SEPTEMBER/OCTOBER 1997

ECONOMIC PERSPECTIVES

A review from the
Federal Reserve Bank
of Chicago

The role of banks in monetary policy: A survey with implications for the European monetary union

Understanding aggregate job flows

FEDERAL RESERVE BANK
OF CHICAGO

Contents

This article begins with a review of the growing literature on the role of banks in the transmission of monetary policy. The authors then discuss the implications of this literature for the operation of monetary policy in the European monetary union.

The authors describe how evidence on aggregate job flows challenges standard business cycle theory and discuss recent developments in business cycle theory aimed at accounting for the evidence.

ECONOMIC PERSPECTIVES

September/October 1997, Volume XXI, Issue 5

ECONOMIC PERSPECTIVES is published by the Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors' and do not necessarily reflect the views of the management of the Federal Reserve Bank.

Single-copy subscriptions are available free of charge. Please send requests for single- and multiple-copy subscriptions, back issues, and address changes to the Public Information Center, Federal Reserve Bank of Chicago, P.O. Box 834, Chicago, Illinois 60690-0834, telephone (312) 322-5111 or fax (312) 322-5515.

ECONOMIC PERSPECTIVES is also available on the World Wide Web at http://www.frbchi.org.

ISSN 0164-0682

The role of banks in monetary policy: A survey with implications for the European monetary union

Anil K Kashyap and Jeremy C. Stein

Much of the debate about European monetary union (EMU) has focused on the likely macroeconomic effects. On the benefits side, there is clearly the reduction in transactions costs that comes from eliminating all the competing currencies. For some countries there is also the possibility that the shift to the new European central bank will bring enhanced inflation-fighting credibility. If so, these countries will enjoy lower nominal interest rates and perhaps even lower real interest rates if they can eliminate an inflation risk premium. On the cost side, some countries may see their inflation-fighting credibility decline. In addition, all countries will presumably have less freedom to use monetary policy to stimulate their own economy.

While these issues are important, we believe another crucial factor is being overlooked: the banking system aspects of monetary policy under the EMU. This article reviews some recent work, which suggests that monetary policy has significant distributional effects that operate through the banking system. We briefly discuss how this bank transmission channel may operate in the EMU.

First, we describe the conceptual differences between the bank-centric view of monetary transmission and the conventional view, in which banks do not play a key role. The bank-centric theory hinges on two key propositions: that monetary interventions do something special to banks; and that once banks are affected, so are firms and/or consumers. Then we review the empirical evidence, which tends to support the bank-centric view. Finally, we look at how a common monetary policy will affect banks throughout Europe and how this, in turn, might influence real economic activity in different countries.

A byproduct of our work is that we have developed a large amount of documentation and experience working with U.S. bank-level data, which we describe in the appendix at the end of this article. The appendix also provides details of how researchers can access these data via the Federal Reserve Bank of Chicago's Web site.

Contrasting views of monetary transmission

Conventional monetary economics

The classic textbook treatment of monetary policy focuses on how the central bank's actions affect households' portfolios. In simple terms, household portfolios are allocated between

Anil K Kashyap is a professor of economics at the University of Chicago Graduate School of Business, a consultant to the Federal Bank of Chicago, and a research associate at the National Bureau of Economic Research (NBER). Jeremy C. Stein is the J.C. Penney Professor of Management at the MIT Sloan School of Management and a research associate at the NBER. The authors thank Magda Bianco, Giovanni Ferri, Dario Focarelli, and Luigi Guizo for providing data and suggestions, as well as Christopher Tang of Thomson BankWatch and Sharon Standish of the OECD for supplying unpublished data. The Web site for the Call Report data could not have been built without the help of many Bank staff members; however, Nancy Andrews, Larry Chen, and Peter Schneider deserve special recognition for their efforts. This research was supported by a National Science Foundation grant made to the NBER.

"bonds," shorthand for all types of financial assets that are not used for transactions purposes, and money (which is the asset used in transactions). Importantly, money can be more than just currency, with checking accounts being the obvious substitute to include in narrow measures of money.

It is assumed that central banks can control the quantity of money. If the central bank can control one of the two asset types in household portfolios, it follows that by adjusting the relative supply of the two asset types, the central bank can control their relative prices. For simplicity, we often assume that transaction-facilitating assets do not pay interest. In this case, the relative price of money and bonds is the nominal interest rate. If we alter the characterization to allow transactions accounts to pay interest, the central bank will be able to influence the gap between this rate and the rate on assets with no transactions services.

Regardless of whether transactions accounts pay interest, the conventional view rests on two assumptions. First, there must be some well-defined asset called money, which is essential for transactions. Second, the monetary authority must be able to control (with some precision over intermediate horizons) the supply of money.

Historically, when demand deposits and currency were about the only assets used in transactions, it was easy to see how this control might work. Because the central bank is the only entity that can create currency, it can determine how much currency comes into circulation. Furthermore, the ability of banks (and other financial institutions) to create checking accounts has typically been limited by the requirement that banks hold reserves (which can be thought of as vault cash) against these accounts. By managing the rules regarding reserves, the monetary authority indirectly controls the non-currency component of transaction balances.

Typically, the central bank decides both the level of reserves to be held against a given level of transactions balances and the types of assets that can be used as reserves. When the central bank wants more money in the economy, it provides the banks with more currency that can be used as reserves (say by trading reserves for other bank securities). Banks then lever up the reserves through lending and crediting the checking accounts of the borrowers who receive the funds. In this framework, the willingness of banks to lend matters only to the extent to which it influences the creation of transaction-facilitating assets, that is, deposits.

Once the supply of transactions accounts has been adjusted following the central bank's reserve injection, interest rates respond in a predictable manner. When more transactions balances become available to households, the valuation of these balances falls and money becomes cheaper to hold than before—that is, nominal interest rates fall. For this change in nominal rates to matter, one must assume that prices do not adjust instantly to the change in the money supply. Then with more money, people will have more *real* purchasing power, and the nominal interest decline will correspond to a lower real interest rate.

The major problem with the conventional theory of monetary policy is the sharp two-asset dichotomy that underlies the model. There is an increasing proliferation of assets, which, from the household perspective, mimic checking accounts but are not controllable by the central bank (for example, mutual funds with check writing privileges). As these non-reservable transactions-type accounts become more prevalent, the central bank's power over currency and transaction deposits becomes less relevant in the determination of interest rates. This does not mean the central bank will no longer be able to influence rates; however, we believe that the basic logic underlying the textbook model is becoming much less compelling.

The bank-centric view

In view of the above limitation of the conventional theory, a large literature has developed based on the assumption that there are three important asset types: money, bonds, and bank loans. In this context, the special response of banks to changes in monetary policy is their *lending* response (not just their role as deposit creators). Thus, the ambiguity over what constitutes money is much less important. For this mechanism to operate, it is essential that some spending that is financed with bank loans will not occur if the banks cut the loans (that is, there is no perfect substitute available for a bank loan). The assumed sensitivity of bank loan supply to monetary policy together with the assumed dependence of some spending on bank lending generate a number of predictions about how monetary policy will work.[1]

One basic prediction is that the firms and individuals whose creditworthiness is most difficult to gauge (that is, those borrowers about whom information is imperfect) will be most dependent on banks for financing. Because these borrowers face the extra cost of raising funds from third parties, they are not indifferent about the composition of their liabilities. Banks have a particular advantage in lending to such borrowers because they can specialize in information gathering to determine creditworthiness. Moreover, by developing repeat business, banks can stay informed about their customers. They are therefore better able to make prudent lending decisions than lenders that don't have access to this information.[2]

The question of who will fund the banks remains. Banks that lend to relatively small, little known borrowers will have collections of assets that are difficult to value. This implies that individual investors are not as well informed as bank management about the value of the bank's existing assets. Depending on the type of liability the bank issues to finance itself, this may create an adverse selection problem. Banks with high levels of opaque assets need to pay a relatively high interest rate to offset the risk associated with these assets. Some banks may prefer to make fewer loans than to pay the rates required to attract funds.

One way to overcome this problem is through deposit insurance. If banks can issue insured deposits, account holders need not worry about the lending decisions made by their bank. To fund themselves with insured deposits, banks typically have to allow the entity that is providing the deposit guarantee to oversee their lending decisions. In addition, they are usually required to put aside reserves (generally currency) against the insured deposits. This link between deposit insurance and reserve requirements gives the monetary authorities a powerful lever. In effect, the reserves allow banks to raise funds without having to generate comprehensive information about the quality of their own assets. (See Stein, 1995, for the formal model.)

In this context, a reduction in the supply of reserves has an impact beyond those emphasized in the conventional textbook description: It pushes the banks toward a more costly form of financing. Because of the extra premium that banks will have to pay to bring in noninsured deposits, the banks will make fewer loans after the reserve outflow. If the borrowers that lose their loans cannot obtain new funds quickly, their spending levels may fall. Because these consequences can be partially anticipated, banks and firms will hedge this risk. Banks will not fully loan out their deposits, holding some securities as a buffer stock against a reserve outflow. Similarly, firms will hold some liquid assets on their books in case a loan is withdrawn.

Nevertheless, there are good reasons to believe that such buffer stocks will not fully offset the effects of contractionary monetary policy. For one thing, buffer stocks are costly for the banks. Banks make money by making loans, not by sitting on securities that offer returns close to the rates the banks pay on deposits. Moreover, the tax code makes it inefficient for the banks to hold securities. As with any equity financed corporation, holding these types of assets imposes double taxation on the bank's shareholders.

In summary, unlike the traditional theory that emphasizes households' preferences between money and other less liquid assets, the new theory of monetary policy asserts that the role of the banking sector is central to the transmission of monetary policy. Specifically, two key factors shape the way in which monetary policy works: 1) the extent to which banks rely on reservable deposit financing and adjust their loan supply schedules following changes in bank reserves; and 2) the extent to which certain borrowers are bank-dependent and cannot easily offset these shifts in bank loan supply.

Empirical evidence on the role of banks in monetary policy

A growing literature tests the bank-centric theory described above. Although relatively little of this research has been done using European data, we will explain in a later section why the existing results suggest there may be powerful effects in Europe.[3]

The work (which mostly focuses on the U.S.) can be summarized by the following picture of monetary policy transmission. When the Federal Reserve tightens policy, aggregate lending by banks gradually slows down and there is a surge in nonbank financing, such as commercial paper. When this substitution of financing is taking place, aggregate investment is reduced by more than would be predicted solely on the basis of rising interest rates.

Small firms that do not have significant buffer cash holdings are most likely to trim investment (particularly inventory investment) around the periods of tight money. Similarly, small banks seem more prone than large banks to reduce their lending, with the effect greatest for small banks with relatively low buffer stocks of securities at the time of the tightening. Overall, the results suggest that monetary policy may have important real consequences beyond those generated by standard interest rate effects. Below, we review this evidence in detail.

Do banks change their supply of loans when monetary policy changes?

Perhaps the simplest aggregate empirical implication of the bank-centric view of monetary transmission is that bank loans should be closely correlated with measures of economic activity. Following changes in monetary policy, there is a strong correlation between bank loans and unemployment, GNP, and other key macroeconomic indicators (see Bernanke and Blinder, 1992). However, such correlations could arise even if the "bank lending channel" is not operative. The correlations may be driven by changes in the demand for bank loans rather than the supply of bank loans. For example, bank loans and inventories might move together because banks always stand willing to lend and firms finance desired changes in inventory levels with bank loans.

Kashyap, Stein, and Wilcox (KSW, 1993) use macroeconomic data to overcome the difficulty of separating the role of loan demand from loan supply. According to KSW, movements in substitutes for bank financing should contain information about the demand for bank financing. For example, if bank loans are falling while commercial paper issuance is rising, one can infer that bank loan supply has contracted.[4] KSW examine movements in the mix between bank loans and loan substitutes following changes in monetary policy. They find that when the Fed tightens, commercial paper issuance surges while bank loans (slowly) decline.

Hoshi, Scharfstein, and Singleton (1993) conduct an analogous set of tests using aggregate Japanese data. Specifically, they compare the behavior of bank loans subject to informal control by the Bank of Japan with loans from insurance companies that are the main alternative to bank financing. As predicted by the lending channel theory, they find that when the Bank of Japan tightens, the fraction of industrial loans coming from banks drops noticeably. Arguably, the Japanese evidence is less surprising because the Bank of Japan appears to exert some direct control over loan volume in addition to any indirect control that might come from changing reserves.

Evidence relying on changes in the aggregate financing mix has been questioned because alternative explanations exist that do not rely on bank loan supply shifts. For instance, one could argue that *large* firms that typically use commercial paper financing might tend to increase *all* forms of borrowing, while smaller firms that are mostly bank-dependent receive less of all types of financing. In this case, heterogeneity in loan demand rather than differences in loan supply would explain the results above. In response to this criticism, however, Kashyap, Stein, and Wilcox (1996) show that even among a composite of large U.S. firms, there is considerable substitution away from bank loans toward commercial paper.

Calomiris, Himmelberg, and Wachtel (1995) use data on individual firms to make a similar point. Using a sample of firms that are issuing commercial paper, Calomiris et al. show that when monetary policy tightens, commercial paper issuance rises and so does the trade credit extended by these firms. This finding suggests that these larger firms are taking up some of the slack created as their smaller customers lose their bank loans. While this mechanism partially offsets the impact of the loan supply shock, it does not eliminate the shock.

Recently, Ludvigson (1996) developed a test for loan supply effects that is immune to the loan demand explanation. Comparing the extension of auto credit by banks and finance companies, the author finds that bank lending to consumers declines relative to finance company lending when monetary policy tightens, as predicted by the lending channel. The vast majority of the borrowers in this case are individuals, so it is not possible to appeal to differences in large and small buyers to explain the pattern. Furthermore, Ludvigson finds that finance company borrowers default more than bank borrowers, so finance companies are not lending more after a monetary contraction simply because they have higher-quality customers. Thus, Ludvigson's findings strongly indicate a loan supply effect of monetary policy.

The search for loan supply responses to monetary policy has also been carried out using disaggregated bank data. The theory outlined above suggests that banks that have trouble raising external finance respond differently to a monetary policy tightening from banks that can easily raise uninsured external funds. One natural proxy for the ability to raise such financing is bank size. Particularly in the U.S. where there are thousands of banks, small banks tend not to be rated by credit agencies and, therefore, have trouble attracting uninsured nondeposit financing.

In Kashyap and Stein (1995), we created a composite of small and large banks to study this question. As predicted by the theory, we find that banks of different sizes use different forms of financing. Only the larger banks have much success in securing nondeposit financing. More importantly, we find that small banks' lending is more sensitive to Fed-induced deposit shocks than that of large banks.

While these results are consistent with the idea that policy shifts induce changes in loan supply, there is also a loan demand interpretation. In this case one would have to argue that the customers of small banks differ from the customers of large banks and that loan demand drops more for customers of small banks. To take account of this possibility, we conducted further tests at the individual bank level, comparing the behavior of different small banks (Kashyap and Stein, 1997). Because most U.S. bank-level data are collected for regulatory purposes rather than for use in research, bank-level analysis requires a considerable amount of effort to get the data into usable form. As mentioned earlier, one of the byproducts of this effort is that we have developed a large amount of documentation and experience working with these data. The appendix provides a description of the data, available on the Bank's Web site; table 1 also summarizes some of the data.

At the individual bank level, the theory predicts that banks that have difficulty making up for deposit outflows should typically hold a buffer stock of securities, so that they can reduce securities holdings rather than having to cut back loans. Consistent with this prediction, table 1 shows strong evidence that small banks hold a higher fraction of assets in cash and securities than large banks. The data in table 1 also bear out other predictions of the imperfect information theory, such as small banks not being able to borrow in the federal funds market (where collateral is not used).

In terms of the search for loan supply effects, the buffer stocks will make it more difficult to find lending responses to shifts in monetary policy. Nevertheless, our research suggests that securities holdings do not seem to completely insulate bank lending from monetary policy. Even among small banks where the tendency to hold buffer stocks is most pronounced, banks with more cash and securities at the onset of a monetary contraction respond differently from less liquid banks (Kashyap and Stein, 1997). Specifically, the liquid banks are much less prone to reduce their lending following a tightening of monetary policy. Gibson (1996) shows that this pattern holds over time: When the aggregate bank holdings of securities are low, lending is more responsive to monetary policy.

The accumulated evidence shows that the bank loan supply shifts when monetary policy changes. However, there are various ways in which this loan supply shock could be neutralized. For instance, borrowers could find other nonbank lenders to fully offset the shortfall in bank lending. As a result, we must go beyond data on the volume of lending alone to see if the lending channel has any real effects on economic activity.

Does spending respond to changes in bank loan supply?

KSW check whether the financing mix has any additional explanatory power for investment once other fundamental factors, such as the cost of capital, are taken into account. The authors find that the mix does seem to have independent predictive power for investment, particularly inventory investment. Similarly, Hoshi, Scharfstein, and Singleton (1993) find that in a four-variable vector autoregression (which includes interest rates), the credit mix variable is a significant determinant of both fixed investment and finished goods inventories. Thus, the Japanese and U.S. data give the same basic message.

Working at a lower level of aggregation, Ludvigson looks at whether the financing mix (which in this case separates bank loans and finance company lending) is an important predictor of automobile sales. The author finds that the mix is a significant predictor even controlling for income, auto prices, and interest rates. This evidence strikes us as particularly

TABLE 1

Composition of bank balance sheets

As of 1976:Q1	Below 75th percentile	75th to 90th percentile	90th to 95th percentile	95th to 98th percentile	98th to 99th percentile	Above 99th percentile
Number of banks	10,784	2,157	719	431	144	144
Mean assets *(1993 $ millions)*	32.8	119.1	247.7	556.6	1,341.5	10,763.4
Median assets *(1993 $ millions)*	28.4	112.6	239.0	508.1	1,228.7	3,964.6
Fraction of total system assets	0.128	0.093	0.064	0.087	0.070	0.559
Fraction of total assets in size category						
Cash and securities	0.426	0.418	0.418	0.408	0.396	0.371
Fed funds lent	0.049	0.040	0.038	0.045	0.045	0.025
Total domestic loans	0.518	0.531	0.531	0.531	0.539	0.413
Real estate loans	0.172	0.191	0.106	0.179	0.174	0.087
C&I loans	0.102	0.131	0.153	0.160	0.168	0.171
Loans to individuals	0.147	0.162	0.148	0.147	0.138	0.059
Total deposits	0.902	0.897	0.890	0.969	0.841	0.810
Demand deposits	0.312	0.301	0.301	0.313	0.327	0.248
Time and savings deposits	0.590	0.596	0.589	0.554	0.508	0.326
Time deposits > $100 K	0.067	0.095	0.119	0.139	0.143	0.156
Fed funds borrowed	0.004	0.010	0.019	0.039	0.067	0.076
Subordinated debt	0.002	0.003	0.004	0.005	0.006	0.005
Other liabilities	0.008	0.012	0.013	0.014	0.017	0.057
As of 1993:Q2						
Number of banks	8,404	1,681	560	336	112	113
Mean assets *(1993 $ millions)*	44.4	165.8	370.1	1,072.6	3,366.0	17,413.4
Median assets *(1993 $ millions)*	38.6	155.7	362.7	920.8	3,246.3	9,297.7
Fraction of total system assets	0.105	0.078	0.060	0.101	0.106	0.551
Fraction of total assets in size category						
Cash and securities	0.399	0.371	0.343	0.333	0.325	0.311
Fed funds lent	0.045	0.040	0.035	0.041	0.041	0.040
Total loans	0.531	0.562	0.596	0.594	0.599	0.587
Real estate loans	0.296	0.331	0.337	0.302	0.252	0.209
C&I loans	0.087	0.101	0.111	0.117	0.132	0.183
Loans to individuals	0.086	0.098	0.120	0.144	0.166	0.097
Total deposits	0.879	0.868	0.850	0.794	0.760	0.690
Transaction deposits	0.258	0.257	0.254	0.240	0.258	0.193
Large deposits	0.174	0.207	0.225	0.248	0.244	0.212
Brokered deposits	0.022	0.004	0.008	0.017	0.016	0.013
Fed funds borrowed	0.010	0.021	0.039	0.063	0.097	0.093
Subordinated debt	0.000	0.000	0.001	0.002	0.004	0.017
Other liabilities	0.013	0.021	0.026	0.054	0.059	0.129
Equity	0.098	0.090	0.084	0.086	0.080	0.072

Source: Kashyap and Stein (1997).

strong, because the mix variable is added to a structural equation that is already supposed to account for monetary policy.

Among other work using disaggregated data, perhaps the most intriguing studies focus on inventory investment. Inventory reductions are large during recessions and monetary policy is typically tight prior to recessions. However, the simple story that tight money and high carrying costs lead to inventory runoffs is undermined by the difficulty in documenting interest rate effects on inventories. The previously discussed aggregate findings provide some support for the view that monetary policy and financial factors may be important for inventory movements, even though standard security market interest rates do not have much predictive power for inventories.

Gertler and Gilchrist (1994) compare the aggregate investment of a sample of large firms with that of a sample of small firms, which are presumably more bank-dependent.

They find that the small firms' inventory investment is much more sensitive to changes in monetary policy than that of the large firms. The differences are large enough that as much as half of the aggregate movement in inventory investment two years after a major monetary tightening may be attributable to the small firms. The authors find similar effects in terms of sales.

Using individual firm data, Kashyap, Stein, and Lamont (1994) look at the differences in inventory investment between publicly traded companies with bond ratings and those without bond ratings. The non-rated companies are typically much smaller than the rated companies and are more likely to be bank-dependent. The authors find that during the 1982 recession, prior to which Federal Reserve policy was restrictive, the inventory movements of the non-rated companies were much more sensitive to their own cash holdings than were the inventory movements of the rated companies. (In fact, there was no significant liquidity effect for the rated companies.) They find a similar pattern for the 1974–75 recession, which also followed a significant tightening of monetary policy by the Fed.

In contrast, in other "easy money" periods there is little relation between cash holdings and inventory movements for the non-rated companies. For instance, during 1985 and 1986, when many argue that U.S. monetary policy was particularly loose, the correlation between inventory investment and cash holdings is completely insignificant. The difference in the cash sensitivity of inventory investment for the bank-dependent firms is precisely to be expected if loan supply is varying with monetary policy.

Subsequent work by Carpenter, Fazzari, and Petersen (1994) confirms these patterns using a sample that includes information on quarterly (rather than annual) adjustments in inventories. Similarly, Milne (1991) finds similar credit availability effects on inventory investment for British firms. Thus, several independent pieces of evidence now point toward the importance of loan supply effects.

Other work with disaggregated data shows cross-sectional differences among firms involving margins other than inventory investment. As mentioned above, Gertler and Gilchrist find differences in the sales response of large and small firms following a monetary policy shock. Gertler and Hubbard (1988) find differences in the correlation between fixed investment and cash flow for firms that pay dividends and those that do not pay dividends in recessions and normal periods. If we accept a low dividend payout ratio as a proxy for bank dependence and assume that monetary policy shifted prior to the recessions, we can read these results as supporting a bank lending channel.

Focusing on Japanese firms that are not part of bank-centered industrial groups and, therefore, are susceptible to being cut off from bank credit, Hoshi, Scharfstein, and Singleton (1993) find that when monetary policy is tight, liquidity is more important for independent firms' investment than in normal times.

Finally, Sharpe (1994) contrasts the employment adjustment of different sized firms to changes in the real federal funds rate. He finds that small firms' employment is more responsive than that of large firms. Furthermore, firms that are more highly leveraged tend to show greater sensitivity to funds rate shocks. If we assume that more highly leveraged firms are more bank-dependent, this finding is also consistent with the lending channel.

Taken together, these findings strongly support the view that banks play an important role in the transmission of monetary policy. The evidence from different countries, different time periods, and for different agents suggests that 1) restrictive monetary policy reduces loan supply by banks and 2) this reduction in loan supply depresses spending.

Implications for monetary transmission under the EMU

We believe the work reviewed above answers a number of questions about the ways consumers, firms, and banks respond to monetary policy. Furthermore, it implies that the degree of bank dependence in the economy and the extent to which central bank actions move loan supply are the key factors determining the importance of the lending channel. In light of the vast differences in institutions across Europe, this story could have important implications for how monetary policy operates under the EMU.

Consider a uniform tightening of monetary policy. Suppose one country has a set of mostly creditworthy banks and relatively few bank-dependent firms. In this case, the banks may be able to offset the contraction in reserves by picking up uninsured nondeposit financing in the capital markets. Accordingly, bank lending will not fall by much. Moreover, if most firms can continue producing even if some bank loans are cut, the aggregate lending channel effect will be fairly weak.

In a country with many bank-dependent firms and a weak banking system, the impact might be quite different. Banks with poor credit ratings may not be able to attract uninsured funds to offset their deposit outflow. As the banks are driven to cut their lending, their customers will need to find other funding. If this funding is not available in the short run, a sizable spending drop may occur. Thus, a uniform contraction in monetary policy across the two countries may lead to a very asymmetric response, raising potentially problematic distributional issues.

This hypothetical comparison focuses on the differences in the aggregate conditions in the two countries. A key lesson from the work on the U.S. is that the banking-related effects of monetary policy are subtle and that micro-level studies are often required. Nevertheless, in light of the difficulty of getting reliable micro data for a large number of countries, we make an illustrative first pass at the problem with some, admittedly crude, aggregate-level calculations. We infer the degree of bank dependence in different countries by looking at the size distribution of firms and the availability of nonbank finance. To gauge loan supply effects, we study the size distribution of the banking industry and the health of the banks. These are no doubt highly imperfect proxies. We hope this exercise, which we view as a somewhat speculative first step, will spur researchers who have access to better data to build on our results.

Cross-country responsiveness of loan supply to policy changes

Since it is still too early to be certain which countries initially join the monetary union, we work with data for the following countries in the European Union: Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, and the UK. We report similar statistics for the U.S. and Japan, wherever possible.

As mentioned above, Kashyap and Stein (1995) show that small banks are more responsive to monetary tightening than large banks. If bank size is an appropriate proxy for the ability to access noninsured sources of funds, this contrast makes sense (in the context of the lending channel). In some European countries, even large banks may find it difficult to obtain nondeposit financing. We have not been able to find any good data on differences in bank financing options across countries, however, and must therefore rely on size proxies to infer the sensitivity of loan supply to monetary policy.

Our first size distribution indicator (shown in column 4 of table 2) is the three-firm concentration ratio for commercial banks (that is, the share of total commercial bank assets controlled by the three largest commercial banks) as reported by Barth, Nolle, and Rice (BNR, 1997). Although the statistics are a bit dated (from 1993), they cover all of the countries. However, the ratio covers only commercial banks and for some countries, such as Germany, commercial banks are of limited overall importance. The data shown in column 5 of table 2 have been rescaled to correct for this coverage effect; where BNR report the share of commercial bank assets relative to total bank assets, we restate the three-firm concentration ratio in terms of all bank assets.

Even after making this adjustment, looking at only the top three firms may be misleading. For example, consider a country with ten roughly equally sized banks versus a country that has three dominant banks and hundreds of small banks. Depending on the size of the large banks in the second country, small banks might appear to be more or less important than in the first country, even though there may be no small banks in the first country. This problem can occur where there is a sharp discontinuity in the size distribution of banks. To partially address it, we show five- and ten-firm concentration ratios, based on data for 1995 from the Bank for International Settlements (BIS). The BIS data are broader (relative to all banks not just commercial banks) and more current than the BNR measure but, regrettably, they are not available for all countries.

For the most part, the different size distribution statistics paint a similar picture. In Belgium, Netherlands, and the UK, the large banks appear to hold a dominant position. Conversely, Italy, Germany, and Luxembourg stand out as countries in which the smaller banks control a significant fraction of the assets. The limitations of the data preclude drawing any sharp distinctions among the remaining countries.

TABLE 2

Size distribution of banks in selected countries

Country	Range of banks covered by OECD	Banks covered by OECD in 1995	Total assets of 1995 OECD reporting banks (billions U.S. $)[a]	1993 commercial bank assets in the 3 largest commercial banks[b]	1993 assets of all credit institutions in 3 largest commercial banks[b]	1995 assets in 5 largest institutions[c]	1995 assets in 10 largest institutions[c]
				(- - - - - - -	- - - - - - percent	- - - - - - -	- - - - - - -)
Belgium	All banks	143	843.3	44	44	59[e]	73[e]
Denmark	Commercial and savings banks	114	166.7	64	NA	NA	NA
France	All banks	1,453	3797	64	33	47	63
Germany	All banks	3,500	4,151.4	89	24	17	28
Greece	Commercial banks	18	69.7	98	73	NA	NA
Ireland	Commercial banks[d]	434[d]	71.24[d]	94	76	NA	NA
Italy	All banks	269	1,519.7	36	28	29	45
Luxembourg	Commercial banks	220	612.9	17	17	NA	NA
Netherlands	All banks	173	916.5	59	59	81	89
Portugal	All banks	37	201.1	38	NA	NA	NA
Spain	All banks	318	951	50	34	49	62
UK	Commercial banks	40	1,184.1	29	NA	57[e]	78[e]
U.S.	Commercial banks	9,986	4,149.3	13	10	13	21
Japan	Commercial banks	138	6,733.9	28	NA	17[e]	28[e]

[a]Exchange rates are taken from 1996 IMF *Financial Statistics Yearbook*, p. 15. All domestic figures are converted into special drawing rights and then into dollars.
[b]Source is Barth, Nolle, and Rice (1997), table 3.
[c]Source is Bank for International Settlements, *Annual Report* (1996).
[d]These data are for 1993 and are taken from Barth, Nolle, and Rice (1997), table 3.
[e]These data are for 1994.

In addition to the data on bank size, we use a number of measures of bank profitability and capital. In principle, the uninsured liabilities of banks with high levels of capital should have lower credit risk. Thus, well-capitalized (or highly rated) banks should have a much easier time going to securities markets to raise funds in the face of a deposit shock. This implies that monetary policy would have less of an impact when banks are well capitalized.[5] However, for most countries, data on capitalization and creditworthiness are available only for the major institutions; smaller banks tend not to be monitored by the rating agencies that collect most of these statistics. Our benchmark measure of creditworthiness comes from Thomson BankWatch, one of the leading global bank rating agencies. According to its Web page, Thomson constructs ratings which:

"Incorporate a combination of pure credit risk with performance risk looking over an intermediate horizon. These ratings indicate the likelihood of receiving timely payment of principal and interest, and an opinion on the company's vulnerability to negative events that might alter the market's perception of the company and affect the marketability of its securities."[6]

Because these ratings (shown in column 1 of table 3) do not cover all the banks in each country, we supplemented the Thomson data with another measure of bank health. The OECD publishes a stylized income statement for banks in its member countries. The processing lags required to generate comparable data are such that 1995 data are just becoming available. To calibrate the Thomson sample to the broader OECD sample, we calculated the return on average assets (ROA) for both samples. To control for year-to-year volatility, we averaged the numbers over three years and the results are shown in table 3. The ROA estimates from the two sources are very similar. Table 3 also shows loan losses relative to loans from the Thomson data.

Looking across table 3, the countries seem to fall into three fairly distinct groups. The evidence for the first group, Netherlands, Luxembourg, and the UK, suggests that the banks are in good shape. (The U.S. is also in this group.) In the case of the second group, France and Italy, the numbers consistently show that the banking sectors are relatively weak, with high levels of bad loans and low profit rates. (Japan also belongs in this group.) The third group, comprising all remaining countries, falls somewhere in between.

Options for substituting toward nonbank financing

Our first measure of bank dependence is culled from employment data. Using information from the European Commission, we compare the importance of small firms in different countries. The data exclude the self-employed, but include very small firms employing between one and nine people. We believe monitoring costs for these micro firms are likely to be so high that they will have trouble attracting nonbank financing.[7] Because of the processing lags, the data we analyze are from 1990, but a comparison with similar statistics from 1988 suggests that these employment patterns are fairly stable over time.

Table 4 shows that the smallest firms generally account for a larger fraction of employment in Europe than they do in the U.S., although they vary significantly in importance from one European country to another. In Spain and Italy, more than 40 percent of the work force is attached to these firms, while in Belgium, Germany, and Luxembourg, they are of much more limited significance.[8] Similar heterogeneity exists for mid-sized and large firms.

The last column in table 4 reports the ratio of each country's share of total European employment to its share of the total number of enterprises. A ratio of one would be the typical size distribution for European countries. Ratios below one indicate a preponderance of smaller firms, while ratios above one indicate more larger firms.

Again, these data can be used to sort the countries into three categories. In Greece, Italy, Netherlands, Portugal, and Spain, smaller firms are most important. Germany, Luxembourg, and the UK are dominated by larger firms, with employment distributions that look much more like those of the U.S. The remaining cases are not clear cut.

The second indicator of bank dependence is based on the structure of capital markets across Europe. Ideally, we would like to have a measure of the switching costs firms would incur if they lost their bank financing. We would not expect these firms to be able to issue publicly traded securities directly. However, through trade credit, they may have access to funds raised in the securities markets (see Calomiris, Himmelberg, and Wachtel, 1995). Similarly, although equity financing is rarely an important source of funding for most firms, deep equity markets are often correlated with the existence of other public markets that might be tapped when bank credit contracts.[9]

Accordingly, table 5 provides information from the World Bank on stock market capitalization across different countries. The table also shows OECD data on the public bond markets for each country. However, these data are only for firms listed on the specific exchanges shown in the table and, in some cases, this significantly understates the size of the bond market (for example, in the U.S. where only bonds of the NYSE firms are counted). The bottom line is in the last two columns of the table, which show the ratio of stock market capitalization to gross domestic product (GDP) and the ratio of public bonds to GDP. Subjectively weighting these two measures, we conclude that the availability of nonbank finance is greatest in Belgium, Denmark, and the UK. Conversely, Greece, Italy, and Portugal appear to be the least developed by this metric.

TABLE 3

Bank health in selected countries

Country	Fiscal 1995 Thomson average rating of tracked banks (no. of banks)[a]	Thomson estimated ROA for major banks, 1993–95 (average no. of major banks)	OECD profit before tax relative to assets, 1993–95 (average no. of rated banks)	1995 Thomson estimated loan losses relative to loans for major banks (no. of major banks)
Belgium	B (8)	0.28 (54)	0.23 (147)	NA (NA)
Denmark	B/C (3)	0.55 (74)	0.52 (113)	0.91 (86)
France	B/C (22)	0.15 (298)	0 (1,569)	2.56 (269)
Germany	B/C (24)	.22 (205)	0.26 (3,627)	0.17 (204)
Greece	B (9)	0.39 (22)	0.84 (19)	0.57 (23)
Ireland	B (3)	1.03 (29)	NA (NA)	0.78 (28)
Italy	C (30)	–0.01 (57)	0.11 (296)	7.47 (57)
Luxembourg	B (3)	0.6 (128)	0.36 (220)	0.14 (127)
Netherlands	A/B (3)	0.57 (52)	0.5 (174)	NA (NA)
Portugal	B/C (4)	0.46 (48)	0.62 (36)	3.61 (46)
Spain	B/C (14)	0.20 (101)	0.45 (317)	4.09 (105)
United Kingdom	B (25)	1.84[b] (6)	0.67 (38)	1.21[b] (6)
United States	B (29)	1.23 (29)	1.18 (10493)	0.74 (29)
Japan	C (10)	–0.06[c] (10)	–0.07 (139)	3.96[c] (10)

[a]Thomson normally requires banks to pay to be evaluated. In some cases struggling banks decide not pay for the rating but Thomson assigns a rating anyway (although it may not store all of the financial information for these banks). The country averages pertain to all banks for which a rating was assigned.

The Thomson rating scale is as follows:

A—Company possesses an exceptionally strong balance sheet and earnings record, translating into an excellent reputation and very good access to its natural money markets. If weakness or vulnerability exists in any aspect of the company's business, it is entirely mitigated by the strengths of the organization.
A/B—Company is financially very solid with a favorable track record and no readily apparent weakness. Its overall risk profile, while low, is not quite as favorable as for companies in the highest rating category.
B—Company is strong with a solid financial record and is well received by its natural money markets. Some minor weaknesses may exist, but any deviation from the company's historical performance levels should be limited and short–lived. The likelihood of significant problems is small, yet slightly greater than for a higher rated company.
B/C—Company is clearly viewed as a good credit. While some shortcomings are apparent, they are not serious and/or are quite manageable in the short term.
C—Company is inherently a sound credit with no serious deficiencies, but financial statements reveal at least one fundamental area of concern that prevents a higher rating. Company may recently have experienced a period of difficulty, but those pressures should not be long term in nature. The company's ability to absorb a surprise, however, is less than that for organizations with better operating records.
C/D—While still considered an acceptable credit, the company has some meaningful deficiencies. Its ability to deal with further deterioration is less than that of better rated companies.
D—Company financials suggest obvious weaknesses, most likely created by asset quality considerations and/or a poorly structured balance sheet. A meaningful level of uncertainty and vulnerability exists going forward. The ability to address further unexpected problems must be questioned.
D/E—Company has areas of major weakness that may include funding and/or liquidity difficulties. A high degree of uncertainty exists about the company's ability to absorb incremental problems.
E—Very serious problems exist for the company, creating doubt about its continued viability without some form of outside assistance, regulatory or otherwise.

[b]United Kingdom data are averaged for two years only.

[c]Japanese data cover fiscal years 1995 through 1997.

Predicted potency of the lending channel under the EMU

Given the noisy nature of our data, it is not possible to make strong claims about how important the lending channel might be in different countries. However, we believe the proxies reviewed above provide some interesting information, particularly at the extremes of their respective distributions. To summarize these results, we assigned each country a letter grade (from A to C) for each of our four factors. A grade of "A" indicates the *least* sensitivity to monetary policy.

Table 6 shows these grades and an overall grade (shown in the last column) based on a subjective weighting of the factors. The UK emerges as the country for which the evidence most clearly suggests a relatively weak lending channel. UK banks are in relatively good

TABLE 4

1990 size distribution of employment in selected countries

	% of total Euro 12 employment	% in firms with fewer than 10 people	% in firms with 10–499 people	% in firms with 500+ people	% of total enterprises in Euro 12	Ratio of share of employment to share of enterprises
EURO 12	100	30.3	39.4	30.3	100	1.00
Belgium	3.0	17.0	47.7	35.3	3.5	0.86
Denmark	1.8	31.6	49.1	19.3	1.8	1.00
France	15.5	28.0	41.0	31.0	13.9	1.12
Germany	23.2	18.3	45.6	36.1	14.8	1.57
Greece	NA	NA	82.7	17.3	NA	NA
Ireland	0.25	NA	NA	NA	.072	3.46
Italy	15.7	42.5	37.8	19.7	21.5	0.73
Luxembourg	0.2	15.1	40.6	25.5	0.1	2.00
Netherlands	NA	30.1	45.4	24.5	NA	NA
Portugal	3.0	24.3	54.7	21.0	4.2	0.71
Spain	10.5	45.8	38.9	15.3	17.0	0.62
United Kingdom	20.9	27.1	39.1	33.8	17.2	1.22
United States	107	12.0	41.4	46.6	NA	NA

Notes: Greek data only cover NACE 1–4 and 67; employment figures only cover establishments with an average of 10 or more employees. Irish data only cover enterprises in NACE 1–4 averaging 3 employees or more and NACE establishments averaging 20 employees or more. Data are reported for 3–19 employees or 20 plus employees. NA indicates not available.
Source: Commission of the European Communities, *Enterprises in Europe, Third Report*, Brussels, Belgium (1994).

shape, there are not a lot of small firms, and firms have many other financing options. Belgium and Netherlands also appear to be on the relatively insensitive end of the spectrum. Netherlands has large, creditworthy banks, and Belgium appears to be in moderately good shape in terms of both loan supply sensitivity and bank dependence.

At the opposite end of the spectrum, Italy is clearly the country in which we would expect strong effects of monetary policy, based on each of the factors we have studied. Portugal also fits into this part of the distribution.

In the remaining countries, the picture is less clear. For example, in Germany and Luxembourg there are many small banks, but bank health appears at least adequate and large firms are relatively important. Our data are not sufficiently precise to identify more than the extreme cases.

Conclusions

Research strongly suggests that banks play a role in the transmission of monetary policy. The factors that determine the significance of this role are the degree of bank dependence on the part of firms and consumers and the ability of banks to offset monetary-policy-induced deposit outflows. Based on the best available data, we find considerable differences in these dimensions across member countries of the European Union.

When it goes into effect, the EMU may provide answers to key questions regarding the potency of the bank lending channel. Given the wide heterogeneity in bank health, a sudden shift in monetary conditions (such as an increase in interest rates by the European central bank) would provide a live test of this mechanism. In the meantime, our research suggests that it would be desirable to consider integration in banking and securities markets in tandem with the move to a single currency. European

TABLE 5

Nonbank financing options

Country	Stock exchange tracked by World Bank	1995 listed firms on exchange tracked by World Bank	1995 market capitalization (world rank)	1995 GDP	Exchange for bond market data	Public bonds of traded firms	Equity value as a % of GDP	Public bonds as a % of GDP
			(- - - - -U.S. $ billions- - - - -)		*(year)*	*(U.S. $ billions)*		
Belgium	Brussels	143	104.96 (22)	269.2	Brussels (1995)	235.0	0.39	0.87
Denmark	Copenhagen	213	56.22 (27)	175.2	Copenhagen (1995)	301.1	0.32	1.72
France	Paris	450	522.05 (5)	1,549.2	Paris (1993)	662.9	0.34	0.43
Germany	German Stock Exchange Inc.	678	577.37 (4)	2,420.5	Frankfurt (1995)	1,223.8	0.24	0.51
Greece	Athens	99	10.16 (NA)	111.8	Athens (1989)	17.5	0.09	0.16
Ireland	Irish Stock Exchange	80	25.82 (37)	60.1	Not shown separately	NA	0.43	NA
Italy	Italian Stock Exchange Council	250	209.52 (13)	1,091.1	Milan (1994)	760.5	0.19	0.70
Luxembourg	Luxembourg	61	30.44 (36)	16.8	Luxembourg (1989)	1.5	1.81	0.09
Netherlands	Amsterdam	387	356.48 (8)	396.9	Amsterdam (1995)	0.294	0.90	0.00
Portugal	Lisbon	169	18.36 (39)	103.2	NA	NA	0.18	NA
Spain	Madrid	362	197.79 (14)	557.4	Madrid (1995)	27.7	0.35	0.05
United Kingdom	London	2,078	1,407.74 (3)	1,099.7	Ireland and UK (1993)	554.4	1.28	0.50
United States	Combined NYSE, AMEX, NASDAQ	7,671	6,857.62 (1)	6,981.7	New York (1995)	2,495.9	0.98	0.36
Japan	Combined all major exchanges[a]	2,263	3,667.29 (2)	4,960.7	Tokyo (1994)	1,789.6[b]	0.74	0.36

[a]The Japanese exchanges include Fukoka, Hiroshima, Kyoto, Nagoya, Niigata, Osaka, Sapporo, and Tokyo.
[b]Japanese bond data cover both domestic and foreign firms.

Sources: *Emerging Stock Markets Factbook*, International Finance Corporation (1996); OECD, "OECD financial statistics, part 1," *Financial Statistics Monthly*, various issues, and OECD, *Non-Financial Enterprises Financial Statements* (1995).

TABLE 6

Summary of factors affecting the lending channel

Country	Importance of small banks	Bank health	Importance of small firms	Availability of nonbank finance	Overall predicted potency
	(Table 2)	(Table 3)	(Table 4)	(Table 5)	
Belgium	A	B	B	A	A/B
Denmark	B	B	B	A	B
France	B	C	B	B	B/C
Germany	C	B	A	B	B
Greece	B	B	C	C	B/C
Ireland	B	B	B	B	B
Italy	B	C	C	C	C-
Luxembourg	C	A	A	B	B
Netherlands	A	A	C	B	A/B
Portugal	B	C	C	C	C
Spain	B	B	C	B	B
United Kingdom	A	A	A	A	A

Note: A grade of "A" indicates low effect of lending channel sensitivity to monetary policy; "C" indicates high sensitivity.

banking regulations have officially been harmonized for several years. However, the health of the banking system varies significantly from one country to another, and few banks have begun lending outside their own borders. Countries with weak banking systems might benefit from the entry of foreign banks into their markets. The development of deeper securities markets that would be available to all European firms could also help offset a potential credit crunch.

APPENDIX

The data shown in table 1 and used in Kashyap and Stein (1995 and 1997) are taken from the quarterly regulatory filings made by all U.S. commercial banks. These reports, commonly referred to as Call Reports, contain detailed quarterly balance sheet and income statement data for all banks. In addition to this basic information, the reports contain data on a variety of off-balance-sheet items, a special supplement on small business lending that is collected as part of the June Call Report, geographic information, and the holding company status of the banks.

The Federal Reserve Bank of Chicago is now making the most popular items from the Call Reports available through its Web site. Initially, the post-1990 data will be available; eventually data going back to 1976 will be online. The data for each quarter are stored in a SAS transport data set, which has been compressed in a zip format. The zipped files are typically 4.5 megabytes and expand to about 48 megabytes when they are uncompressed. It took us about 15 minutes to download the 1995 fourth quarter file in our tests. You can access the data at www.frbchi.org/rcri/rcri_database.html. (The site also shows current reporting forms filled out by the banks.)

To supplement the raw Call Report data, the Bank's research staff is making a file available that lists all the mergers between U.S. commercial banks from 1976 onward. This merger file can easily be combined with the Call Report data for a number of projects, for example, an event study analysis. We have used the file to screen out banks for which mergers make the accounting statements discontinuous.

The Bank's Web site also contains a simple data access program. This program allows a user to create consistent time series for several of the major items on the banks' balance sheets. Similarly, there is documentation describing the known breaks in all of the series.

A picture of the Web site appears on the following page.

Go To: http://www.frbchi.org/rcri/rcri_database.html

Report of Condition and Income Database

The Report of Condition and Income database contains selected data for all banks regulated by the Federal Reserve System, Federal Deposit Insurance Corporation, and Comptroller of the Currency. The financial data are on an individual bank basis and were selected from the following schedule: assets and liabilities, income, capital, off-balance-sheet transactions, risk-based capital, and other memoranda items. Files are available quarterly and only for downloading purposes.

About the Data

Documentation files:

Data Description contains a list of all the variables in this database.

Data Definitions contains the definitions of the variables and notes on forming consistent time series.

Data Access contains information on how to import the zipped SAS files into various software packages and a sample SAS program.

Sample Form shows the reporting form currently used to collect the data.

Merger Data

The merger file contains information that can be used to identify all bank acquisitions and mergers since 1976. These data can be merged into the Call Report data.

Quarterly Call Report Data

Each quarterly data file contains income and balance sheet items for all the banks. The files are zipped using PKZIP. The files are in SAS transport data file format. The files are about 4.5 megabytes in compressed form and about 48 megabytes when expanded.

Year	1st quarter	2nd quarter	3rd quarter	4th quarter
1990	1st	2nd	3rd	4th
1991	1st	2nd	3rd	4th
1992	1st	2nd	3rd	4th
1993	1st	2nd	3rd	4th
1994	1st	2nd	3rd	4th
1995	1st	2nd	3rd	4th
1996	1st	2nd	3rd	4th

Document: Done

NOTES

[1]Throughout all of what follows we are implicitly relying on the conventional assumption that there is imperfect price adjustment. See Bernanke and Gertler (1995), Cecchetti (1995), Hubbard (1995), and Kashyap and Stein (1994) for other surveys of this literature.

[2]See Diamond (1984) for a formal treatment of this problem.

[3]See Borio (1996) and Berran, Coudert, and Mojon (1996) for two exceptions.

[4]A finding that these forms of financing move in opposite directions following a monetary contraction should not be taken as an indication that those firms that are cut off from banks are the same ones that begin issuing commercial paper. A much more realistic mechanism is that smaller firms that are cut off from bank lending receive increased trade credit, and the trade credit is supplied by larger firms that can access the commercial paper market.

[5]For the U.S., Kashyap and Stein (1994) note that things may have worked differently in the credit crunch of the early 1990s. If a regulatory risk-based capital standard binds banks at the margin, then the banks' loan supply can become disconnected from changes in monetary policy. In this case, the binding capital requirement can generate a "pushing on a string" problem for the central bank, in which monetary policy becomes *less* effective.

[6]Description of Thomson issuer ratings from www.bankwatch.com, as of July 17, 1997. We thank Christopher Tang for supplying the BankWatch data and answering our questions about them.

[7]One caveat to this assumption is that if firms are part of a holding company structure that creates the appearance of many small firms in order to skirt certain regulations, then it is possible that these firms may have access to the internal capital market of the holding company.

[8]Of course, the Gertler and Gilchrist numbers shown earlier demonstrate that small firms generally may account for a much larger fraction of fluctuations than suggested by their average share of the aggregate economy.

[9]For example, Demirgüç-Kunt and Levine (1996) show that stock market capitalization tends to be fairly highly correlated with the ratio of domestic credit to GDP.

REFERENCES

Bank for International Settlements, *Annual Report*, Basel, Switzerland: BIS, 1996.

Barth, James R., Daniel E. Nolle, and Tara N. Rice, "Commercial banking structure, regulation, and performance: An international comparison," Office of the Comptroller of the Currency, economics working paper, No. 97-6, 1997.

Berger, Allen N., Anil K Kashyap, and Joseph M. Scalise, "The transformation of the U.S. banking industry: What a long, strange trip it's been," *Brookings Papers on Economic Activity*, 1995, pp. 55–218.

Bernanke, Ben S., and Alan S. Blinder, "The federal funds rate and the channels of monetary transmission," *American Economic Review*, Vol. 82, September 1992, pp. 901–921.

Bernanke, Ben S., and Mark Gertler, "Inside the black box: The credit channel of monetary policy transmission," *Journal of Economic Perspectives*, Vol. 9, Fall 1995, pp. 27–48.

Berran, Fernando, Virginie Coudert, and Benoit Mojon, "The transmission of monetary policy in European countries," Centre D'Études Prospectives et D'Informations Internationales, working paper, 1995.

Borio, Claudio, "Credit characteristics and the monetary transmission mechanism in fourteen industrial countries: Facts, conjectures, and some econometric evidence," in *Monetary Policy in a Converging Europe*, Koos Alders, Kees Koedijk, Clemens Kool, and Carlo Winder (eds.), Amsterdam: Kluwer Academic Press, 1996, pp. 77–115.

Calomiris, Charles W., Charles P. Himmelberg, and Paul Wachtel, "Commercial paper, corporate finance, and the business cycle: A microeconomic perspective," *Carnegie-Rochester Conference Series on Public Policy*, Vol. 42, 1995, pp. 203–250.

Carpenter, Robert E., Steven M. Fazzari, and Bruce C. Petersen, "Inventory investment, internal-finance fluctuations, and the business cycle," *Brookings Papers on Economic Activity*, 1994, pp. 75–138.

Cecchetti, Stephen G., "Distinguishing theories of the monetary transmission mechanism," *Review*, Federal Reserve Bank of St. Louis, Vol. 77, May/June 1995, pp. 83–97.

Demirgüç-Kunt, Asli, and Ross Levine, "Stock market development and financial intermediaries: Stylized facts," *The World Bank Economic Review*, Vol. 10, No. 2, 1996, pp. 291–321.

Diamond, Douglas W., "Financial intermediation and delegated monitoring," *Review of Economic Studies*, Vol. 51, No. 166, July 1984, pp. 393–414.

Gertler, Mark, and Simon Gilchrist, "Monetary policy, business cycles, and the behavior of small manufacturing firms," *Quarterly Journal of Economics*, Vol. 109, May 1994, pp. 309–340.

Gertler, Mark, and R. Glenn Hubbard, "Financial factors in business fluctuations," *Financial Volatility: Causes, Consequences, and Policy Recommendations*, Washington, DC: Federal Reserve Board, 1988.

Gibson, Michael, "The bank lending channel of monetary policy transmission: Evidence from a model of bank behavior that incorporates long-term customer relationships," Federal Reserve Board, working paper, 1996.

Hoshi, T., David Scharfstein, and Kenneth J. Singleton, "Japanese corporate investment and Bank of Japan guidance of commercial bank lending," in *Japanese Monetary Policy*, Kenneth J. Singleton (ed.), NBER, 1993, pp. 63–94.

Hubbard, R. Glenn, "Is there a 'credit' channel for monetary policy?," *Review*, Federal Reserve Bank of St. Louis, Vol. 77, May/June 1995, pp. 63–77.

International Finance Corporation, *Emerging Stock Markets Factbook*, Washington, DC, 1996.

Kashyap, Anil K, and Jeremy C. Stein, "Monetary policy and bank lending," in *Monetary Policy*, N. Gregory Mankiw (ed.), Chicago: University of Chicago Press, 1994, pp. 221–256.

__________, "The impact of monetary policy on bank balance sheets," *Carnegie-Rochester Conference Series on Public Policy*, Vol. 42, 1995, pp. 151–195.

__________,"What do a million banks have to say about the transmission of monetary policy," National Bureau of Economic Research, working paper, No. 6056, June 1997.

Kashyap, Anil K, Jeremy C. Stein, and Owen Lamont, "Credit conditions and the cyclical behavior of inventories," *Quarterly Journal of Economics*, Vol. 109, August 1994, pp. 565–592.

Kashyap, Anil K, Jeremy C. Stein, and David W. Wilcox, "Monetary policy and credit conditions: Evidence from the composition of external finance," *American Economic Review*, Vol. 83, March 1993, pp. 78–98.

__________, "Monetary policy and credit conditions: Evidence from the composition of external finance: Reply," *American Economic Review*, Vol. 86, March 1996, pp. 310–314.

Ludvigson, Sydney, "The channel of monetary transmission to demand: Evidence from the market for automobile credit," Federal Reserve Bank of New York, working paper, No. 9625, 1996.

Milne, Alistair, "Inventory investment in the UK: Excess volatility, financial effects, and the cost of capital," University of London, unpublished Ph.D. thesis, 1991.

Organization for Economic Cooperation and Development, *Bank Profitability: Financial Statements of Banks 1986–1995* (preliminary), Paris: OECD, 1997.

________, *Non-Financial Enterprises Financial Statements*, Paris, 1995.

________, "OECD financial statistics, part 1," *Financial Statistics Monthly*, Paris, various issues.

Sharpe, Steven A., "Bank capitalization, regulation, and the credit crunch: A critical review of the research findings," Federal Reserve Board, working paper, No. 95-20, 1995.

Stein, Jeremy C., "An adverse selection model of bank asset and liability management with implications for the transmission of monetary policy," National Bureau of Economic Research, working paper, No. 5217, 1995.

Understanding aggregate job flows

Jeffrey R. Campbell and Jonas D.M. Fisher

Recent empirical work on plant-level employment dynamics, described in Davis, Haltiwanger, and Schuh (DHS, 1996), represents a challenge to conventional ways of thinking about business cycles. The plant-level data provide concrete evidence against the broad applicability of the representative agent construct. Moreover, the behavior of the macro aggregates based on the plant-level data seem hard to reconcile with predictions of the models that dominate the literature on business cycles, which are based on the representative agent paradigm.

Although DHS present evidence at the micro and aggregate levels, most of the literature that has developed in response has focused on the aggregate-level evidence. Two of the aggregate variables that have attracted the most attention are the rates of job creation, that is, positive plant-level employment growth, and job destruction, that is, negative plant-level job growth. DHS find that the variance of job destruction in the U.S. manufacturing sector is greater than the variance of job creation and that these variables are negatively correlated (albeit imperfectly).

A variety of models have been developed to explain the above observations, which are difficult to reconcile with standard representative-agent models of the business cycle. Examples include Caballero (1992), Caballero and Hammour (1994), Foote (1995), and Mortenson and Pissarides (1994). While this work has provided important insights into business cycles, for the most part it does not simultaneously account for the significant heterogeneity in the intensity of job growth at the plant level documented in DHS. Thus, it does not bring us any closer to establishing a direct connection between detail at the micro level and the behavior of important macro aggregates.

In Campbell and Fisher (1996), we present a model that has the potential of accounting for both the aggregate and the cross-sectional evidence. We believe that knowledge of the microeconomic decision rules suggested by the plant-level employment data enhances our understanding of the aggregate evidence. A significant feature of the plant-level employment data is that large numbers of plants do not change employment over a quarter or even a year, and there is considerable heterogeneity among plants that do change, with changes occurring over a fairly broad range. These results suggest a microeconomic interpretation: that plants face idiosyncratic uncertainty and employment adjustment costs which are non-differentiable at the point of zero change. This structure captures the qualitative features of the cross-sectional evidence. Moreover, we find that the same friction that underlies the adjustment cost formulation may imply that average

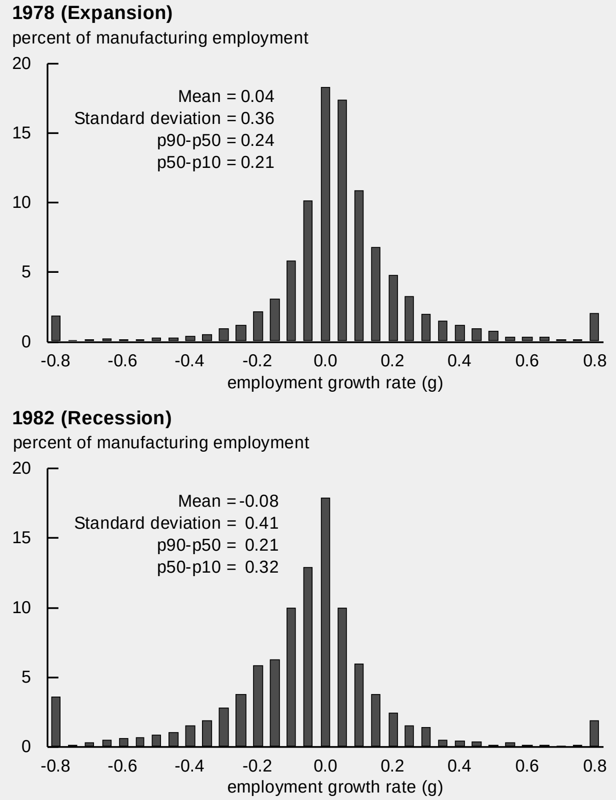
Jeffrey R. Campbell is an assistant professor of economics at the University of Rochester and a faculty research fellow at the National Bureau of Economic Research. Jonas D.M. Fisher is an economist at the Federal Reserve Bank of Chicago and an assistant professor at the University of Western Ontario.

job destruction by plants that reduce employment is more variable than average job creation by plants that increase employment. This helps us account for the aggregate evidence on employment flows. That is, we are able to establish a *direct* connection between micro and aggregate fluctuations.

In this article, we review our work in Campbell and Fisher (1996). We describe the evidence in DHS that has generated the recent theoretical interest, discuss the reasons this evidence represents a challenge to standard models, and briefly outline the recent theoretical responses to this challenge. We develop a benchmark model that captures key features of standard business cycle theory, which we use to demonstrate the difficulties standard models have in accounting for the DHS evidence. We then use a model based on Caballero (1992) to demonstrate the main mechanism at work in our model.

FIGURE 1

Plant-level employment growth-rate distributions: 1978 and 1982



Notes: (p90–p50) is the 90th percentile minus the 50th employment percentile. (p50–p10) is the 50th percentile minus the 10th employment percentile. The growth-rate distributions show the number of occurrences of each observed employment rate weighted by each plant's employment. The bars thus indicate the share of employment associated with each rate. In this figure, the growth rate, *g*, is measured as the change in employment divided by the average of current and lagged employment. (See technical appendix in source publication.)

Source: Davis, Haltiwanger, and Schuh (1996)

Implications of evidence on job flows for business cycle analysis

The evidence presented in DHS is based on the Longitudinal Research Database compiled by the U.S. Bureau of the Census. This database contains detailed quarterly and annual plant-level employment data for the U.S. manufacturing sector from 1972 to 1988. First, we describe the evidence on job flows at the plant level. Second, we describe various aggregate variables which are based on the plant-level data.[1] Finally, we discuss how some of this evidence represents a challenge to conventional ways of modeling business cycles and review leading theoretical responses to this challenge.

Evidence on plant-level heterogeneity in job growth

Figure 1 displays two snapshots of employment growth for the U.S. manufacturing sector. DHS measure date *t* employment growth at the plant level as the change in employment between date *t*–1 and date *t* divided by the average of date *t* and *t*–1 employment. Formally,

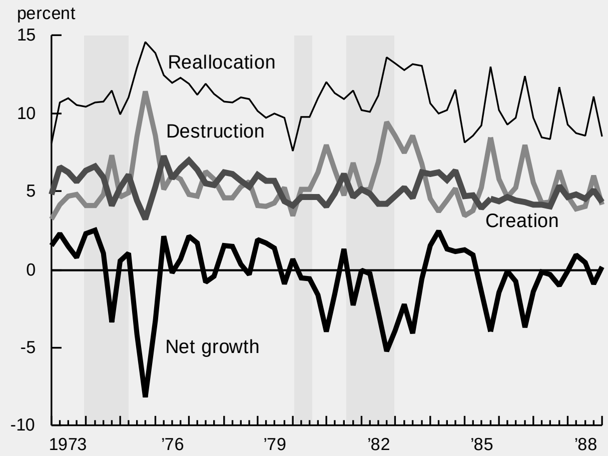
$$\text{employment growth at plant } i = \frac{(n_{i,t} - n_{i,t-1})}{(n_{i,t} + n_{i,t-1})\,/2},$$

where $n_{i,t}$ denotes the level of employment at plant *i* at date *t*. Both panels in the figure display cross-sectional histograms of employment growth, where individual plant-level employment growth rates are weighted by the plant's share of total employment. Hence, the height of a bar is the percentage of total employment accounted for by plants within the growth rate interval on the horizontal axis. The top panel shows the employment-weighted cross-sectional distribution of plant-level employment growth rates for 1978, an expansion year, and the bottom panel shows the same for 1982, a recession year.

As the histograms illustrate, job creation and job destruction are pervasive. Moreover, the scale of employment changes at the plant level displays considerable heterogeneity. Further, as we would expect, changing from an expansion to a recession involves a drop in the mean of the job-growth distribution (see panel inset). Notice that the recession distribution appears more skewed to the left (toward destruction) than seems warranted by a change in the mean of the distribution alone. Indeed, the variance of the distribution increases in a recession relative to a boom. Finally, both panels show that a large fraction of employment is at plants that do not change employment or change employment by a very small amount.[2]

FIGURE 2

Aggregate job flows, 1972:Q4–1988:Q4



Note: Shaded areas indicate recessions.
Source: Davis, Haltiwanger, and Schuh (1996).

Evidence on aggregate job flows

Figure 2 plots quarterly data for aggregate job creation, job destruction, job reallocation, and net job growth or the difference between job creation and destruction from the fourth quarter of 1972 through the fourth quarter of 1988. Due to the non-stationarity in the levels of these variables, the data plotted are rates and not levels. DHS define the aggregate *rate of job creation* at date *t* as total job creation between dates *t*–1 and *t* divided by the average of current and lagged aggregate employment:

$$\text{aggregate rate of job creation at date } t = \frac{\sum_{\{i:n_{i,t} > n_{i,t-1}\}} (n_{i,t} - n_{i,t-1})}{(n_t + n_{t-1})/2}.$$

Here, n_t denotes aggregate employment at date *t*. Similarly, the aggregate *rate of job destruction* at date *t* is defined as total job destruction between date *t*–1 and *t* divided by the average of current and lagged aggregate employment:

$$\text{aggregate rate of job destruction at date } t = \frac{\sum_{\{i:n_{i,t} < n_{i,t-1}\}} (n_{i,t} - n_{i,t-1})}{(n_t + n_{t-1})/2}.$$

According to figure 2, job destruction is clearly more variable than job creation. Destruction in particular tends to rise sharply around times of recessions (shaded areas in figure). Although there is some negative covariation between job creation and destruction, it is not perfect. Job destruction seems to be quite cyclical, while job creation seems virtually acyclical. Finally, both reallocation and net job growth appear quite cyclical, moving in opposite directions over the business cycle.

Another way of looking at time series data is to examine summary statistics derived from the data. Various statistics summarizing the cyclical characteristics of the aggregate variables plotted in figure 2 are displayed in table 1, with standard errors in parentheses.[3] These confirm our main impressions from figure 2. Note that the variance of job creation is less than one-third that of job destruction, and the difference is significant at the 1 percent level. Note also that creation and destruction are significantly negatively correlated, but the absolute value of the correlation is significantly different from unity. Another feature of the

TABLE 1

Cyclical characteristics of quarterly job flows, U.S. manufacturing sector, 1972:Q4–1988:Q4

Variances	
Creation	0.79 (0.10)
Destruction	2.70 (0.60)
Reallocation	1.52 (0.12)
Growth	2.15 (0.24)
Correlations	
Creation and growth	0.72 (0.05)
Destruction and growth	–0.93 (0.02)
Reallocation and growth	0.58 (0.09)
Creation and destruction	–0.40 (0.11)

Source: Authors' calculations based on data in figure 2.

data is that reallocation and net job growth display a significant negative correlation. This evidence of "countercyclical job reallocation" has been the focus of a lot of theoretical attention. (However, it is logically indistinct from the observation that destruction is more variable than creation. This follows from the definitions of job reallocation and net job growth and the definition of a covariance.)

***Challenging the conventional view*[4]**

The evidence presented above represents a challenge to conventional approaches to modeling business cycles. Of particular relevance are the following observations: 1) plant-level job creation and destruction display considerable heterogeneity (including many plants that do not change employment for extended periods) and are ongoing phenomena that occur at all stages of the business cycles; 2) the variance of the cross-sectional employment growth distribution rises in a recession; 3) aggregate job destruction is more variable than aggregate job creation (or aggregate job reallocation is countercyclical); and 4) aggregate job creation and job destruction are negatively correlated, albeit imperfectly.

Standard business cycle models are built around three main tenets: representative agents, symmetric aggregate shocks, and frictionless markets. Aggregate variables are considered to be determined by the optimal decisionmaking of a representative household and a representative firm, each subject to random disturbances. These agents are assumed to interact in competitive goods and factor markets. The representative agent assumption is valid in these models because all households and firms behave identically. The random disturbances are shocks that disturb the economy as a whole. Examples used in recent business cycle studies include government spending shocks, technology shocks, monetary policy shocks, or shocks to marginal tax rates.

Standard models with these features have difficulty with the evidence summarized in the four observations above. First, standard models do not exhibit any heterogeneity at the plant level. All firms are identical and behave exactly as the representative firm. When employment at the representative firm changes, it changes by the same amount at all firms. Thus these models are unable, at first glance at least, to account for the heterogeneity observation. Second, since creation and destruction are not pervasive at the plant level in these models, they cannot account for the variance of the cross-sectional employment growth distribution in recessions compared with periods of economic growth. Third, with symmetric aggregate shocks, aggregate job creation and aggregate job destruction at the representative firm occur with similar frequency and magnitude. Therefore, aggregate job creation and destruction are equally variable, which contradicts the third observation above. Finally, because all firms act identically, when these models display aggregate job creation, aggregate job destruction must be zero and vice versa. Given the assumption of symmetric aggregate shocks, it follows that aggregate job creation and destruction are perfectly negatively correlated in these models, so they fail to account for the evidence of imperfect negative correlation.

Recent responses to the challenge

As mentioned earlier, most of the literature has focused on the aggregate-level evidence, in particular the evidence of greater variability in aggregate job destruction relative to aggregate job creation (observation number three in the previous section). We have taken the response a step further by attempting to make a direct connection between the micro- and aggregate-level evidence. Our work shows that the same friction that can help account for the plant-level data also helps to account for the evidence on aggregate job flows. To clarify these points, we briefly summarize the recent literature.

In the model developed by Caballero and Hammour (1994), aggregate disturbances influence the incentives to create and destroy plants. These disturbances affect the rate at which new vintages of capital render older vintages obsolete and so determine the rates at which plants are created and destroyed. Since it is assumed that a fixed number of workers is used to operate a plant, variation in the numbers of plants being shut down or coming on line translates directly into numbers of jobs destroyed or created. Caballero and Hammour account for the relative variability of creation and destruction by introducing a friction into the process of plant creation. In particular, they assume that costs of plant creation are increasing in aggregate creation activity, but that destruction costs are not.

Mortensen and Pissarides (1994) develop a model in which the key departure from the conventional model is that the labor market is no longer frictionless. In their model, production takes place at plants in which one worker operates one unit of capital. Workers are matched with plants and sometimes these matches are broken, in which case it takes time for new job–worker matches to be formed. Measured variation in employment occurs as the number of plants matched with a worker varies over time. If a match is broken, a job is said to be destroyed; if a match is formed, a job is created. Variation in the number of new job–worker matches or new job–worker separations translates directly into measures of aggregate creation and destruction. In this model, periods of low aggregate productivity are also periods in which the opportunity costs of reallocating workers are low. Hence, reallocation activity tends to be high in recessions relative to booms and, therefore, destruction is more variable than creation.

Caballero (1992) studies a model of lumpy employment adjustment. Fixed costs of adjustment prevent employment from being always at its frictionless optimum level, as in conventional models. If employment falls below a threshold relative to the frictionless optimum, the plant increases employment by a fixed amount; if employment exceeds some threshold relative to the frictionless optimum, the plant reduces employment by a fixed amount. Aggregate disturbances influence the distribution of plants relative to their frictionless optimum levels, leading to variablitity in aggregate creation and destruction. Caballero demonstrates that if the aggregate disturbances are symmetric, movements in the numbers of creators and destroyers are such that the variance of creation equals the variance of destruction, regardless of the amounts created and destroyed by individual plants. If, on the other hand, the aggregate shocks are assumed to be asymmetric, the author shows that it is possible to reproduce the excess variability of destruction found in DHS. In particular, if *bad* shocks are more severe but occur less frequently than *good* shocks, there is a tendency for the variance of the number of job destroyers to exceed the variance of the number of job creators.

Foote (1995) presents another explanation for the empirical evidence that builds on the same basic structure studied by Caballero (1992). This analysis also focuses on generating movements in the numbers of job creators and job destroyers, holding fixed the amounts created and destroyed by individual plants. The mechanism emphasized by Foote involves the trend downward in average plant size in the U.S. manufacturing sector over the sample period studied by DHS. The downward trend is modeled in terms of a trend downward in the frictionless level of employment at the plant level. This tends to lead to the bunching of plants near their job destruction thresholds, which means that bad aggregate shocks have a larger impact on job destruction than good shocks have on job creation. The net result is higher variation in job destruction than in job creation, driven entirely by variation in the numbers of job creators and job destroyers.

Although the above models achieve some success in providing a theoretical grounding for the DHS evidence on aggregate employment flows, they leave the plant-level evidence largely unexplained. In these models, there is no heterogeneity in creation and destruction at the plant level, and the amounts created and destroyed at the plant level are invariant over the business cycle. All variation in aggregate creation and destruction is derived from model features that influence the numbers of plants creating and destroying. Our contribution is to show how the same friction that helps to account for the plant-level evidence may also imply variation in the amounts created and destroyed at the plant level, which in turn may account for the evidence on aggregate job flows.

In our model, plants are subject to idiosyncratic technology shocks and we assume that it is costly to adjust employment at the plant level, with these costs being nondifferentiable at the point of zero adjustment. In the following two sections, we illustrate the potential of these model elements to simultaneously account for the micro and aggregate evidence. Below, we present a benchmark macro model without employment adjustment costs, but with idiosyncratic uncertainty at the plant level. This illustrates how minor modifications to a standard model can help it account for some of the plant-level evidence. However, without employment adjustment costs, this model still has difficulties with the evidence presented by DHS. Next, we use Caballero's (1992) model of employment adjustment to demonstrate the basic mechanism driving the findings for aggregate job flows in our work.

Benchmark business cycle model

Our benchmark business model includes the three main elements of standard models described earlier: representative agents, symmetric aggregate shocks, and frictionless markets. The model departs from standard models in that it incorporates idiosyncratic technology shocks. However, it incorporates these shocks in a way that retains the validity of the representative agent assumption for aggregate analysis. Our purpose is to develop a concrete example to illustrate the extent of the failure of this class of models with respect to the DHS evidence.

Consider an economy composed of a single infinitely lived household and a continuum (very large number) of productive establishments called plants, which interact in competitive goods and labor markets in order to maximize utility and profits, respectively. To connect with the plant-level evidence on job creation and destruction, we assume that plants are subject to plant-specific random technology shocks, but otherwise are identical. These shocks include a common aggregate component; we make assumptions so that the behavior of the plants when considered in the aggregate corresponds to that of a *stand-in* representative plant that faces the common aggregate shock alone.

The representative household chooses consumption and work effort to maximize the present discounted value of utility subject to a budget constraint. Its decision problem is:

$$\max_{\{h_t,\, n_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \log\left(h_t - n_t^{\gamma}/\gamma\right)$$

$$\text{subject to } h_t \leq w_t n_t + \int_0^1 \pi_{i,t} di,\ t = 0,1,2,....$$

Here E_0 is the mathematical expectations operator conditional on information at date 0; h_t and n_t denote the date t consumption of the household and date t labor supply, respectively; $0 < \beta < 1$ is the household's subjective time discount factor; $\gamma > 1$ is an exogenous parameter governing the elasticity of labor supply; and w_t is the wage rate in consumption units. In addition, $\pi_{i,t}$ denotes time t profits of firm $i \in [0,1]$, also in consumption units, which the household receives by virtue of its ownership of plants. Hence, the last term on the right hand side of the budget constraint is the sum of profits at all plants.

Household optimization yields a first order condition relating labor supply to the wage at each date t. This can be rearranged to arrive at the following labor supply schedule for the household:

$$1)\quad n_t^s = w_t^{\frac{1}{\gamma-1}} \equiv S(w_t).$$

Since there is only one household, this equation also determines the economy-wide labor supply schedule, summarized by S (.).

Plant $i \in [0,1]$ produces output, $y_{i,t}$, for sale in the goods market using the technology $y_{i,t} = \theta_{i,t}^{1-\alpha} n_{i,t}^{\alpha}$. Here $0 < \alpha < 1$ and $\theta_{i,t}$ is the time t random technology disturbance for firm i. The random technology disturbance has the form

$$\theta_{i,t} = \eta_{i,t} + \theta_t.$$

Here $\eta_{i,t}$ is an idiosyncratic shock that follows a stationary stochastic process with support $[-\eta, \eta]$, $\eta > 0$, and $\theta_t > \eta$, $\forall t \geq 0$, is an aggregate disturbance that is common to all plants, which follows a stationary stochastic process. Two assumptions guarantee the existence of a stand-in representative plant: 1) θ_t is independent of $\eta_{i,t}$ for each i, and 2) $E_t\eta_{i,t} = 0$, that is, the idiosyncratic terms sum to zero at each date t.

The manager of the plant is assumed to maximize profits on a period-by-period basis, so its optimization problem is

$$2) \quad \max \pi_{i,t} = \theta_{i,t}^{1-\alpha} n_{i,t}^{a} - w_t n_{i,t}.$$

Optimization at plant i yields a first order condition for labor demand which must hold at each date t. Solving this for $n_{i,t}$, we have the labor demand schedule for the ith plant,

$$3) \quad n_{i,t}^{d} = \theta_{i,t} \left[\frac{\alpha}{w_t}\right]^{\frac{1}{1-\alpha}}.$$

Adding over all plants and making use of assumptions 1 and 2 above, we have the aggregate labor demand schedule

$$4) \quad n_t^{d} = \theta_t \left[\frac{\alpha}{w_t}\right]^{\frac{1}{1-\alpha}} \equiv D\,(w_t;\theta_t).$$

A competitive equilibrium in this model consists of a sequence of wages $\{w_t\}$ and quantities $\{h_t, n_t^s, (n_{i,t}^d : i \in [0,1])\}$ such that 1) given the wages, $\{h_t, n_t^s\}$ solve the household's problem, and for each i, $\{n_{i,t}^d\}$ solves plant i's problem, and 2) at these quantities, the goods market clears $h_t = \int_0^1 y_{i,t}\,di$ and the labor market clears $n_t^s = n_t^d$ at each date t.

The equilibrium quantities and wage rate at each date t are found as follows. First, substituting for n_t^d using equation 4 and for n_t^s using equation 1 in the labor market clearing condition and solving for w_t, we find the equilibrium wage rate at date t:

$$5) \quad w_t = \left(\alpha\theta_t^{1-\alpha}\right)^{\frac{\gamma-1}{\gamma-\alpha}}.$$

This says that the wage rate is increasing in the aggregate technology shock due to the assumptions made above on the magnitudes of γ and α. Using equation 5 to substitute for the wage in equation 4, we can find equilibrium aggregate labor input:

$$6) \quad n_t = A\theta_t^{\frac{1-\alpha}{\gamma-\alpha}},$$

where $A = \alpha^{1/(\gamma-\alpha)}$. We follow convention and interpret labor input as employment.[5] Then, equation 6 indicates that equilibrium employment is also an increasing function of the aggregate technology shock. Notice also that since the number of plants sums to unity, total employment corresponds to average employment across plants. Equilibrium employment at plant i is found similarly using equation 3:

$$7) \quad n_{i,t} = \theta_{i,t} A\theta_t^{\frac{1-\gamma}{\gamma-\alpha}}$$

$$= A\eta_{i,t}\theta_t^{\frac{1-\gamma}{\gamma-\alpha}} + A\theta_t^{\frac{1-\alpha}{\gamma-\alpha}}$$

$$8) \quad = A\eta_{i,t}\theta_t^{\frac{1-\gamma}{\gamma-\alpha}} + n_t.$$

This indicates that in equilibrium, employment at firms is heterogeneous and varies about average labor input. Equilibrium consumption is derived using the goods market clearing condition as follows:

$$h_t = \int_0^1 \theta_{i,t}^{1-\alpha} n_{i,t}^{\alpha}\,di$$

$$= \theta_t^{1-\alpha} n_t^{\alpha} + \int_0^1 \eta_{i,t}\theta_t^{-\alpha} n_t^{\alpha}\,di$$

$$= \theta_t^{1-\alpha} n_t^{\alpha} + \theta_t^{-\alpha} n_t^{\alpha}\int_0^1 \eta_{i,t}\,di$$

$$9) \quad = \theta_t^{1-\alpha} n_t^{\alpha}.$$

The first line of this derivation is just the goods market clearing condition and the second line follows after substituting for $n_{i,t}$ using equation 7 and rearranging the resulting expression using the definition of $\theta_{i,t}$ and equation 6. We arrive at the third line by using assumption 1 and the last line follows from assumption 2.

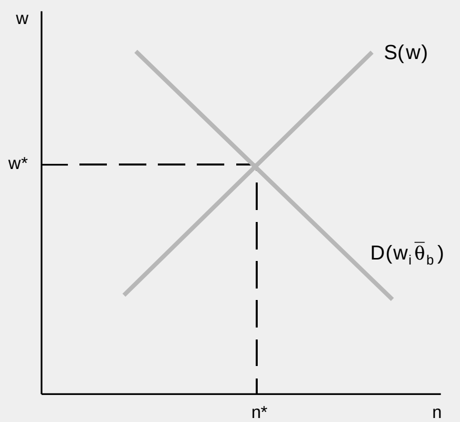
Note that the detail of firm-level heterogeneity in the model is unnecessary if we are only interested in aggregate consumption and employment. First, we could have derived equation 4 by considering the problem of a representative plant identical to that in equation 2, with $\theta_{i,t}$ replaced by θ_t. Second, equation 6 would be the correct equilibrium labor input in such a model. Third, equation 9 would continue to hold in this model. Thus, in terms of its predictions for aggregate consumption and employment, this model is identical to a model involving a representative plant facing only an aggregate technology shock.

Job creation and destruction in the benchmark model

To analyze the model's implications for creation and destruction, we discuss a steady-state scenario in which the aggregate disturbance is a constant. Figure 3 depicts equilibrium

FIGURE 3

Steady state labor market equilibrium

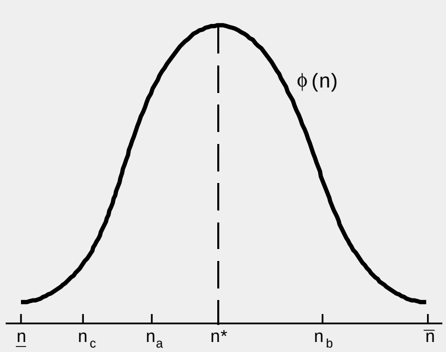


in the labor market for this case; we assume $\theta_t = \bar{\theta}$, $\forall\, t$. Equilibrium employment is given by the intersection of the labor demand and supply schedules at employment n^* and wage rate w^*. This diagram is useful for studying the model's implications for aggregate employment. However, the job creation and destruction data involve counting employment changes at the plant level. To investigate our model's implication for creation and destruction, therefore, we must study the model's implications for plant-level employment.

Figure 4 shows the distribution of employment across plants for the constant aggregate shock case. We assume that the idiosyncratic shocks are independently and identically distributed according to a truncated normal distribution,

FIGURE 4

Steady state distribution of employment across plants



with the truncation points determined by the bounds for the idiosyncratic shocks stated above. Employment at the plant level is distributed according to the density ϕ (.), which has mean n^* and lower and upper bounds $\bar{n}$ and $\underline{n}$, respectively.[6]

In the current example, individual plants receive a new idiosyncratic shock each period, so employment is always changing at the plant level. For example, a plant at a given level of employment in figure 2, say n_a, at date t–1 is subject to a new idiosyncratic disturbance at date t. The realization of this disturbance could be higher or lower than the level underlying n_a. A higher realization of technology might involve the plant in question choosing $n_b > n_a$ and a lower realization might involve the plant choosing $n_c < n_a$. In the former case $n_b - n_a$ jobs are created and in the latter case $n_a - n_c$ jobs are destroyed. There are many similar plants, all of which get different realizations of the idiosyncratic technology disturbance.

To connect this model with the DHS evidence, we need to investigate measures of aggregate job creation and destruction. Following DHS, aggregate job creation at date t is the sum of all jobs created at plants that increase employment between dates t and t–1:

$$\text{total job creation} = \sum_{\{i: n_{i,t} > n_{i,t-1}\}} (n_{i,t} - n_{i,t-1}).$$

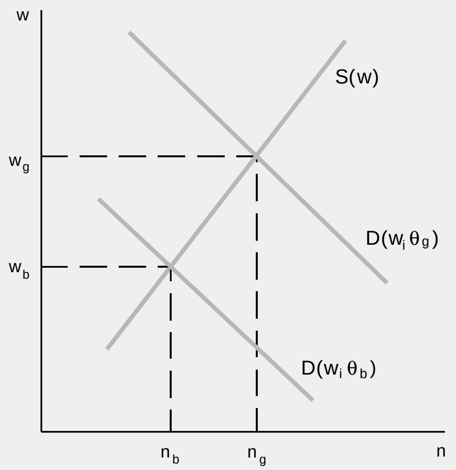
Similarly, aggregate job destruction at date t is the sum of all jobs destroyed at plants that decrease employment between dates t and t–1:

$$\text{total job destruction} = \sum_{\{i: n_{i,t} < n_{i,t-1}\}} (n_{i,t-1} - n_{i,t}).$$

Let N^c and N^d denote the total number of plants that create and destroy at each date, respectively. Also, let c and d denote the *average* amount that each job-creating plant creates and each job-destroying plant destroys at each date, respectively. Since aggregate employment, n^*, is constant in a steady state, aggregate job creation and destruction must be equal at every date, $N^c c = N^d d$. Furthermore, due to the symmetry in the distribution of idiosyncratic disturbances and the fact that all plants will either create or destroy, we have $N^c = N^d = 1/2$, and therefore, $c = d$. We use $a > 0$ to denote the common value taken by c and d.

FIGURE 5

Aggregate labor market fluctuations



To address the DHS evidence on aggregate job creation and destruction, we need to modify the current model specification to allow the aggregate technology shock to vary. To keep it simple, suppose that the aggregate shock can take on only two values, $\theta_g > \theta_b$, where g is good and b is bad. Figure 5 depicts equilibrium in the labor market for the two possible technology shocks. When $\theta_t = \theta_g$, employment is n_g and the wage rate is w_g; when $\theta_t = \theta_b$, employment is n_b and the wage rate is w_b. A given sequence of θ_t determines the dynamics of aggregate employment as the labor demand schedule shifts up and down the labor supply schedule.

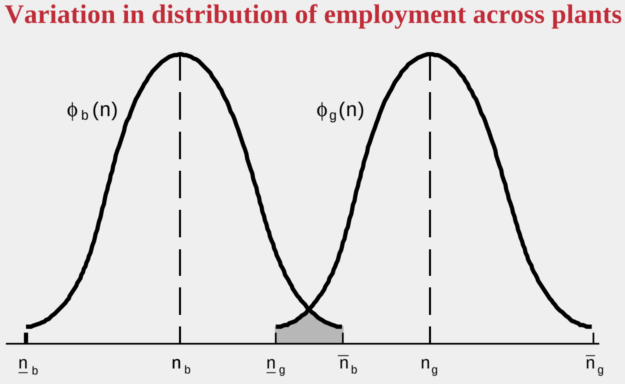
Figure 6 displays the distribution of employment across plants for the two possibilities of the aggregate shock. We make the same distributional assumption for the idiosyncratic shocks as in the steady state analysis above. We assume the distribution is constant over time and, in particular, that it does not depend on the realization of the aggregate technology disturbance. When $\theta_t = \theta_g$, employment is distributed according to the density $\phi_g(.)$, which has mean n_g, and lower and upper bounds $\bar{n}_g$ and $\underline{n}_g$, respectively. Similarly, when $\theta_t = \theta_b$, employment is distributed according to the density $\phi_b(.)$ which has mean n_b and lower and upper bounds $\bar{n}_b$ and $\underline{n}_b$, respectively.[7] The fact that the two densities overlap (shaded region in the figure) shows that the variance of the aggregate shock is small relative to the variance of the idiosyncratic shock.

We can use figure 6 to study the model's business cycle implications for job creation and destruction. If the aggregate shock at date t is the same as at date t–1, the cross-sectional pattern of creation and destruction is the same as for the steady state example. This follows because the pattern is determined by the location of lagged plant-level employment relative to the optimal current level. With the distribution of idiosyncratic shocks being time invariant, the distribution of plants' lagged employment relative to their optimal current levels must be the same each period. The implication of this observation is that when the aggregate shock remains the same as its lagged value, $N_t^c = N_t^d = 1/2$ and $c_t = d_t = a$, as in the steady state case.

Next, observe that changes in creation and destruction at the aggregate level only occur when the aggregate level of technology changes. Suppose that the aggregate shock changes from θ_b at date t–1 to θ_g at date t. Aggregate job creation must necessarily increase since the employment distribution shifts to the right and average employment rises (see figure 6.) This change is accomplished by an increase in the number of plants creating jobs, to $N_t^c = 1/2 + \delta$, where $0 < \delta < 1/2$, and an increase in the average amount each job-creating plant creates, to $c_t = a + \Delta\theta$, $\Delta\theta = \theta_g - \theta_b$. The chances of getting a higher realization of technology at a given plant at date t than at date t–1 have increased, so in the aggregate there must be more plants

FIGURE 6

Variation in distribution of employment across plants



creating jobs. Furthermore, the increase in the mean of the disturbances influencing plant employment implies the level to which a typical job creator creates must also increase.

Conversely, aggregate job destruction must fall at date t compared to date t–1. This is a result of both a fall in the number of plants destroying and a fall in the average amount a job-destroying plant destroys. Only the plants that had employment in the interval $(\underline{n}_g, \overline{n}_b)$ at date t–1 destroy jobs at date t, whereas at date t–1 any plant with employment in the interval $(\underline{n}_b, \overline{n}_b)$ at date t–2 is a candidate for job destruction. It follows that the number of plants that could possibly destroy jobs at date t is given by the shaded area in figure 6, and the number of plants that could possibly destroy jobs at date t–1 is given by the area under $\phi_b(.)$. Since the former is smaller than the latter, the number of job-destroying plants must fall at date t compared to date t–1. Moreover, since the shaded region has smaller support than for the $\phi_b(.)$ density as a whole, the typical amount destroyed by a job-destroying plant when the aggregate shock switches from θ_b to θ_g must also fall. Due to the symmetry in the model, we have $N_t^d = 1/2 - \delta$ and $d_t = a - \Delta\theta$.

Clearly, the impact of an increase in aggregate technology on job creation and destruction is reversed when there is a decrease in aggregate technology. In this case, the numbers of job creators and destroyers are $N_t^c = 1/2 - \delta$ and $N_t^d = 1/2 + \delta$, respectively, and the average amounts created and destroyed are $c_t = a - \Delta\theta$ and $d_t = a + \Delta\theta$, respectively. This leads us to conclude that the model predicts that aggregate job creation and destruction are equally variable and perfectly negatively correlated, contradicting our earlier observations based on DHS.

However, the model achieves some success at replicating evidence on the cross-sectional distribution of employment growth. As in DHS, plant-level job creation and destruction are pervasive, display considerable heterogeneity, and occur in booms and recessions (although all plants change employment every period in this model). In addition, changing from an expansion to a recession involves an increase in the variance of employment growth, consistent with DHS. Of course, if the aggregate shock equals θ_b for several periods, the variance of employment growth will be the same as it would be if the aggregate shock equaled θ_g for several periods, so the success here is limited.

Several authors have interpreted the change in this variance over the business cycle as evidence of a prominent business cycle role for idiosyncratic disturbances. While this may be the case, it may still be possible to abstract from such disturbances when considering business cycles. If the variance of the idiosyncratic disturbances is countercyclical but the symmetry in the distribution of idiosyncratic shocks is retained, the analysis of aggregate consumption and employment developed above may still apply. In this case, it would be legitimate to abstract from the microeconomic detail when considering aggregate employment fluctuations, as is the practice in conventional approaches to studying business cycles. One case in which it would not be legitimate to abstract from the microeconomic detail would be if labor market search frictions impede the process of reallocating workers across plants.[8]

The above discussion suggests that by introducing idiosyncratic uncertainty into an otherwise standard business cycle model, it is possible to account for some of the qualitative features of the cross-sectional distribution of employment growth. Nevertheless, the benchmark model does have difficulty accounting for the DHS evidence on aggregate job flows.

Moving the model closer to the data

The (moderate) success of the benchmark model at accounting for the plant-level observations in DHS raises the possibility that, with further modifications, the model might account for the evidence on aggregate job flows without dropping the main assumptions of standard business cycle models. It is important to recognize that simple changes to the stochastic structure of the benchmark will not change our main qualitative findings with respect to aggregate job flows. For example, introducing persistence into either the idiosyncratic or aggregate technology implies small differences in the job flow variances and a less than perfect negative job flow correlation. However, the differences with the observed magnitudes will remain stark. Adding more aggregate shocks to the benchmark model will not have a substantive impact on its predictions for aggregate job flows either.[9] Finally, allowing the distribution of idiosyncratic shocks to be asymmetric about their mean has no impact on the main conclusion.

The assumptions underlying the failure of the benchmark model are those that characterize conventional views of the business cycle: representative agents, symmetric aggregate shocks, and frictionless markets. The validity of using representative agents to model aggregate employment in the benchmark model relies on the special assumptions we made for the idiosyncratic shocks. If the idiosyncratic shocks are correlated with the aggregate shocks in a particular way, it may be possible to move the model closer to the data. However, this kind of assumption would likely disallow a representative agent formulation for the model. The importance of symmetric aggregate shocks is highlighted, for example, by Caballero's (1992) findings. Caballero showed that if aggregate shocks have a particular form of asymmetry, it is possible to reproduce some of the aggregate job flow evidence. The role of frictionless markets is less obvious, but Mortensen and Pissarides (1994) find that the implications of a particular kind of friction in the labor market may account for some of the evidence on aggregate job flows.

We conclude that the three main elements of the benchmark model, which are shared by a broad class of models in macroeconomics, contribute to its failures with respect to the DHS evidence. This is why so much work has been done introducing model elements that deviate from the conventional to try to account for the DHS evidence. As discussed earlier, much of this work has focused on aggregate job flows and has not attempted to make a connection between this evidence and the cross-sectional evidence. This may be justified to some extent by the finding, described above, that accounting for the cross-sectional evidence is not necessarily a challenge to a model that shares most of the features of standard models. However, it remains possible that there is a connection between the cross-sectional evidence and the evidence on aggregate job flows. Making this connection is one of the main contributions of our work.

Building on plant-level evidence to explain aggregate job flows

The evidence on heterogeneity in plant-level job growth, including the prevalence of plant-level inactivity in employment adjustment, helps to motivate our research. We examine a model in which it is costly to change plant-level employment, where the marginal costs of changing employment are discontinuous at the point of zero change. This implies that it is sometimes optimal to keep employment constant, even as the level of technology changes at the plant level. We find that the same friction which gives rise to the nondifferentiable costs of employment adjustment may also account for the evidence on aggregate employment flows. In contrast to the models discussed earlier, the employment-adjustment technology we study implies variation in the average amounts created and destroyed by employment-changing plants. The connection between the micro and aggregate evidence arises because the employment-adjustment technology, which helps account for the micro evidence, may also imply that the average amount of job destruction by job-destroying plants is more variable than the average amount of job creation by job-creating plants. This helps account for the evidence on aggregate job flows.

Below, we describe a simple version of our model based on Caballero's (1992) model of employment adjustment. We use this example to illustrate the basic mechanism driving our success at accounting for the DHS observations on aggregate employment flows. Then we describe the economics underlying the mechanism and discuss how our model may also account for the plant-level evidence.

Caballero's model of employment adjustment

Caballero's (1992) mechanical model of employment adjustment captures key features of fully articulated economic models, in which employment adjustment is infrequent and lumpy.[10] Consider an industry with a fixed number of plants subject to idiosyncratic and, possibly, aggregate disturbances. Let each individual plant *i* have some desired or frictionless level of employment at time *t*, $n^*_{i,t}$. We can imagine this frictionless level of employment being determined as in the benchmark model. The plant's frictionless level of employment is assumed to evolve exogenously as follows:

$$10)\quad n^*_{i,t} = n^*_{i,t-1} + \begin{cases} +1 \text{ with probability } 1/2 \\ -1 \text{ with probability } 1/2 \end{cases}.$$

The realization of the increment to $n^*_{i,t}$ is the idiosyncratic disturbance to plant *i*. Actual employment at the plant level, $n_{i,t}$, is not always equal to the frictionless optimal level. Let $\delta_{i,t} = n_{i,t} - n^*_{i,t}$ denote the deviation of actual employment from its frictionless level.

The rule governing employment decisions at the plant level, or the plant-level *employment policy*, is specified exogenously as follows. An *employment action*, which means a change in actual employment at the plant, occurs whenever $\delta_{i,t}$ will cross a threshold *in the absence of employment action*. If, in the absence of employment action, $\delta_{i,t} > D > 0$, the plant reduces employment to a level such that $\delta_{i,t}$ does not actually cross the threshold. Similarly, if, in the absence of employment action, $\delta_{i,t} < C < 0$, the plant increases employment to a level such that $\delta_{i,t}$ does not actually cross the threshold. If, in the absence of employment action, $D < \delta_{i,t} < C$, no employment action is taken by the plant. Employment typically changes by an amount that depends on 1) whether the change involves job creation or job destruction; and 2) the realizations of aggregate shocks to the economy. Here, we assume that the aggregate state of the economy is constant, so employment changes only depend on whether jobs are being created or destroyed. We denote the amount employment changes at a job-creating plant by c and at a job-destroying plant by d. This threshold employment policy is a stylized version of what would emerge if the plants in the benchmark model were to face employment adjustment costs that are nondifferentiable at the point of zero change.

The following example shows the evolution of employment at the plant level, assuming the employment policy described in the previous paragraph. We assume $D = 1$, $C = -1$, $d = 2$, and $c = 1$. Then, according to the employment policy, $\delta_{i,t}$ can take on only three values: –1, 0, and 1. Next, we describe the various possible outcomes for $\delta_{i,t+1}$ and the probabilities of these outcomes given the three possible date t values for $\delta_{i,t}$.

Suppose $\delta_{i,t}$ equals –1. According to equation 10, there is a probability equal to 1/2 that the frictionless level of employment will increase by 1 at date $t + 1$. In this case, if no employment action is taken, $\delta_{i,t+1} < C$. The employment policy requires that employment at the plant increases by $c = 1$. Therefore,

$$\begin{aligned}\delta_{i,t+1} &= \delta_{i,t} + \text{increment due to } n^*_{i,t+1} + \text{increment due to employment policy}\\ &= -1 -1 + 1\\ &= -1.\end{aligned}$$

There is also a probability equal to 1/2 that the frictionless employment level drops by 1. In this case $\delta_{i,t} = -1 + 1 + 0 = 0$ since no employment action is taken.

Now suppose $\delta_{i,t} = 0$. In this case no employment action is taken, since neither of the possible changes in the frictionless employment level leads to a threshold being crossed in the absence of employment action. There are two possible outcomes for $\delta_{i,t+1}$. With probability 1/2 $n^*_{i,t+1}$ increases by 1, so that $\delta_{i,t+1} = 0 -1 + 0 = -1$, and with probability 1/2 $n^*_{i,t+1}$ decreases by 1, in which case $\delta_{i,t+1} = 0 + 1 + 0 = 1$.

Finally, suppose $\delta_{i,t} = 1$. There is a probability equal to 1/2 that the destruction threshold will be crossed next period in the absence of employment action. In this case, $d = 2$ jobs will be destroyed, so that $\delta_{i,t+1} = 1 + 1 - 2 = 0$. There is also a probability equal to 1/2 that the frictionless employment level will increase by 1 at date $t + 1$, in which case no employment adjustment occurs and we have $\delta_{i,t} = 1 - 1 + 0 = 0$. Hence, when $\delta_{i,t} = 1$, it follows that $\delta_{i,t+1} = 0$ with certainty.

To summarize this, we use a transition equation for a vector that describes the fraction of plants at each possible level of $\delta_{i,t}$. Let p_t be a 1 x 3 vector where the jth column indicates the probability that for any plant i, $\delta_{i,t} = j - 2$. With a large number of plants, these probabilities equal the fraction of plants at each of the three possible values for $\delta_{i,t}$. Below, we use the notation $p_t(\delta)$ to denote the fraction of plants at the state $\delta_{i,t} = \delta$. The evolution of the vector p_t depends on the plants' employment policy and is given by

$$11) \quad p_{t+1} = p_t P,$$

where

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}.$$

The rows and columns of P represent possible values for $\delta_{i,t}$ and $\delta_{i,t+1}$, respectively. For example, the (3,2) position in this matrix says that starting from $\delta_{i,t} = 1$, $\delta_{i,t+1} = 0$ with probability 1. Equation 11 defines a Markov chain on the vector of probabilities p_t. It describes how the fraction of plants at each possible level of $\delta_{i,t}$ evolves over time.

The matrix P satisfies the assumptions required for p_t to converge to a constant vector.[11] That is, from any initial vector p_0, whose elements are non-negative and sum to unity, iterating on equation 11 implies $p_t \rightarrow p^*$ as $t \rightarrow \infty$, where the elements of p^* are non-negative and sum to unity. The vector p^* is called the vector of stationary probabilities, since it has the property that

$$p^* = p^* P.$$

That is, given an initial vector p^*, the system is stationary in that the fraction of plants at each possible level of $\delta_{i,t}$ will not change. (The vector of stationary probabilities for our example is $p^* = [2/5\ 2/5\ 1/5]$.) This stationary situation is analogous to the steady state discussed for the benchmark model, and we have $N^c c = N^d d$, using the same notation as before. In particular, while aggregate numbers at each level of $\delta_{i,t}$ do not change, employment change at individual plants is an ongoing phenomenon. Unlike the benchmark model, however, here in every period some plants neither create nor destroy jobs. Thus, in qualitative terms, this example seems to fit more closely the cross-sectional distribution of employment growth discussed in DHS.

To study variation in creation and destruction, we need to introduce some form of aggregate uncertainty. We assume that the probabilities governing the evolution of the frictionless level of employment, $n^*_{i,t}$, can take on two sets of values. Specifically, in *good times*

$$n^*_{i,t} = n^*_{i,t-1} + \begin{cases} +1 \text{ with probability } \lambda_g \\ -1 \text{ with probability } 1 - \lambda_g \end{cases}$$

and in *bad times*

$$n^*_{i,t} = n^*_{i,t-1} + \begin{cases} +1 \text{ with probability } \lambda_b \\ -1 \text{ with probability } 1 - \lambda_b \end{cases}.$$

We assume that good times and bad times occur with probability 1/2 each and that

$$\lambda_g = (1 + \Delta)/2,$$
$$\lambda_b = (1 - \Delta)/2.$$

Notice that λ_g and λ_b equal the fraction of plants whose frictionless employment increases by 1 in good times and bad times, respectively. Here, Δ represents the fraction of the total uncertainty faced by an individual plant that is due to aggregate uncertainty.

With this form of aggregate uncertainty, the transition matrix of the Markov chain described by equation 11 is no longer time invariant. The transition matrix now takes on two values, P_g and P_b, depending on the aggregate state. Using the three-state example developed above we have

$$P_g = \begin{bmatrix} \lambda_g & 1 - \lambda_g & 0 \\ \lambda_g & 0 & 1 - \lambda_g \\ 0 & 1 & 0 \end{bmatrix}$$

and

$$P_b = \begin{bmatrix} \lambda_b & 1 - \lambda_b & 0 \\ \lambda_b & 0 & 1 - \lambda_b \\ 0 & 1 & 0 \end{bmatrix}.$$

Now that the aggregate state may vary, we must consider how the amounts changed by individual job creators and destroyers may vary with the aggregate state of the economy. Caballero (among others) only considers cases in which these values are held constant. However, we argue that there are good reasons to expect variation in employment policies and that the amounts changed by job destroyers may be more variable than the amounts changed by job creators. Next, we present examples that summarize these two possibilities and discuss our intuition that the variable employment policies case may be a more plausible assumption.

Job creation and destruction with constant and variable employment policies

To facilitate comparisons with Caballero's (1992) analysis, we borrow the basic structure of our examples from his paper. Enlarging the state space from the cases considered above, we assume $D = 7$ and $C = -7$ so that $\delta_{i,t}$ now takes on values between –7 and 7. This reduces the impact of state–space discreteness. We also assume $\Delta = 0.30$ so that $\lambda_g = 0.65$ and $\lambda_b = 0.35$. The examples we consider share these features, but differ in the assumptions we make on how c and d depend on the aggregate state.

In the first set of examples, employment policies are constant in the presence of aggregate uncertainty, that is, c and d equal constants. First, we consider $c = d = 1$, so that creation and destruction at the plant level are symmetric. Second, we consider $c = 1$ and $d = 6$, so that destruction at the plant level is larger than creation. Third, we consider $c = 6$ and $d = 1$, so that creation at the plant level is larger than destruction.

In the second set of examples, employment policies are variable, so that c and d depend on the aggregate state. We use the subscripts g and b to denote the amounts created and destroyed in good times and bad times, respectively. We consider three separate cases to facilitate comparison with the first set of examples and to explore the idea that the amounts destroyed at job-destroying plants may be more variable than the amounts created at job-creating plants. First, we suppose $c_g = c_b = 1$, $d_g = 1$, and $d_b = 2$. Second, we suppose $c_g = c_b = 1$, $d_g = 3$, and $d_b = 6$. Third, we suppose $c_g = c_b = 6$, $d_g = 1$, and $d_b = 2$.

In all these examples, aggregate job creation and destruction are measured as $\lambda_t\, p_t\,(-7)\, c_t$ and $(1 - \lambda_t)\, p_t\,(7) d_t$, respectively. Here, λ_t equals λ_g in good times and λ_b in bad times. Also c_t and d_t equal c and d, respectively, in the first three cases. In the second three cases, $c_t = c_g$ and $d_t = d_g$ in good times and $c_t = c_b$ and $d_t = d_b$ in bad times. The analysis below is based on statistics involving these measures of creation and destruction based on 1,000 replications of samples of 200 periods each.

The implications of these parameterizations of the Caballero (1992) model for the cyclical behavior of job creation and destruction are summarized in table 2. The first two columns, reported in Caballero, show the volatility of aggregate job creation and destruction. The third column shows the correlation between creation and destruction. The first three rows refer to the constant employment policy cases and the second three rows refer to the cases with variable employment policies.

In the constant policy cases, creation and destruction are roughly equally variable, regardless of the relative magnitudes of c and d. Caballero (1992) described this as a “fallacy of composition,” since it says that even if adjustment at the plant level displays an asymmetry, it need not translate to aggregate variables. We also note that the absolute values of the correlation statistic in these examples are roughly double those in table 1 for the U.S. manufacturing sector.

TABLE 2

Aggregate job creation and destruction using the Caballero model

	σ (creation) / $\bar{x}$ (creation)	σ (destruction) / $\bar{x}$ (destruction)	ρ(creation, destruction)
Constant policies			
$c = 1$ $d = 1$	0.567 (0.005)	0.560 (0.005)	–0.809 (0.001)
$c = 1$ $d = 6$	0.567 (0.004)	0.563 (0.005)	–0.809 (0.001)
$c = 6$ $d = 1$	0.569 (0.004)	0.560 (0.005)	–0.810 (0.001)
Variable policies			
$c_g = c_b = 1$ $d_g = 1,\ d_b = 2$	0.567 (0.005)	0.780 (0.007)	–0.633 (0.001)
$c_g = c_b = 1$ $d_g = 3,\ d_b = 6$	0.566 (0.004)	0.818 (0.007)	–0.700 (0.001)
$c_g = c_b = 6$ $d_g = 1,\ d_b = 2$	0.572 (0.004)	(0.780 (0.006)	–0.630 (0.001)

Notes: In the column headings, σ(y) denotes the average across samples of the within-sample standard deviations of aggregate variable y; $\bar{x}(y)$ denotes the average across all samples of the mean (over time) of aggregate variable y, respectively; $\rho(y,z)$ is the average across samples of the within-sample correlation between aggregate variables y and z. The numbers in parentheses are Monte Carlo standard errors for the associated statistic. These equal the standard deviation of the relevant statistic across samples divided by the square root of the number of samples (1,000).

Source: Authors’ calculations based on Caballero’s (1992) model of employment dynamics.

In the variable employment policy cases, we see an improvement in the empirical implications of the model versus the constant policy cases. For all three examples, job destruction is clearly more volatile than job creation. This might seem obvious, given that we assume that d_t is more variable than c_t. However, the structure of the transition matrices P_g and P_b is influenced by the plants' employment policy. This means assumptions regarding the variability of c_t and d_t influence the evolution of the numbers of creators, $\lambda_t p_t(-7)$, and destroyers, $(1-\lambda_t)\, p_t(7)$. In principle, movements in the numbers of agents engaged in employment action can interact with the amounts actually created and destroyed to undo microeconomic asymmetries at the aggregate level of measurement. Another thing to notice from table 2 is that variable employment policies tend to reduce the strong negative correlation between job creation and destruction that constant employment policies imply.

These examples show the potential for excess variability in job destruction over job creation at the plant level to translate into phenomena that are more consistent with empirical evidence on aggregate job flows than if employment policies are assumed to be constant. Next, we assess whether this a reasonable assumption.

Justifying variable employment policies

Consider a plant with similar production technology to that considered in the benchmark model. Suppose the wage rate is exogenous and the plant takes the price, normalized at unity, as given. The key change to the plant-level production environment we introduce is that when employment changes at the plant, the owner incurs a cost associated with reorganizing work to accommodate a larger or smaller work force. Specifically, for plant $i \in [0,1]$ if $n_{i,t} > n_{i,t-1}$, revenue is reduced by $\tau^c (n_{i,t} - n_{i,t-1})$, $\tau^c \geq 0$; if $n_{i,t} < n_{i,t-1}$, revenue is reduced by $\tau^d (n_{i,t-1} - n_{i,t})$, $\tau^d \geq 0$, $\tau^c\,\tau^d > 0$; and if employment is unchanged, revenue is unaffected.

The optimal employment policy in this environment is hard to compute, because the adjustment costs make the plant owner's problem dynamic. For example, in deciding whether to destroy a job in response to a low technology shock, the owner must take into account the possibility that technology will improve, which would make it desirable to keep employment at a high level. Since these dynamic considerations are not crucial to the main argument, we assume that the plant owner infinitely discounts the future, choosing current employment to maximize current profits without regard to the impact of the decision on future actions.

We characterize the optimal employment policy at plant $i \in [0,1]$ at some date t. Let $n_{i,t-1} > 0$ denote employment last period, let $c_{i,t} \geq 0$ denote job creation in the current period, and let $d_{i,t} \geq 0$ denote job destruction in the current period. Date t employment is $n_{i,t} = n_{i,t-1} + c_{i,t} - d_{i,t}$. Then, the plant owner's objective is

$$\max_{c_{i,t}, d_{i,t} \geq 0} \; \theta_{i,t}^{1-\alpha} (n_{i,t-1} + c_{i,t} - d_{i,t})^{\alpha} - w_t (n_{i,t-1} + c_{i,t} - d_{i,t}) - \tau^c c_{i,t} - \tau^d d_{i,t}.$$

The relevant first order conditions for this problem are:

$$12)\quad \alpha\, \theta_{i,t}^{1-\alpha} (n_{i,t-1} + c_{i,t} - d_{i,t})^{\alpha-1} - w_t - \tau^c \leq 0,$$

$$13)\quad -\alpha\, \theta_{i,t}^{1-\alpha} (n_{i,t-1} + c_{i,t} - d_{i,t})^{\alpha-1} + w_t - \tau^d \leq 0,$$

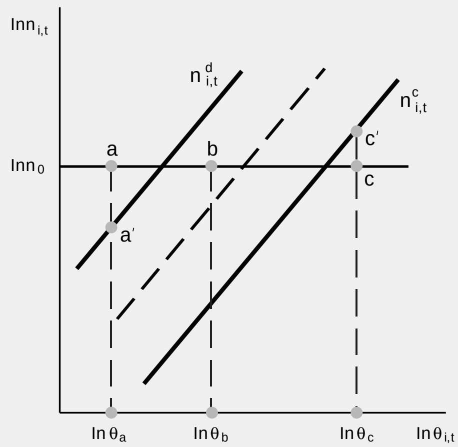
where the first condition applies to the choice for creation and holds with equality if $c_{i,t} > 0$ and the second condition applies to the choice for destruction and holds with equality if $d_{i,t} > 0$. We note from equations 12 and 13 that only one of $c_{i,t}$ and $d_{i,t}$ is ever strictly positive. Second, there may be no positive value of either choice variable which sets the relevant first order condition to zero. In this case, it is optimal to keep current employment at last period's level, $n_{i,t-1}$.

Figure 7 characterizes the optimal employment policy. The *frictionless schedule* (dashed line) is the locus of points $(\ln \theta_{i,t}, \ln\, n_{i,t})$, such that $n_{i,t} = \theta_{i,t}\, [\alpha/w_t]^{1/(1-\alpha)}$. The *creation schedule*, denoted $n_{i,t}^c$, is the locus of points $(\ln\, \theta_{i,t}, \ln n_i)$, such that equation 12 holds with equality. The *destruction schedule*, denoted $n_{i,t}^d$, is the locus of points $(\ln \theta_{i,t}, \ln n_{i,t})$, such that equation 13 holds with equality. The vertical distance between the creation and the frictionless schedules is the same as the vertical distance between the destruction and the frictionless schedules. This reflects an implicit assumption that $\tau^c = \tau^d > 0$.[12]

To understand the employment policy, consider three possible realizations of technology at plant i with a lagged employment value equal to n_0. Optimal current employment if current technology is θ_α involves destroying

FIGURE 7

Steady state creation, destruction, and frictionless optimal employment schedules



jobs so that employment is at the point on the destruction schedule consistent with this level of technology. The quantity of jobs destroyed in this case is the vertical distance between point *a* and point *a′*. Optimal current employment if current technology is θ_b is to leave it at n_0. In this case, no job creation or destruction occurs at the plant. Finally, the optimal employment policy if current technology is equal to θ_c is to create jobs equal to the vertical distance between point *c* and point *c′*.

Suppose we introduce aggregate uncertainty by assuming the real wage, w_t, is a random variable which can take on two values, $w_h > w_l > 0$. Furthermore, assume for now that $\tau^c = \tau^d = \lambda w_t$, $\lambda > 0$. This implies that when the wage changes, the adjustment costs change by the same percentage amount, as would be the case if the reorganization costs associated with changing employment were all absorbed in lost production time. It is easy to establish that

$$14)\quad \left|\partial \ln n^c_{i,t}(\theta_{i,t};w_t)/\partial \ln w_t\right| = \frac{1}{1-\alpha} = \left|\partial \ln n^d_{i,t}(\theta_{i,t};w_t)/\partial \ln w_t\right|.$$

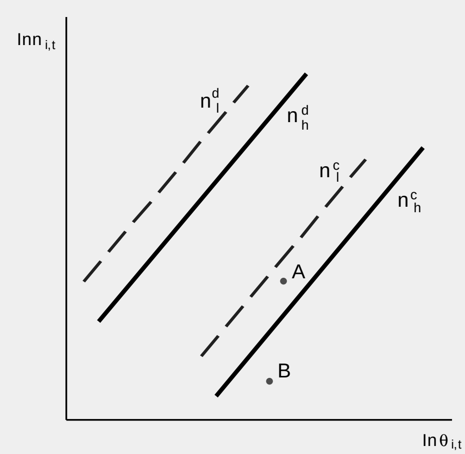
This says that, at each level of technology, the percentage change in the creation schedule due to a unit percent change in the wage is identical to the percentage change in the destruction schedule due to a unit percent change in the wage.

Figure 8 shows the implications of this for aggregate creation and destruction. Lines n^c_h and n^d_h are the creation and destruction schedules, respectively, associated with $w_t = w_h$; and lines n^c_l and n^d_l are the creation and destruction schedules, respectively, associated with $w_t = w_l$. The vertical distance between the two pairs of schedules is identical; the schedules shift by the same amount when the wage changes. This is a direct implication of equation 14.

Consider a change from $w_t = w_h$ to $w_t = w_l$. In figure 8, we see that the creation and destruction schedules are at a higher position in the state space compared with the high-wage case. Since the creation schedule when $w_t = w_l$ lies above the creation schedule when $w_t = w_h$, the number of job-creating plants must be greater than before. For example, take a plant with lagged employment and current technology such that its position in figure 8 is between the two creation schedules, say at point *A*. When $w_t = w_h$, this plant would neither create nor destroy jobs. However, when $w_t = w_l$, this plant becomes a job creator. Since there are many such plants, the number of job-creating plants must rise relative to the high-wage case. To see what happens to average creation, take a plant at position *B* in figure 8. This plant creates jobs regardless of the wage. However, the vertical distance from point *B* to n^c_l is greater than the vertical distance to n^c_h. This tells us that average creation must be larger in the low-wage state compared with the high-wage state. An analogous logic holds for job destruction.

FIGURE 8

Fluctuations in creation and destruction schedules with adjustment costs proportional to wage



Although employment policies are variable in this example, the fact that the creation and destruction schedules shift by the same amount in response to a wage disturbance suggests that this model is likely to imply roughly equal variation in aggregate creation and destruction (with standard assumptions regarding the process governing the wage). We aim to demonstrate that the destruction policy may be more variable than the creation policy, which is the key assumption underlying the examples in table 2.

In the analysis above, we assume that the adjustment costs are proportional to the wage, meaning that the costs associated with adjusting employment are perfectly correlated with the wages paid to production workers. This is unlikely, since part of the costs of reorganization involve capital and nonproduction workers. Suppose the adjustment costs do not depend on wages at all. In particular, suppose they are constant, as would be the case if they reflected a pure drain on output. This assumption delivers our desired result. To see why, we recalculate the elasticities presented above:

$$15)\quad \left|\partial \ln n^{c}_{i,t}(\alpha_{i,t};w_t)/\partial \ln w_t\right| = \frac{1}{1-\alpha}\,\frac{w}{w+\tau^{c}}\,.$$

$$16)\quad \left|\partial \ln n^{d}_{i,t}(\alpha_{i,t};w_t)/\partial \ln w_t\right| = \frac{1}{1-\alpha}\,\frac{w}{w-\tau^{d}}\,.$$

These expressions indicate that job creation costs tend to dampen variation in the job creation schedule and job destruction costs tend to amplify variation in the job destruction schedule. What is the intuition for this? The job creation schedule is the locus of current (log) employment and (log) technology, such that the marginal benefit of adding a worker is equated to the marginal cost (see equation 12). The dampening effect of the job creation cost arises because it adds to the marginal cost of creating a job. Along the job destruction schedule, the marginal benefits and costs of *keeping* a worker are equated. Job destruction costs enter this calculation as a benefit because, at the margin, keeping a worker involves saving the costs associated with destroying the job. The cost saving acts like a reduction in the wage for the marginal worker; hence, job destruction costs enter with a minus sign in equation 16 and act to amplify fluctuations in the destruction schedule.

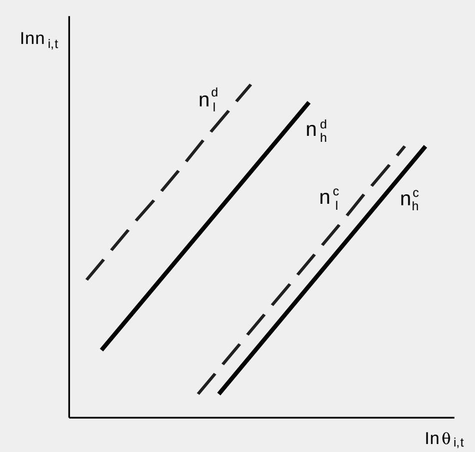
Notice from equations 15 and 16 that the dampening effect of the creation cost and the amplifying effect of the destruction cost do not depend on the relative magnitudes of the costs. Put another way, asymmetry in the way the schedules fluctuate does not depend on asymmetry in the magnitude of the costs. All that matters is that the costs are present.

Figure 9 shows the constant adjustment cost case. In contrast to figure 8, the vertical distance between the schedules in figure 9 is different. In particular, the displacement of the creation schedules is less than in figure 8 and that of the destruction schedules is greater. Clearly, average creation will be less variable than average destruction in figure 9. Working out the implications of this for aggregate creation and destruction is quite difficult even in this simple example. However, the results for the employment adjustment model described in the previous section suggest that this kind of variation in the employment policies may be sufficient to account for the DHS observations on aggregate employment flows.

We now discuss briefly the model's implications for the cross-sectional evidence presented by DHS. With a large number of plants all subject to idiosyncratic uncertainty, creation and destruction at the plant level are pervasive and occur in booms and recessions (when the

FIGURE 9

Fluctuations in creation and destruction schedules with constant adjustment costs



wage is low and high, respectively), which is consistent with the DHS findings. Furthermore, the vertical distance between the employment schedules in figure 9 is smaller in a recession (high wage) than in a boom (low wage.) This suggests that the model should exhibit greater cross-sectional variability in employment changes in recessions compared with booms, which is also consistent with the DHS evidence.

This analysis establishes the potential for asymmetries in how the creation and destruction margins behave over the business cycle to account for both the plant-level and aggregate evidence on employment flows. The model sketched above was necessarily simple and abstracts from many important considerations. In the article summarized here, we built a more empirically appealing model to analyze the plausibility of the variable employment policy mechanism. Our analysis takes into account the dynamic nature of the plant owner's problem and our results are based on a well-defined industry equilibrium. Also, since the DHS evidence shows births and deaths of plants accounting for a significant fraction of creation and destruction, we allow for entry to and exit from the industry, whereas here we keep the number of plants fixed. Our findings confirm that the intuition presented above extends beyond the very simple environments we have studied, and that the basic mechanism of asymmetric fluctuations in the creation and destruction schedules may help account for other features of the aggregate employment flow data not emphasized here.

Conclusion

The evidence presented by DHS has been provocative not only because it has challenged standard theories of the business cycle, but also because the aggregate variables it describes are built directly from micro data; hence, the DHS evidence provides the opportunity to build and test models that describe genuine microeconomic foundations for macroeconomic analysis. However, much of the theoretical work developed in response to the DHS evidence has taken a distinctly conventional approach, focusing on models in which the policies of micro agents do not display the degree of heterogeneity found in the data.

The main manifestation of this is the common assumption in the theoretical literature that the average amounts created and destroyed by employment changing plants are invariant to the aggregate state of the economy. This has led researchers to emphasize model features that lead to changes in the numbers of creating and destroying plants, at the expense of model features that might influence the amounts created and destroyed at individual plants. The plant-level empirical evidence presented by DHS suggests that these averages do change over the business cycle and the version of our model described here suggests that taking into account these changes may be important for understanding the evidence.

One of the longstanding motivations of macroeconomic research is the desire to develop microeconomic foundations for macroeconomic phenomena. Our model presents a positive development in this regard, because our analysis suggests that the same friction that helps to account for the cross-sectional evidence on employment changes also seems able to account for the behavior of job creation and destruction in the aggregate. That is, the presence of proportional employment adjustment costs, which is a simple explanation for the cross-sectional evidence, may also imply that the job creation and destruction margins respond asymmetrically to aggregate shocks, which in turn may account for the aggregate evidence. Thus we have been able to establish a direct connection between detail at the micro level and the behavior of important macro aggregates.

NOTES

[1]The data on aggregate job flows are available electronically via anonymous ftp from haltiwan.econ.umd.edu.

[2]The mode of both histograms is at the growth-rate interval including zero change. The set of plants that fall into this interval include a substantial fraction that do not change employment at all. See Hammermesh (1989) and Hammermesh, Hassink, and van Ours (1994) for more evidence on the sizable fraction of establishments that fail to adjust employment over extended periods of time.

[3]These are computed using a generalized method of moments procedure. For this procedure a Bartlett window with four lags was used to estimate the spectral

density matrix at frequency zero. See Hamilton (1994, chapter 14).

[4]See DHS, chapter 5, for a similar discussion.

[5]It is straightforward to add assumptions to the household and plant problems so that labor input and employment are equivalent.

[6]Using equation 6, we have $\bar{n} = A(\bar{\theta} + \eta)^{(1-\alpha)/(\gamma-\alpha)}$ and $\underline{n} = A(\bar{\theta} - \eta)^{(1-\alpha)/(\gamma-\alpha)}$.

[7]Using equation 6, we have $\bar{n}_g = A(\theta_g + \eta)^{(1-\alpha)/(\gamma-\alpha)}$, $\underline{n}_g = A(\theta_g - \eta)^{(1-\alpha)/(\gamma-\alpha)}$, $\bar{n}_b = A(\theta_b + \eta)^{(1-\alpha)/(\gamma-\alpha)}$, and $\underline{n}_b = A(\theta_b - \eta)^{(1-\alpha)/(\gamma-\alpha)}$.

[8]See Hall (1995) for a discussion of this possibility.

[9]For example, suppose we introduce i.i.d. preference shocks that shift the aggregate labor supply curve. The main impact here would be to change the number of possibilities for aggregate outcomes for mean employment. Nevertheless, the general behavior of creation and destruction outlined above would continue to hold since this is driven by the cross-sectional distribution of employment growth.

[10]See Bertola and Caballero (1990) for a justification of the microeconomic decision rules assumed in this section.

[11]See Stokey and Lucas (1989), chapter 13.

[12]If we had assumed $\tau^c > \tau^d$, for example, then the vertical distance from the creation to the frictionless schedule would have been larger than the vertical distance between the destruction and frictionless schedules. Notice also that if one of τ^c or τ^d were zero, then the associated schedule would coincide with the frictionless schedule.

REFERENCES

Bertola, G., and R. Caballero, "Kinked adjustment costs and aggregate dynamics," in *NBER Macroeconomic Annual*, O. Blanchard and S. Fischer (eds.), 1990, pp. 237–88.

Caballero, R., "A fallacy of composition," *American Economic Review*, Vol. 82, 1992, pp. 1279–1292.

Caballero, R., and M. Hammour, "The cleansing effect of recessions," *American Economic Review*, Vol. 84, 1994, pp. 1350–1368.

Campbell, J., "Entry, exit, embodied technical change and business cycles," National Bureau of Economic Research, working paper, No. 5955, 1997.

Campbell, J., and J. Fisher, "Aggregate employment fluctuations with microeconomic asymmetries," National Bureau of Economic Research, working paper, No. 5767, 1996.

Davis, S., and J. Haltiwanger, "Gross job creation and destruction: Microeconomic evidence and macroeconomic implications," in *NBER Macroeconomics Annual*, O. Blanchard and S. Fischer (eds.), Cambridge, MA.: Massachusetts Institute of Technology Press, 1990, pp. 123–168.

Davis, S., J. Haltiwanger, and S. Schuh, *Job Creation and Destruction*, Cambridge, MA: Massachusetts Institute of Technology Press, 1996.

Foote, C., "Trend employment growth and the bunching of job creation and destruction," Harvard University, mimeo, 1995.

Hall, R., "Lost jobs," *Brookings Papers on Economic Activity*, Vol. 1, 1995, pp. 221–273.

Hamilton, J., *Times Series Analysis*, Princeton, NJ: Princeton University Press, 1994.

Hammermesh, D., "Labor demand and the structure of adjustment costs," *American Economic Review*, Vol. 79, No. 4, 1989, pp. 674–689.

Hammermesh, D., W.H.J. Hassink, and J.C. van Ours, "Job turnover and labor turnover: A taxonomy of employment dynamics," Applied Labor Economics Research Team, Vrije Universiteit Amsterdam, research memorandum, No. 94-50, 1994.

Mortensen, D., and C. Pissarides, "Job creation and destruction in the theory of unemployment," *Review of Economic Studies*, Vol. 61, 1994, pp. 397–415.

Stokey, N., and R. Lucas with E. Prescott, *Recursive Methods in Economic Dynamics*, Cambridge MA: Harvard University Press, 1989.