivation for using real as opposed to nominal

interest rates as the cost of working capital. Schwartzman (2010) shows that this channel allows a multi-sector small open economy model to account for a substantial part of the sectoral reallocation that takes place in the aftermath of emerging market crises.

Working capital requirements often appear in the firm's problem by requiring firms to pay for labor one period in advance, thus borrowing in order to pay for the wage bill. In a setup where labor factor is used one period before production takes place, the problem of a firm that faces decreasing marginal returns to labor input (suppose for simplicity its capital stock is fixed at 1) is

$$\max_{\{l_{t+s+1}\}_{s=0}^{\infty}} \sum_{s=0}^{\infty} \frac{1}{R_{t,t+s}} \left[A l_{t+s+1}^{\omega} - w_{t+s+1} l_{t+s+2} \right].$$

The first-order condition is

$$\omega A l_{t+s}^{\omega-1} = w_{t+s-1} R_{t+s-1,t+s}.$$

Hence, an increase in the one-period interest rate has a similar impact on labor demand as an increase in wages in the same proportion. However, households are only compensated for their labor through wages. In effect, because labor demand drops, in equilibrium wages must drop for the labor market to clear. The higher interest rate functions as a tax on labor, leading to a drop in employment and consumption.

The demand for working capital and capacity utilization decisions reinforce each other. The point is clear in Schwartzman's (2010) study of emerging market crises in the presence of demand for working capital, where he finds that adding a capacity utilization margin to a model with working capital almost doubles the aggregate output reduction from an interest rate increase.

Firm Heterogeneity

One key benefit of well-functioning credit markets is that they direct resources to the most productive uses. If credit markets malfunction, aggregate productivity in the economy may suffer. Models with firm heterogeneity capture that idea. Credit frictions typically imply larger interest rates (or shadow cost of funds) for borrowers than for savers. Since borrowing firms tend to also be the most productive ones, this means that an exacerbation of credit frictions will lead capital and labor to move from high productivity units that borrow a lot, to low productivity ones that borrow little if at all. This reallocation reduces the average total factor productivity in the economy. Because

productivity drops, wages drop, leading to reduced incentives for labor supply and to a recession.

The misallocation is at the heart of the output drop in Kiyotaki and Moore (1997) and Kiyotaki (1998), and more recently in Gilchrist, Sim, and Zakrajsek (2010). While intuitively appealing, the capital reallocation view has not fared particularly well in some quantitative studies. One notable example is Cordoba and Ripoll (2004). The authors show that in order for capital reallocation to have a large impact on output, it is necessary for capital to account for a large share of output. In their parameterizations, they find that large amplification requires capital shares of output close to 0.8. This, they argue, is too large in the face of an aggregate capital share of close to $\frac{1}{3}$.

Cordoba and Ripoll (2004) may exaggerate the difficulties of the capital reallocation model by focusing too narrowly on the reallocation of fixed capital while keeping labor reallocation in the background. To see this, consider the following firm problem:

$$\begin{aligned} &\max Y_t - (R_t - (1-\delta))\, K_t - w_t L_t \\ &Y_t = A_t K_t^{\alpha} L_t^{\beta}, \end{aligned}$$

where K_t is capital and L_t is labor. Solving out the firm's problem yields

$$Y_t = A_t \left(\frac{\alpha}{R_t - (1-\delta)} \right)^{\frac{\alpha}{1-\alpha-\beta}} \left(\frac{\beta}{w_t} \right)^{\frac{\beta}{1-\alpha-\beta}}.$$

With credit frictions, interest rates are firm-specific. In many models, increases in credit frictions imply that interest rates are higher for firms with large productivity (high A_t) relative to those with low productivity (low A_t). Thus, there is a decrease in output of high productivity firms relative to that of low productivity, leading to a drop in aggregate output.

How much of a change in relative output there is for a given change in relative interest rates depends on the elasticity of output to the user cost of capital $R_t - (1-\delta)$. This elasticity is given by the exponent $\frac{\alpha}{1-\alpha-\beta}$. Cordoba and Ripoll (2004) assume that firms have a fixed labor input, which insofar as the firm problem is concerned, is equivalent to assuming $\beta = 0$. It follows that the elasticity of firm-level output to the user cost of capital $R_t - (1-\delta)$ is $\frac{\alpha}{1-\alpha}$. Cordoba and Ripoll's preferred calibration has α close to $\frac{1}{3}$, so that $\frac{\alpha}{1-\alpha} = \frac{1}{2}$. In comparison, if firms can choose how many workers to hire, then the elasticity is $\frac{\alpha}{1-\alpha-\beta}$. Supposing that $\alpha + \beta = 0.9$, which is not far from common estimates of the degree of decreasing returns to scale, and keeping the capital share to $\frac{1}{3}$, then the elasticity of firm-level output to the user cost of

capital rises to a much more substantial $\frac{10}{3}$. This is an effect almost seven times as large.

Another, more recent quantitative study of a model with firm heterogeneity and credit frictions that allows for full labor mobility across firms is Khan and Thomas (2011). In the article, the authors introduce firms that face a quantitative constraint in their ability to borrow and fixed investment costs. The quantitative constraint on borrowing implies that firms for which the borrowing constraint binds face an infinite borrowing rate on the margin, and firms for which the constraint does not bind may refrain from borrowing to preserve financing capacity. Khan and Thomas (2011) study the impact of a shock to the maximum leverage that firms can hold. While they find that a persistent shock to leverage can have a sizeable impact on productivity after several quarters, the shock does not have any immediate impact on average productivity and leads, in fact, to a short-lived consumption boom. The reason is that with realistic investment adjustment costs at the firm level, capital reallocation takes time. Over the short run, productive firms keep their capital even in the face of tighter constraints and unproductive firms do not expand even in the face of lower interest rates.

Sticky Prices

Fluctuations in the intensity of credit market frictions generate the correct patterns of business cycle co-movement in the presence of an unrelated but widely used friction: sticky prices. Examples of models with financial frictions that use sticky prices are Bernanke, Gertler, and Gilchrist (1999); Del Negro et al. (2009); Gertler and Karadi (2011); and Christiano, Motto, and Rostagno (2013)

Sticky prices do not change in any way the direct impact of changes in the borrowing rate on investment, consumption, or labor supply decision. Rather, what they do is translate changes in the demand for investment or consumption goods into changes to the real wage. In this class of models, monopolistic firms commit to matching whatever demand they face at a price they have previously set, irrespective of what this implies to their marginal costs. If both consumption and investment drop in a given period, firms keep their price constant but hire fewer workers, thus paying lower real wages and increasing their markups. This lower real wage leads workers to want to work less and consume less.[7]

[7] With sticky wages the workers pre-commit to supplying as much labor as needed to satisfy demand at the pre-determined prices, so that over the short run they lose the ability to adapt labor supply decisions to credit conditions.

Closing sticky price models requires a policy rule adopted by the central bank, such as the Taylor rule, to determine the nominal interest rate. In principle, the policy rule could be chosen so as to keep markup fluctuation at a minimum, thus essentially replicating the allocation of a flexible price model. However, if the central bank is constrained by a zero lower bound on the nominal interest rate, then the central bank does not have any option but to allow markups to vary a lot. In such an environment, fluctuations in borrowing costs can be particularly potent (see Del Negro et al. [2009] and Gertler and Karadi [2011]).

Recently, New Keynesian models with credit spreads as a main driving force have been used to suggest that credit frictions are important to explain regular business cycles. In terms of making the quantitative case, the most well-developed model is the one by Christiano, Motto, and Rostagno (2013). There, the authors find that volatility shocks (which, in their model have a direct impact on credit spreads) account for about 60 percent of business cycle fluctuations.

Risk Management and Bargaining

The bulk of the survey was concerned with models where the action occurs because of changes in the cost of borrowing and lending faced by firms or households. These are not the only way in which credit markets affect the economy. In this section we give two examples from the recent literature where credit frictions operate indirectly by affecting risk management decisions or bargaining relationships.

Borrowing limits when combined with incomplete insurance can lead to significant risk management concerns that distort allocations. This is the focus of the article by Arellano, Bai, and Kehoe (2010). There, all of production takes place within the same period and there is no need to borrow in order to pay the wage bill. Rather, the friction is that firms pre-commit to using a certain amount of labor before they learn what their production will be. Firms normally borrow because there is a tax advantage for debt, but if output turns out to be low, they need to borrow an additional amount in order to pay for their previous commitments. There is a possibility that, at the end of the period, firms could find themselves in default because they face a borrowing limit. When this happens, they have to close, thus losing future production opportunities. In order to avert this inefficient outcome, firms may decide to restrict hiring ex-ante in order to reduce the risk of default. The increased cost of default increases the cost of hiring, thus acting like a tax on labor and reducing wages.

The second example relies on the fact that credit contracts are commitments to the transfer of future income between particular agents.

Such a pre-commitment can have implications for future bargaining with third parties. This is the route taken by Monacelli, Quadrini, and Trigari (2011). Their idea is that firms pre-commit to paying creditors before bargaining with workers. Hence, by increasing their indebtedness, firms are able to take part of the surplus out of the negotiation when negotiating wages. This leads to lower wages. While this should lead to an incentive for more job creation, it also decreases the incentive for workers to supply labor, with the latter more likely to happen over the short run.

Both of these examples serve as reminders that credit frictions can matter for business cycles even if they are not directly distorting intertemporal decisions. Exploring such possibilities is a particularly promising avenue for future research.

4. CONCLUSION

The study of quantitative macroeconomic models with credit frictions has come a long way since the seminal contributions of Bernanke and Gertler (1989), Carlstrom and Fuerst (1997), and Kiyotaki and Moore (1997). The case for an important quantitative role for such frictions is still unsettled for various reasons. On the one hand, a first brush approach using standard growth models may lead researchers to discount heavily how important such shocks can be. On the other hand, recent research shows that a number of more or less reasonable modifications can help amplify their role and imply better behaved predictions.

Many of the modifications may make models more cumbersome to write down but can be justified. That firms need to finance working capital, and that they have an important capacity utilization margin, should be uncontroversial. Firm level heterogeneity is also a well-documented fact. Finally, while the importance of sticky prices is still a matter of some controversy, it is routinely accepted as an important mechanism by a very large fraction of applied macroeconomists. Other modifications such as the adoption of preferences that shut down wealth effects on labor supply might be harder to justify.

The financial crisis of 2008–2009 has highlighted for many economists the importance of taking the financial sector seriously when thinking about macroeconomic dynamics. However, establishing that the financial sector matters for business cycles involves close attention to the seemingly unrelated issues surrounding the details of preferences, technology, and the importance of other frictions. This attention should be an important focus of future research.

REFERENCES

Arellano, Cristina, Yan Bai, and Patrick Kehoe. 2010. "Financial Markets and Fluctuations in Uncertainty." Mimeo, Federal Reserve Bank of Minneapolis (April).

Barro, Robert J., and Robert G. King. 1984. "Time Separable Preferences and Intertemporal-Substitution Models of Business Cycles." *The Quarterly Journal of Economics* 99 (November): 817–39.

Bernanke, Ben, and Mark Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79 (March): 14–31.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, edition 1, volume 1, number 1, edited by J. B. Taylor and M. Woodford. Philadelphia: Elsevier, 1–21.

Brunnermeier, Markus K., Thomas M. Eisenbach, and Yuliy Sannikov. 2012. "Macroeconomics with Financial Frictions: A Survey." Working Paper 18102. Cambridge, Mass.: National Bureau of Economic Research (May).

Buera, Francisco J., and Benjamin Moll. 2012. "Aggregate Implications of a Credit Crunch." Working Paper 17775. Cambridge, Mass.: National Bureau of Economic Research (January).

Campello, Murillo, John R. Graham, and Campbell R. Harvey. 2010. "The Real Effects of Financial Constraints: Evidence from a Financial Crisis." *Journal of Financial Economics* 97 (September): 470–87.

Carlstrom, Charles, and Timothy S. Fuerst. 1997. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review* 87 (December): 893–910.

Chari, V. V., Patrick Kehoe, and Ellen R. McGrattan. 2005. "Sudden Stops and Output Drops." *American Economic Review* 95 (May): 381–7.

Chari, V. V., Patrick Kehoe, and Ellen R. McGrattan. 2007. "Business Cycle Accounting." *Econometrica* 75 (3): 781–836.

Christiano, Lawrence J., and Martin Eichenbaum. 1992. "Liquidity Effects and the Monetary Transmission Mechanism." *American Economic Review* 82 (May): 346–53.

Christiano, Lawrence J., Roberto Motto, and Massimo Rostagno. 2013. "Risk Shocks." Working Paper 18682. Cambridge, Mass.: National Bureau of Economic Research (January).

Chung, Hess T., Michael T. Kiley, and Jean-Philippe Laforte. 2010. "Documentation of the Estimated, Dynamic, Optimization-based (EDO) Model of the U.S. Economy: 2010 Version." Finance and Economics Discussion Series, Division of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington D.C. (March 25).

Coen-Pirani, Daniele. 2005. "Margin Requirements and Equilibrium Asset Prices." *Journal of Monetary Economics* 52 (March): 449–75.

Cordoba, Juan-Carlos, and Marla Ripoll. 2004. "Credit Cycles Redux." *International Economic Review* 45 (November): 1,011–46.

Cúrdia, Vasco, and Michael Woodford. 2009. "Credit Spreads and Monetary Policy." Working Paper 15289. Cambridge, Mass.: National Bureau of Economic Research (August).

Deaton, Angus. 1992. *Understanding Consumption*. Oxford: Oxford University Press.

Del Negro, Marco, Gauti Eggertsson, Andrea Ferrero, and Nobuhiro Kiyotaki. 2009. "The Great Escape? A Quantitative Evaluation of the Fed's Non-Standard Policies." Federal Reserve Bank of New York Working Paper (December).

Fuerst, Timothy S. 1992. "Liquidity, Loanable Funds, and Real Activity." *Journal of Monetary Economics* 29 (February): 3–24.

Gertler, Mark, and Peter Karadi. 2011. "A Model of Unconventional Monetary Policy." Journal of Monetary Economics 58 (January): 17–34.

Gertler, Mark, Simon Gilchrist, and Fabio Nataluci. 2007. "External Constraints on Monetary Policy and the Financial Accelerator." *Journal of Money, Credit and Banking* 39 (2-3): 295–330.

Gilchrist, Simon, Jae Sim, and Egon Zakrajsek. 2010. "Uncertainty, Financial Frictions, and Investment Dynamics." Mimeo, Boston University.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffmann. 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review* 78 (June): 402–17.

Greenwood, Jeremy, Richard Rogerson, and Randall Wright. 1995. "Household Production in Real Business Cycle Theory." In *Frontiers of Business Cycle Research*, edited by Tom Cooley. Princeton, N.J.: Princeton University Press, 157–74.

Guerrieri, Veronica, and Guido Lorenzoni. 2011. "Credit Crises, Precautionary Savings, and the Liquidity Trap." Working Paper 17583. Cambridge, Mass.: National Bureau of Economic Research (November).

Ivashina, Victoria, and David S. Scharfstein. 2010. "Bank Lending During the Financial Crisis of 2008." *Journal of Financial Economics* 97 (September): 319–38.

Jaimovich, Nir, and Sergio T. Rebelo. 2009. "Can News about the Future Drive the Business Cycle?" *American Economic Review* 99 (September): 1,097–118.

Jermann, Urban, and Vicenzo Quadrini. 2012. "Macroeconomic Effects of Financial Shocks." *American Economic Review* 102 (February): 238–71.

Kehoe, Timothy J., and Edward. C. Prescott. 2002. "Great Depressions of the Twentieth Century." *Review of Economic Dynamics* 5 (January): 1–18.

Khan, Aubhik, and Julia K. Thomas. 2011. "Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity." Working Paper 17311. Cambridge, Mass.: National Bureau of Economic Research (August).

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo. 1988. "Production, Growth and Business Cycles." *Journal of Monetary Economics* 21: 195–232.

Kiyotaki, Nobuhiro. 1998. "Credit and Business Cycles." *Japanese Economic Review* 49 (March): 18–35.

Kiyotaki, Nobuhiro, and John Moore. 1997. "Credit Cycles." *Journal of Political Economy* 105 (April): 211–48.

Kiyotaki, Nobuhiro, and Mark Gertler. 2010. "Financial Intermediation and Credit Policy in Business Cycle Analysis." In *Handbook of Monetary Economics*, volume 3, edited by Benjamin M. Friedman and Michael Woodford. Philadelphia: Elsevier (January).

Kocherlakota, Narayana R. 2000. "Creating Business Cycles Through Credit Constraints." Federal Reserve Bank of Minneapolis *Quarterly Review* 24 (Summer): 2–10.

Mendoza, Enrique G. 2010. "Sudden Stops, Financial Crises, and Leverage." *American Economic Review* 100 (December): 1,941–66.

Monacelli, Tommaso, Vicenzo Quadrini, and Antonella Trigari. 2011. "Financial Markets and Unemployment." Working Paper 17389. Cambridge, Mass.: National Bureau of Economic Research (September).

Neumeyer, Pablo A., and Fabrizio Perri. 2005. "Business Cycles in Emerging Economies: The Role of Interest Rates." *Journal of Monetary Economics* 52 (March): 345–80.

Perri, Fabrizio, and Vincenzo Quadrini. 2011. "International Recessions." Working Paper 17201. Cambridge, Mass.: National Bureau of Economic Research (July).

Puri, Manju, Jörg Rocholl, and Sascha Steffen. 2011. "Global Retail Lending in the Aftermath of the U.S. Financial Crisis: Distinguishing Between Supply and Demand Effects." *Journal of Financial Economics* 100 (June): 556–78.

Quadrini, Vincenzo. 2011. "Financial Frictions in Macroeconomic Fluctuations." Federal Reserve Bank of Richmond *Economic Quarterly* 97 (3): 209–54.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2012. "What's News in Business Cycles." *Econometrica* 80 (November): 2,733–64.

Schwartzman, Felipe. 2010. "Time to Produce and Emerging Market Crises." Federal Reserve Bank of Richmond Working Paper 10-15 (January).

Sprague, Oliver Mitchell Wentworth. 1907. *History of Crises Under the National Banking System*. Washington, D.C.: National Monetary Commission.