Federal Reserve Bank of St. Louis

REVIEW

MAY/JUNE 2011 VOLUME 93, NUMBER 3

Have Acquisitions of Failed Banks Increased the Concentration of U.S. Banking Markets?

David C. Wheelock

Patterns of Interstate Migration in the United States from the Survey of Income and Program Participation

Rubén Hernández-Murillo, Leslie S. Ott, Michael T. Owyang, and Denise Whalen

Dynamics of Externalities: A Second-Order Perspective

Yi Wen and Huabin Wu

Regional Aggregation in Forecasting: An Application to the Federal Reserve's Eighth District

Kristie M. Engemann, Rubén Hernández-Murillo, and Michael T. Owyang

What Do You Think?

Dear Review Subscriber,

The Federal Reserve Bank of St. Louis has published unbiased economic analysis free of charge for decades. We truly value our 16,000+ *Review* subscribers worldwide.

As a subscriber, you can assist us in keeping the *Review* useful and engaging.

Please give us your opinions and advice in this brief online survey:

https://www.surveymonkey.com/s/PCJ2FRV

You read our publication, so your input is essential; it will help guide our future efforts.

Thank you for your time and your readership.

Visit http://research.stlouisfed.org/ for other free publications, econ ed courses, and data & information services.

If you have any questions, please contact George Fortier, Managing Editor of *Review*: george.e.fortier@stl.frb.org.

REVIEW

Director of Research
Christopher J. Waller

Senior Policy Adviser
Robert H. Rasche

Deputy Director of Research
Cletus C. Coughlin

Review *Editor-in-Chief*
William T. Gavin

Research Economists
Richard G. Anderson
David Andolfatto
Alejandro Badel
Subhayu Bandyopadhyay
Maria E. Canon
Silvio Contessi
Riccardo DiCecio
Thomas A. Garrett
Carlos Garriga
Rubén Hernández-Murillo
Luciana Juvenal
Natalia A. Kolesnikova
Michael W. McCracken
Christopher J. Neely
Michael T. Owyang
Adrian Peralta-Alva
Juan M. Sánchez
Rajdeep Sengupta
Daniel L. Thornton
Yi Wen
David C. Wheelock

Managing Editor
George E. Fortier

Editors
Judith A. Ahlers
Lydia H. Johnson

Graphic Designer
Donna M. Stiller

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

155
Have Acquisitions of Failed Banks Increased the Concentration of U.S. Banking Markets?
David C. Wheelock

169
Patterns of Interstate Migration in the United States from the Survey of Income and Program Participation
Rubén Hernández-Murillo, Leslie S. Ott, Michael T. Owyang, and Denise Whalen

187
Dynamics of Externalities: A Second-Order Perspective
Yi Wen and Huabin Wu

207
Regional Aggregation in Forecasting: An Application to the Federal Reserve's Eighth District
Kristie M. Engemann, Rubén Hernández-Murillo, and Michael T. Owyang

Review is published six times per year by the Research Division of the Federal Reserve Bank of St. Louis and may be accessed through our website: **research.stlouisfed.org/publications/review**. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this website. These data and programs are also available through Inter-university Consortium for Political and Social Research (ICPSR) via their FTP site: www.icpsr.umich.edu/pra/index.html. Or contact the ICPSR at P.O. Box 1248, Ann Arbor, MI 48106-1248; 734-647-5000; **netmail@icpsr.umich.edu**.

Single-copy subscriptions are available free of charge. Send requests to: Federal Reserve Bank of St. Louis, Public Affairs Department, P.O. Box 442, St. Louis, MO 63166-0442, or call (314) 444-8809.

General data can be obtained through FRED (Federal Reserve Economic Data), a database providing U.S. economic and financial data and regional data for the Eighth Federal Reserve District. You may access FRED through our website: **research.stlouisfed.org/fred**.

Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Please send a copy of any reprinted, published, or displayed materials to George Fortier, Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166-0442; **george.e.fortier@stls.frb.org**. Please note: Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis. Please contact the Research Division at the above address to request permission.

© 2011, Federal Reserve Bank of St. Louis.

ISSN 0014-9187

Have Acquisitions of Failed Banks Increased the Concentration of U.S. Banking Markets?

David C. Wheelock

During 2007-10, failures eliminated 318 U.S. commercial banks and savings institutions, about 4 percent of the total number of banks operating at the end of 2006. The assets and deposits of many failed banks were acquired by institutions that already had offices in markets served by the failed banks. This article investigates the impact of in-market acquisitions of failed banks on the concentration of local U.S. banking markets. Most banks that failed during 2007-10 were small, and their acquisitions generally had little impact on market concentration. Acquisitions of larger banks that failed, such as the acquisition of Washington Mutual Bank by JPMorgan Chase Bank, also had only limited impact on the concentration of most banking markets. Among large metropolitan statistical area markets, the Houston and New York City banking markets were most affected by the acquisition of Washington Mutual, but these markets remained relatively unconcentrated after the acquisition. Hence, the article finds that except for a few rural banking markets, acquisitions of failed banks by in-market competitors generally had only a small impact on market concentration. (JEL G21, G28, G34, L11, L41)

Federal Reserve Bank of St. Louis *Review*, May/June 2011, *93*(3), pp. 155-68.

The financial crisis and recession of 2007-09 brought a sharp increase in the number of depository institution failures in the United States and elsewhere. Between January 1, 2007, and December 31, 2010, a total of 318 U.S. commercial banks and savings institutions (hereafter "banks") failed—nearly 4 percent of the total number in operation on December 31, 2006. Failed banks held approximately $436 billion of deposits—nearly 6 percent of the total deposits held in U.S. banks on December 31, 2006.[1]

Compared with firms in most other industries, banks are heavily regulated and supervised by government authorities. Furthermore, their principal form of liabilities—customer deposits—are insured by a government agency, the Federal Deposit Insurance Corporation (FDIC). Government authorities regularly examine all banks, especially those that have sustained substantial losses or show other signs of weakness that might lead to failure. Insolvent banks, or banks that appear in imminent danger of becoming insolvent, may be declared failed by their primary regulator in conjunction with the FDIC. The FDIC acts as the receiver or liquidating agent for federally insured banks that fail and is required by law to resolve each failure in the manner that imposes the least cost on the deposit insurance system.

[1] Data on failed bank and savings institution deposits are from their last available quarterly statement of condition before failure. The data reported here are for failures of U.S. chartered commercial banks and savings institutions located in the 50 states and Washington, D.C., and exclude data for Puerto Rico and other U.S. territories, as well as for one bankers' bank, one cooperative bank, and two industrial banks that failed during the period and appear in the list of bank failures on the FDIC website (www.fdic.gov/bank/individual/failed/banklist.html).

David C. Wheelock is a vice president and economist at the Federal Reserve Bank of St. Louis. The author thanks Alton Gilbert, Rajdeep Sengupta, and Adam Zaretsky for comments on a previous version of this article. David A. Lopez provided research assistance.

© 2011, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

In recent years, most bank failures have been resolved through a "purchase and assumption" (P&A) transaction in which some or all of the assets of the failed bank are sold to a healthy bank, which may also assume some or all of the failed bank's liabilities, including all insured deposits.[2] The FDIC solicits bids from potential acquirers and is required by law to select the qualifying bid that imposes the least cost on the deposit insurance fund. (The FDIC usually pays the acquirer an amount based on the excess of the value of liabilities it assumes over the value of the assets or enters into a loss-sharing agreement with the acquiring bank.) Occasionally, however, failed banks are liquidated, with the FDIC paying off insured depositors and liquidating all of the failed bank's assets. Among the 318 failures noted above, 299 were resolved through P&A transactions and 19 were resolved through a deposit payoff and liquidation or other means.[3]

This article examines how acquisitions of failed banks during 2007-10 affected the concentration of local U.S. banking markets. Concentration measures are used often to gauge the competitiveness of markets. Although the threat of entry by potential competitors may deter firms from exerting market power in industries with a small number of firms—even just one firm—a high degree of concentration is often taken as a sign of a lack of competition in a market.[4] Further, researchers have found that banks in more concentrated local markets tend to charge higher interest rates on loans and pay lower rates on deposits. They also tend to behave more conservatively and be less cost efficient than banks in less concentrated markets.[5] Bank regulators use Department of Justice guidelines for market concentration to evaluate the competitive effects of proposed bank mergers and acquisitions. Proposed transactions that would increase market concentration significantly are subject to more scrutiny and are more likely to be rejected on antitrust grounds than transactions that would not increase concentration significantly.[6] In compiling lists of potential bidders for failed banks, the FDIC considers market competition, as well as other factors. However, the FDIC does not automatically reject bids that would increase market concentration significantly.

Bank regulators use data on deposits held at bank branch offices to measure the concentration of local banking markets. Similarly, this article uses these data to examine how acquisitions of failed banks have affected the concentration of local U.S. banking markets. The merger of two banks with no geographic markets in common will have no effect on the concentration of any local banking markets. However, acquirers of many failed banks have branches in markets served by the failed bank. This article investigates how such transactions have affected the concentration of local banking markets by comparing the actual levels of concentration in those markets before a bank failure with hypothetical, or "pro forma," levels constructed by treating the acquiring and failed banks as having merged before the date of failure. Further, for markets with many bank failures or especially large increases in concentration, the article evaluates how all failed bank acquisitions by in-market competitors since 2007 affected concentration. Although acquisitions of failed banks by in-market competitors produced large increases in concentration in a few, mainly rural, banking markets, such acqui-

[2] Often, however, acquirers do not purchase the brokered deposits, whether insured or not, of failed banks.

[3] A few of the 19 bank failures not resolved through standard P&As were temporarily transferred to bridge banks, which are new, temporary national banks controlled by the FDIC. The FDIC uses bridge banks to provide the time needed to arrange a permanent transaction to resolve a failed bank. The FDIC considers bridge banks a form of P&A transactions. However, this article focuses on the impact of acquisitions of failed banks by in-market competitors and, hence, does not consider the implications of bridge bank resolutions. Additional information about the resolution of failed banks can be found in the FDIC *Resolutions Handbook* (www.fdic.gov/bank/historical/reshandbook/).

[4] If markets are "contestable" (i.e., if barriers to entry and exit are sufficiently low), then even a monopolist is unable to earn excess profits in the long run because other firms will enter and drive down the market price if the incumbent firm sets its price above marginal cost (Baumol, Panzar, and Willig, 1988).

[5] Gilbert (1984) provides a comprehensive review of research on bank market structure and competition through the early 1980s, and Berger et al. (2004) review more recent research on the relationships between market concentration, competition, and bank performance.

[6] See Gilbert and Zaretsky (2003) for more on the use of concentration in the competitive analysis of proposed bank mergers and acquisitions.

sitions did not substantially affect concentration in most U.S. banking markets.

The next section reports summary information about banks that failed during 2007-10 and the institutions that acquired them. Subsequently, the article investigates the impact of acquisitions of failed banks by in-market competitors on the concentration of local banking markets. The article identifies markets where such acquisitions resulted in large increases in concentration, as well as the effect of in-market acquisitions of failed banks on concentration in markets that had many failures or especially large increases in concentration. The final section presents conclusions.

ACQUIRERS OF FAILED BANKS

The majority (94 percent) of the 318 failures of U.S. banks between January 1, 2007, and December 31, 2010, were resolved by P&A transactions. In total, 178 banks acquired some or all of the assets and liabilities of one or more failed banks during this period. In general, the acquirers of failed banks were larger, in terms of both total assets and deposits, and operated more branch offices in more banking markets than failed banks. Table 1 lists the 10 largest failed banks and 10 largest acquirers of failed banks during 2007-10, based on total assets on June 30, 2006. Washington Mutual Bank, a federally chartered savings bank with $350.9 billion of assets on June 30, 2006, was the largest bank that failed during the period. Its acquirer, JPMorgan Chase Bank, had $1.1 trillion of assets on June 30, 2006, and was the largest acquirer of one or more failed banks during the 2007-10 period.

Table 2 reports summary information, as of June 30, 2006, for all failed banks and their acquirers.[7] Information for failed commercial banks and savings institutions is reported separately. Savings institutions traditionally have focused on residential mortgage lending, and many experienced large losses when mortgage delinquencies rose during 2007-10. Several large savings institutions, including Washington Mutual Bank, failed during this period. Although many commercial banks also succumbed to the downturn in real estate markets, failed commercial banks tended to be smaller than failed savings institutions. The average (median) total assets and deposits of failed savings institutions were $9.5 billion ($352 million) and $5.6 billion ($307 million), respectively, whereas the comparable figures for failed commercial banks were $628 million ($223 million) and $478 million ($185 million).

Acquirers of failed institutions, which included both commercial banks and savings institutions, often were much larger than the firms they acquired. The average (median) total assets and deposits of acquiring institutions were $14.7 billion ($649 million) and $7.8 billion ($466 million), respectively. Acquirers also tended to operate more branch offices in more geographic areas than acquired banks. The largest acquirer (in terms of number of branches) operated 2,679 branches in 17 states and another operated branches in 26 states. The median number of branches operated by acquirers was 11 branches. Although the largest failed banks also had many branch offices, the median number of branches operated by failed commercial banks and savings institutions were 3 and 7, respectively, and most failed institutions operated branches in fewer unique zip code regions, counties, and states than did acquirers of failed banks.[8]

IMPACT OF FAILED BANK ACQUISITIONS ON MARKET CONCENTRATION

When it compiles a list of potential acquirers for a failed bank, the FDIC "takes into account the failed institution's geographic location, competitive environment, minority-owned status,

[7] The first bank failure during 2007-10 occurred in February 2007 and, hence, I use data from June 30, 2006, in compiling the summary information presented in Table 2. Although there were 318 failed banks during the period, nine of those banks were chartered after June 30, 2006, and thus the summary information in Table 2 is based on data for only 258 failed commercial banks and 51 failed savings institutions. Likewise, 12 of the 178 acquirers were chartered after June 30, 2006, and hence the summary information for acquirers is based on only 166 observations.

[8] See Aubuchon and Wheelock (2010) for more information about the characteristics of banks that failed during 2007-10.

Table 1

Ten Largest Failed Banks and Acquiring Institutions (January 2007–December 2010), by Total Assets on June 30, 2006

Name of institution	Total assets ($ thousands)	Total deposits ($ thousands)	Total branches	Total zip codes with a branch	Total counties with a branch	Total states with a branch
Failed banks						
Washington Mutual Bank, FSB	350,890,182	210,626,236	2,167	1,745	199	15
Colonial Bank, NA	22,962,845	16,242,689	301	267	74	5
IndyMac Bank, FSB	22,743,262	9,575,579	26	25	4	1
Downey Savings and Loan Association	17,464,594	11,936,431	172	154	20	2
Guaranty Bank	16,920,624	9,362,598	152	149	49	2
Ohio Savings Bank, FSB*	16,605,531	11,188,582	56	52	8	3
BankUnited, FSB	12,866,372	6,014,740	73	62	11	1
First Federal Bank of California, FSB	10,256,842	5,542,113	32	28	2	1
Corus Bank, NA	9,369,988	8,320,397	14	10	1	1
United Commercial Bank	8,280,022	5,497,301	47	36	14	4
Acquiring institutions						
JPMorgan Chase Bank, NA	1,144,680,000	434,752,000	2,679	1,783	283	17
U.S. Bank, NA	212,553,949	117,337,830	2,525	1,873	553	26
SunTrust Bank	181,442,168	117,956,301	1,758	1,130	249	12
PNC Bank, NA	85,946,560	58,134,805	831	588	77	9
Branch Banking and Trust Company	85,214,955	51,246,133	918	718	239	10
Regions Bank	81,954,710	57,231,022	1,397	1,039	398	15
ING Bank, FSB	60,922,260	46,440,495	1	1	1	1
Manufacturers and Traders Trust Company	55,789,617	35,227,118	683	492	101	7
Union Bank, NA	50,054,634	39,743,303	340	296	40	3
Fifth Third Bank	48,259,297	38,388,498	749	547	133	6

NOTE: *On the date of its failure, Ohio Savings Bank was known as AmTrust Bank.

SOURCE: FDIC, *Summary of Deposits*.

Table 2

Descriptive Statistics for Failed Banks and Acquiring Institutions (January 2007–December 2010)

Type of bank	Total assets ($ thousands)	Total deposits ($ thousands)	Total branches	Total zip codes with a branch	Total counties with a branch	Total states with a branch
Failed commercial banks (258 banks)						
Min	11,073	3,264	1	1	1	1
Max	22,962,845	16,242,689	301	267	74	5
Mean	628,081	478,117	7.1	6.2	2.7	1.1
Q1	94,722	74,991	1	1	1	1
Median	223,099	185,124	3	3	2	1
Q3	490,770	380,402	7	6	3	1
SD	1,769,355	1,288,289	20.5	18.0	5.1	0.5
Failed savings institutions (51 banks)						
Min	5,927	3,264	1	1	1	1
Max	350,890,182	210,626,236	2,167	1,745	199	15
Mean	9,494,080	5,648,601	60.3	50.4	9.1	1.7
Q1	112,088	84,525	1	1	1	1
Median	351,608	307,481	7	7	2	1
Q3	1,765,954	1,327,204	20	15	5	2
SD	49,046,846	29,423,941	302.8	244.0	28.5	2.1
Acquiring institutions (166 banks)						
Min	14,709	7,916	1	1	1	1
Max	1,144,680,000	434,752,000	2,679	1,873	553	26
Mean	14,693,641	7,840,964	106.3	77.8	20.9	2.0
Q1	224,488	160,555	4	4	2	1
Median	648,627	465,562	11	9	4	1
Q3	2,664,282	2,345,542	43	33	12	2
SD	91,894,739	36,986,975	353.1	248.8	64.0	3.0

NOTE: Data are for June 30, 2006; Q1 and Q3 are values for the first and third quartiles of the distribution of the data; and SD is the standard deviation of the data.

SOURCE: FDIC, *Summary of Deposits*.

overall financial condition, asset size, capital level, and regulatory ratings" (FDIC, *Resolutions Handbook*, p. 9). A bank with significant operations in the markets served by a failed bank might be willing to bid more than other banks for the assets of the failed bank for several reasons. For example, an in-market bidder might have more information about the local market and the failed bank's customers, and thereby be able to serve those customers at lower cost than other potential bidders. The cost of absorbing the operations of the failed bank might also be lower for a local bidder. Further, a local bank might purchase a failed competitor to deter entry by outside competitors.

Many failed bank resolutions during 2007-10 involved the acquisition of some or all of the assets and deposits of the failed bank by an in-market competitor (i.e., another bank that already operated at least one branch in markets served by the failed bank). Only 2 of 7 failures between January 1, 2007, and June 30, 2008, were resolved by a P&A transaction in which an in-market competitor acquired the failed bank. However, 43 of 65 failures between July 1, 2008, and June 30, 2009; 84 of 175 failures between July 1, 2009, and June 30, 2010; and 42 of 71 failures between July 1, 2010, and December 31, 2010, were resolved by a P&A in which an in-market competitor acquired the failed bank. Thus, a majority of failed bank resolutions during 2007-10 resulted in some consolidation of banking assets and deposits in local banking markets.[9]

Measures of market concentration, such as *n*-firm concentration ratios and the Herfindahl-Hirschman Index (HHI), are used often to help evaluate the competitiveness of banking and other markets. Banks are required to report the amount of deposits held by each of their branch offices on June 30 of each year.[10] To enforce antitrust laws, bank regulators use these data to consider how proposed bank mergers would affect the concentration of bank deposits in local banking markets. Ordinarily, proposed mergers are not challenged on competitive grounds unless they would result in a post-merger HHI value of more than 1800 points and an increase in the index of more than 200 points.[11] These guidelines are not binding on the resolution of failed banks, but as noted previously, the FDIC does consider the competitiveness of banking markets when compiling lists of potential bidders for failed banks.

Similar to the approach used by bank regulators in antitrust enforcement, I estimate the impact of failed bank acquisitions on the concentration of local banking markets by comparing the actual values for each market in which both banks had branches with the pro forma HHI values calculated under the assumption that the acquiring and failed banks had merged before the bank failure. Obviously, the merger of two banks that share no geographic markets will have no impact on the HHI values of any local banking markets. However, the merger of banks that operate in the same markets will, all else equal, reduce the number of banks in those markets and increase the deposits held by the acquiring bank, and thereby increase the value of the HHI.

Estimates of market concentration, such as the HHI, require data on the allocation of customer deposits across bank branches. For banking markets (metropolitan statistical areas [MSAs] or county) in which an acquisition of a failed bank occurred, I use these data to calculate HHI values for the June 30 immediately preceding a bank's failure. For example, I use branch-level deposits data from June 30, 2007, to calculate HHI values for all markets in which one or more banks that failed between July 1, 2007, and June 30, 2008, were acquired by in-market competitors. I then compare actual HHI values for June 30, 2007, with

[9] This article uses the common definition of banking markets as either metropolitan statistical areas (MSAs) or non-MSA counties. However, some U.S. banking markets are defined by regulators to encompass larger or smaller areas, and market definitions are sometimes changed to reflect changes in commuting patterns, population growth, and so on. Current definitions for all U.S. banking markets are available from the Federal Reserve Bank of St. Louis (http://cassidi.stlouisfed.org/).

[10] These data are available from the FDIC (http://www2.fdic.gov/sod/index.asp).

[11] The HHI is calculated as the sum of the squared market shares of each firm competing in a market—that is, $HHI = \Sigma_i \, market\ share_i^2$, where there are $i = 1,\ldots,n$ firms in the market and $market\ share_i$ is the percentage of market output (deposits in the present context) produced by the *i*th firm. Guidelines for the use of the HHI in antitrust enforcement are established by the Department of Justice (www.justice.gov/atr/public/guidelines/6472.htm).

pro forma values calculated under the assumption that the failed and acquiring banks had already been operating as a single institution on that date.[12]

Most acquisitions of failed banks by in-market competitors during 2007-10 did not result in a substantial increase in local market concentration. Comparing actual HHI values on the June 30 before a bank failure with pro forma values derived by assuming that the failed bank and its acquirer had already merged, I find that acquisitions of failed banks by in-market competitors resulted in an average increase in HHI value of 54.1 points and a median increase of just 1.3 points. Further, I estimate that the acquisition of a failed bank by an in-market competitor increased the HHI value by more than 200 points in only 21 markets, which are listed in Table 3. As shown in the table, the 15 markets with the largest difference between the pro forma HHI and pre-failure HHI values (i.e., difference between "Pro Forma HHI" and "HHI Before Acquisition") are all sparsely populated non-MSA counties. Transactions resulting in a difference greater than 200 points also occurred in six MSAs, five of which are relatively small.[13]

Whereas Table 3 lists markets where the differences between the hypothetical and pre-failure HHI values exceed 200 points, Table 4 lists markets where the pre-failure HHI exceeded 1800 points, which is the level at which the Department of Justice considers a market to be concentrated. Terrell County, Texas, was the most concentrated U.S. banking market in which a failed bank was acquired by another in-market bank. Sanderson State Bank, which held 95.4 percent of the county's bank deposits on June 30, 2008, was closed by the Texas Department of Banking on December 12, 2008. The FDIC sold all of the failed bank's assets and deposits to The Pecos County State Bank, which as the only other bank with offices in Terrell County, held 4.6 percent of the county's deposits on June 30, 2008. The acquisition of Sanderson State Bank by its sole in-market competitor increased the HHI value for Terrell County by 875 points, from 9125 to 10000 points.

Besides Terrell County, Texas, 12 other counties had market HHI values greater than 1800 on the June 30 before a bank failure and experienced an increase in HHI value of more than 200 points due to the acquisition of a failed bank by an in-market competitor. Of course, the fact that these counties had highly concentrated banking markets that were made even more concentrated by the acquisition of failed banks does not imply that the FDIC ignored the competitive environment in soliciting bids for failed banks. Market competition is only one consideration in identifying potential bidders for failed banks, and the FDIC is required to pursue the least costly resolution of bank failures.[14]

Next, I examine the impact of acquisitions of failed banks in banking markets that had multiple in-market acquisitions within a 12-month period. Table 5 lists the markets with at least four such acquisitions. For example, between July 1, 2009, and June 30, 2010, there were 15 acquisitions of failed banks with offices in the Chicago MSA by banks that also had offices in Chicago on June 30, 2009. None of the failed banks had a large share of the Chicago banking market. Corus Bank, NA, had the largest share at only 2.5 percent. Among the acquiring banks, Harris Bank, NA, had the largest share of the Chicago market's bank deposits on June 30, 2009, at 8.6 percent. I estimate that

[12] Regulators weight the deposits of savings institutions by 0.5 when calculating HHI values for use in competitive analysis of local banking markets, reflecting a presumption that deposits at savings institutions are similar, but not perfect, substitutes for deposits at commercial banks. For this article, however, I assign full weight to savings institutions' deposits when calculating HHI values. Since the article focuses primarily on *changes* in HHI due to acquisitions of failed banks, the choice of how to weight savings institutions deposits is largely immaterial.

[13] In calculating pro forma HHI values, I assume that acquirers purchase all of the deposits of failed banks, which I estimate to equal deposits held on the most recent June 30 before failure. However, acquirers sometimes purchase less than the total deposits of failed banks, and the deposits of failed banks on the date of failure can differ substantially from the amount held by the bank on the prior June 30. Hence, the pro forma HHI values presented in this article are *estimates* of the effect of in-market acquisitions on market concentration that may overstate or understate the true effect.

[14] As shown in Table 4, some MSA banking markets have exceptionally large HHI values. However, most of these markets include the headquarters of a large national financial institution that assigns a large share of its customer deposits to the headquarters location. Because this skews measures of market concentration for those markets, regulators attempt to estimate the share of locally generated deposits in such organizations when evaluating the competitive implications of proposed mergers in affected markets. This article makes no such attempt.

Table 3

Markets Where Acquisition of a Failed Bank Increased HHI by at Least 200 Points

Market	Failed bank market share (%)	Acquirer market share (%)	HHI before acquisition*	Pro forma HHI†	Difference in HHI
County markets					
Camden, GA	39.2	14.3	2232	3352	1120
McDowell, WV	31.6	17.7	3882	5001	1119
Kay, OK	19.3	25.9	1622	2624	1002
Gilmer, GA	10.2	49.0	3010	4007	997
Terrell, TX	95.4	4.6	9125	10000	875
Crenshaw, AL	11.6	35.6	2349	3178	829
Holt, NE	12.1	32.1	2148	2927	779
Wharton, TX	13.5	28.3	1638	2405	767
Elko, NV	40.1	7.3	2329	2912	583
Montgomery, IA	9.8	28.6	2397	2957	560
Blaine, ID	43.5	5.2	2551	3001	450
Jo Daviess, IL	7.0	28.5	2084	2483	399
Tift, GA	16.7	10.8	2088	2448	360
Banks, GA	11.7	14.9	3066	3416	350
Duplin, NC	12.1	14.2	1990	2334	344
MSA markets					
Bremerton, WA	8.8	18.6	1214	1542	328
Houston, TX	5.0	29.5	1161	1454	293
Anniston, AL	13.6	10.4	1324	1606	282
Columbus, IN	67.4	1.8	4822	5062	240
Beaumont, TX	7.7	14.5	1051	1274	223
Mount Vernon, WA	8.3	13.3	1026	1246	220

NOTE: *HHI before acquisition is the value of the HHI in the market on June 30 of the year before the bank failure. †Pro forma HHI is a hypothetical value of HHI calculated under the assumption that the failed bank and acquirer had operated as a single institution on June 30 of the year before the failure.

the 15 acquisitions as a whole increased the HHI value for the Chicago market by just 24 points, from 563 to 587 points. Among the banking markets noted in Table 5, San Diego had the largest estimated increase in HHI, at just 46 points. Hence, the evidence for markets with multiple in-market acquisitions of failed banks during a 12-month period indicates that the concentration of those markets was not substantially affected by the acquisitions.

Table 6 lists the markets with at least five acquisitions of failed banks by in-market competitors over the entire period from January 1, 2007, through December 31, 2010. To gauge the impact of all acquisitions during the period on market concentration, I compare the actual HHI values for June 30, 2006, with pro forma values constructed under the assumption that the failed banks and their acquirers had merged before that date. The difference between the "HHI After Acquisitions" and "Pro Forma HHI," therefore, provides an estimate of the impact of all in-market acquisitions of failed banks during 2007-10 on market concentration. For example, 22 banks with

Table 4

Highly Concentrated Banking Markets (HHI > 1800) Where a Failed Bank Was Acquired by a Competitor

Market	Failed bank market share (%)	Acquirer market share (%)	HHI before acquisition†	Pro forma HHI‡	Difference in HHI
County markets					
Terrell, TX	95.4	4.6	9125	10000	875
McDowell, WV	31.6	17.7	3882	5001	1119
Banks, GA	11.7	14.9	3066	3416	350
Gilmer, GA	10.2	49.0	3010	4007	997
Blaine, ID	43.5	5.2	2551	3001	450
Chesterfield, SC	3.1	20.1	2397	2522	125
Montgomery, IA	9.8	28.6	2397	2957	560
Crenshaw, AL	11.6	35.6	2349	3178	829
Elko, NV	40.1	7.3	2329	2912	583
Camden, GA	39.2	14.3	2232	3352	1120
Holt, NE	12.1	32.1	2148	2927	779
Tift, GA	16.7	10.8	2088	2448	360
Jo Daviess, IL	7.0	28.5	2084	2483	399
Duplin, NC	12.1	14.2	1990	2334	344
Martinsville, VA	1.5	6.4	1958	1977	19
MSA markets					
Reno, NV	0.0*	0.0*	6803	6803	0
Bloomington, IL	1.8	3.0	6528	6538	10
Ogden, UT	0.0*	0.0*	5632	5633	1
Columbus, IN	67.4	1.8	4822	5062	240
Columbus, IN	5.5	1.8	4822	4841	19
Las Vegas, NV	0.0*	0.0*	4658	4658	0
Las Vegas, NV	0.0*	0.0*	4658	4658	0
San Antonio, TX	0.0*	0.0*	3749	3750	1
San Antonio, TX	0.0*	2.0	3749	3752	3
San Antonio, TX	0.0*	2.2	3590	3592	2
Las Vegas, NV	1.0	1.4	3129	3131	2
Trenton, NJ	0.0*	0.0*	2978	2978	0
Jacksonville, FL	0.0*	0.0*	2665	2665	0
Buffalo, NY	0.0*	1.2	2514	2514	0
Minneapolis, MN	0.0*	0.0*	2254	2254	0
Merced, CA	38.9	2.0	2161	2317	156
Jacksonville, FL	0.0*	1.1	2110	2110	0
Cincinnati, OH	0.0*	0.0*	2028	2028	0
Racine, WI	7.6	0.0*	1967	1974	7
Minneapolis, MN	0.0*	0.0*	1873	1873	0
Minneapolis, MN	0.0*	0.0*	1873	1874	1
Minneapolis, MN	0.0*	0.0*	1873	1873	0
Minneapolis, MN	0.0*	0.0*	1873	1873	0
Minneapolis, MN	0.0*	0.0*	1873	1873	0
Dallas, TX	1.8	2.9	1855	1866	11
Dallas, TX	0.0*	0.0*	1855	1855	0
Cincinnati, OH	1.0	2.7	1813	1818	5

NOTE: *Less than 1 percent. †HHI before acquisition is the value of the HHI in the market on June 30 of the year before the bank failure. ‡Pro forma HHI is a hypothetical value of HHI calculated under the assumption that the failed bank and acquirer had operated as a single institution on June 30 of the year before the failure.

Table 5

Markets with Four or More Acquisitions of Failed Banks by Competitors in a Single Year

Market*	Year†	No. of acquisitions	HHI before acquisitions‡	Pro forma HHI§	Difference in HHI
Chicago, IL	2009-10	15	563	587	24
Los Angeles, CA	2009-10	9	888	915	27
Atlanta, GA	2008-09	7	1270	1279	9
Los Angeles, CA	2008-09	7	774	781	7
Atlanta, GA	2010	7	1248	1249	1
Seattle, WA	2009-10	6	1097	1101	4
Miami, FL	2009-10	5	741	743	2
Minneapolis, MN	2009-10	5	1873	1874	1
Riverside, CA	2009-10	5	1011	1015	4
San Francisco, CA	2009-10	5	1708	1714	6
Chicago, IL	2008-09	4	537	554	17
Atlanta, GA	2009-10	4	1181	1186	5
San Diego, CA	2009-10	4	1029	1075	46

NOTE: *All markets with four or more acquisitions are MSAs. †Years are the 12-month periods July 1, 2008–June 30, 2009, and July 1, 2009–June 30, 2010, and the 6-month period July 1, 2010–December 31, 2010. ‡HHI before acquisitions is the value of the HHI in the market on June 30 of the year before the bank failures. §Pro forma HHI is a hypothetical value of HHI calculated under the assumption that the failed banks and their acquirers had operated as combined institutions on June 30 of the year before the failures.

offices in Chicago that failed during 2007-10 were acquired by banks that also had offices in Chicago. If those failed banks and their acquirers had already merged by June 30, 2006, the estimated HHI for the Chicago market would have been 633 points ("Pro Forma HHI"). By contrast, the actual HHI for the Chicago market on June 30, 2006, was 600 points ("HHI Before Acquisitions"). Hence, I estimate that the acquisitions of the 22 failed banks by in-market competitors increased the HHI of the Chicago market by 33 points, or about 5.5 percent.

Among the markets listed in Table 6, Houston experienced the largest impact of such acquisitions on market concentration. I estimate the acquisitions of failed banks by in-market competitors increased the HHI of the Houston market by 323 points, or 27.8 percent. The next-largest impact was in the New York City market, where failed bank acquisitions increased the HHI by 112 points, or 9.9 percent. In both cases, the acquisition of Washington Mutual Bank by JPMorgan Chase Bank was the main reason for the increase in HHI. The HHI values for both the Houston and New York City markets remained below 1800 after the acquisition, however, indicating that neither market was particularly concentrated.

Finally, I investigate the contribution of acquisitions of failed banks by in-market competitors to changes in the concentration of individual banking markets between 2006 and 2010. The median increase in HHI between June 2006 and June 2010 across all markets in which a failed bank was acquired by an in-market competitor was 90 points. I estimate that, absent acquisitions of failed banks by in-market competitors, the median increase in HHI would have been 10 points.[15] However, such acquisitions account for a much smaller percentage of the increase in HHI among most MSA markets than among rural markets. Among the 89 MSAs with in-market acquisitions

[15] This is the median across all markets of the difference between HHI for June 2010 and a pro forma value for June 2006 constructed under the assumption that all failed banks and their acquirers had merged before June 30, 2006.

Table 6

Markets with at Least Five Acquisitions of Failed Banks by In-Market Competitors (2007-10)

Market*	No. of acquisitions	HHI before acquisitions†	Pro forma HHI‡	Difference in HHI
Chicago, IL	22	600	633	33
Atlanta, GA	18	1412	1421	9
Los Angeles, CA	18	783	799	16
Seattle, WA	10	1330	1335	5
Miami, FL	8	833	836	3
Minneapolis, MN	7	1566	1567	1
San Francisco, CA	7	1372	1375	3
Riverside, CA	7	939	945	6
Kansas City, MO	6	408	409	1
Phoenix, AZ	5	1510	1604	94
Detroit, MI	5	1427	1430	3
Houston, TX	5	1160	1483	323
New York, NY	5	1131	1243	112
San Diego, CA	5	1001	1049	48

NOTE: *All markets with at least five acquisitions are MSAs. †HHI before acquisitions is the value of the HHI in the market on June 30, 2006. ‡Pro forma HHI is a hypothetical value of HHI for June 30, 2006, calculated under the assumption that the failed banks and their acquirers had operated as combined institutions on that date.

of failed banks, the median increase in HHI was 78 points. I estimate that absent such acquisitions, the median HHI would have increased by 50 points. By contrast, among the 27 rural markets where such acquisitions occurred, the median increase in HHI between 2006 and 2010 was 113 points. Absent those acquisitions, I estimate that the median change in HHI across those markets would have been a *decline* of 149 points. Thus, the effect of in-market acquisitions of failed banks on the concentration of rural banking markets was, on average, much greater than the effect of such acquisitions on the concentration of MSA markets.

Table 7 lists markets where the HHI value increased by at least 500 points between June 30, 2006, and June 30, 2010, and at least one failed bank was acquired by an in-market competitor. The table also shows pro forma HHI values on June 30, 2006, calculated under the assumption that banks that failed during 2007-10 had merged with their in-market acquirers before June 30, 2006 ("Pro Forma HHI 2006"). In four markets listed in Table 7, the in-market acquirers of failed banks did not have a presence on June 30, 2006, and, hence, the pro forma HHI values for 2006 are identical to the actual values. In a few rural-county markets, acquisitions of failed banks by in-market competitors can explain a high percentage of the actual increase in HHI value between 2006 and 2010. For example, for the Crenshaw County, Alabama, banking market, the actual HHI values in 2006 and 2010 were 2194 and 3382 points, respectively. The pro forma HHI value for Crenshaw County in 2006 is 3100 points. Thus, I estimate that the in-market acquisition contributed 906 points, or 76 percent, of the 1188-point increase in HHI value for Crenshaw County between 2006 and 2010. Similarly, I estimate that an in-market acquisition contributed 1012 points, or 88 percent, of the 1148-point increase in HHI value for McDowell County, West Virginia, between 2006 and 2010. Other counties for which in-market acquisitions account for large shares

Table 7

Markets with an HHI Increase of at Least 500 Points (2006-10)

Market	HHI 2006*	HHI 2010†	Increase in HHI, 2006-10	Pro forma HHI 2006‡	Increase in HHI due to acquisitions of failed banks
County markets					
Terrell, TX	5104	10000	4896	5104	—
Crenshaw, AL	2194	3382	1188	3100	906
McDowell, WV	3996	5144	1148	5008	1012
Jo Daviess, IL	1821	2574	753	2276	455
Wharton, TX	1735	2363	628	2520	785
Holt, NE	2193	2803	610	2966	773
Montgomery, IA	2415	2940	525	2415	—
Elko, NV	2007	2522	515	2518	511
MSA markets					
Reno, NV	3462	7247	3785	3462	—
Las Vegas, NV	4812	6209	1397	4815	3
San Antonio, TX	2946	4066	1120	2953	7
San Francisco, CA	1372	2332	960	1375	3
Pittsburgh, PA	1520	2431	911	1521	1
Ogden, UT	5762	6658	896	5763	1
Carson City, NV	1060	1934	874	1060	—
Jacksonville, FL	1929	2665	736	1930	1
Minneapolis, MN	1566	2254	688	1567	1
Cincinnati, OH	1442	2028	586	1455	13

NOTE: *HHI 2006 is the value of the HHI on June 30, 2006. †HHI 2010 is the value of the HHI on June 30, 2010. ‡Pro forma HHI 2006 is a hypothetical value of HHI for June 30, 2006, calculated under the assumption that the failed banks and their acquirers had operated as combined institutions on that date.

of the actual increases in HHI values include Jo Daviess County, Illinois; Wharton County, Texas; Holt County, Nebraska; and Elko County, Nevada.[16]

Aside from the few cases noted above, acquisitions of failed banks by in-market competitors can explain relatively little of the increase in HHI value between 2006 and 2010 for most markets, especially MSA markets. Among MSA markets where in-market acquirers of failed banks had offices on June 30, 2006, the largest difference between the actual and pro forma HHI values for 2006 is a mere 13 points for the Cincinnati, Ohio, market.

CONCLUSION

The structure of the U.S. banking industry has changed dramatically since the mid-1980s, when the number of U.S. banks reached a post-World War II peak. Advances in information-processing technology and the removal of most legal barriers to branch banking have been the main drivers of a substantial consolidation of the banking industry (Berger, 2003). Much of the

[16] Acquisitions of failed banks by in-market competitors can explain more than 100 percent of the actual increases in HHI values, as they do for Wharton County, Texas, and Holt County, Nebraska, because other factors, including changes in relative market shares of other banks, offset the impact of acquisitions of failed banks.

consolidation has taken the form of mergers of solvent banks. However, waves of failures, both in the late 1980s and early 1990s and again in the wake of the financial crisis and recession of 2007-09, also contributed to industry consolidation and, at least in some markets, substantial increases in concentration.[17]

Failures eliminated 318 U.S. banks during 2007-10, or about 4 percent of the total number of banks operating at the end of 2006. The assets and deposits of 171 of those failed banks were acquired by institutions that already had branches in markets served by the failed bank. Those acquisitions contributed to increased concentration of local banking markets. However, this article finds that except for a few rural banking markets, acquisitions of failed banks by in-market competitors generally had only a small impact on market concentration. Most banks that failed during 2007-10 were small, and although many of those banks were acquired by much larger institutions, those acquisitions generally had little impact on market concentration. Acquisitions of larger banks that failed during 2007-10, such as the acquisition of Washington Mutual Bank by JPMorgan Chase Bank, also had only limited impact on the concentration in most of the banking markets involved. Among large MSAs, the Houston and New York City banking markets were most affected by the acquisition of Washington Mutual but both remained relatively unconcentrated after the acquisition.[18]

[17] Data from the 1990s indicate that branching deregulation and industry consolidation did not generally increase the concentration of local urban banking markets, though it did increase concentration over larger geographic regions (Dick, 2006).

[18] This article does not investigate the impact on market concentration of acquisitions of certain distressed commercial banks, such as Countrywide Bank; Wachovia Bank, NA; and National City Bank, which were acquired by other banks without being declared failed and closed by regulators. Those acquisitions may have had significant impacts on concentration in some banking markets. Although not the subject of the present article, the data and methods used here could be applied to investigate the effects of these and other acquisitions on the concentration of individual banking markets.

REFERENCES

Aubuchon, Craig P. and Wheelock, David C. "The Geographic Distribution and Characteristics of U.S. Bank Failures, 2007-2010: Do Bank Failures Still Reflect Local Economic Conditions?" Federal Reserve Bank of St. Louis *Review*, September/October 2010, *92*(5), pp. 395-415; http://research.stlouisfed.org/publications/review/10/09/Aubuchon.pdf.

Baumol, William J.; Panzar, John C. and Willig, Robert D. *Contestable Markets and the Theory of Industry Structure*. Revised Edition. San Diego: Harcourt Brace Jovanovich, 1988.

Berger, Allen N. "The Economic Effects of Technological Progress: Evidence from the Banking Industry." *Journal of Money, Credit, and Banking*, April 2003, *35*(2), pp. 141-76.

Berger, Allen N.; Demirgüç-Kunt, Asli; Levine, Ross and Haubrich, Joseph G. "Bank Concentration and Competition: An Evolution in the Making." *Journal of Money, Credit, and Banking*, June 2004, *36*(3 Part 2), pp. 433-51.

Dick, Astrid A. "Nationwide Branching and Its Impact on Market Structure, Quality, and Bank Performance." *Journal of Business*, March 2006, *79*(2), pp. 567-92.

Federal Deposit Insurance Corporation. *Resolutions Handbook*; www.fdic.gov/bank/historical/reshandbook/.

Federal Deposit Insurance Corporation. *Summary of Deposits*; www2.fdic.gov/sod/index.asp.

Gilbert, R. Alton. "Bank Market Structure and Competition: A Survey." *Journal of Money, Credit, and Banking*, November 1984, *16*(4 Part 2), pp. 617-45.

Gilbert, R. Alton and Zaretsky, Adam M. "Banking Antitrust: Are the Assumptions Still Valid?" Federal Reserve Bank of St. Louis *Review*, November/December 2003, *85*(6), pp. 29-52; http://research.stlouisfed.org/publications/review/03/11/gilbert.pdf.

Patterns of Interstate Migration in the United States from the Survey of Income and Program Participation

Rubén Hernández-Murillo, Lesli S. Ott, Michael T. Owyang, and Denise Whalen

The authors describe the Survey of Income and Program Participation (SIPP) as a data source for migration studies. The SIPP is a panel dataset that provides information on income, employment outcomes, and participation in government programs. Survey participants are interviewed for up to four years even if they move to a new household or that household migrates within the United States. This unique longitudinal design gives the survey a strong advantage over traditional data sources. The authors illustrate differences in the propensity for interstate migration among different demographic groups over the 12-year period from 1996 to 2008. They also analyze the relationship between migration choices and life-changing events, such as becoming jobless or dissolution of a marriage. Their findings suggest that future research should consider the migration choices of individuals near retirement age. (JEL J11, J24, J61, R23)

Federal Reserve Bank of St. Louis *Review*, May/June 2011, *93*(3), pp. 169-85.

The continued rise in the unemployment rate in the aftermath of the most recent recession has been a significant concern among policymakers. Analysts have hypothesized that a mismatch in the skills of workers and the skills required for available jobs in a particular location may be hampering the opportunities for economic recovery and contributing to chronic unemployment (see, e.g., Valleta and Kuang, 2010). This problem may be exacerbated by financial difficulties following the housing crisis as unemployed workers may be unable to sell their homes and move to alternative locations where job vacancies demand the skills they can provide (see Chan, 2001, among others). Given these circumstances, understanding the determinants of the migration propensity may prove useful for policy design.

Several studies analyzing the determinants of migration have used data from the Current Population Survey (CPS). The Survey of Income and Program Participation (SIPP) is an alternative source of data that has been relatively underutilized in the migration literature.

The SIPP was designed to obtain information about the income and the participation in government programs of individuals and households in the United States. Until the SIPP was implemented in the early 1980s, the March Income Supplement (Annual Demographic Survey) of the CPS was the main source of data on income and program participation. The SIPP was designed to follow individuals for a longer period and with more detail than the CPS. The SIPP chooses households based on location and then tracks the individual household members, new individuals who join

Rubén Hernández-Murillo is a senior economist and Michael T. Owyang is a research officer at the Federal Reserve Bank of St. Louis. Lesli S. Ott is a senior statistician at the Center for Outcomes Research and Evaluation, Yale University/Yale New Haven Hospital. Denise Whalen is a researcher at Mathematica Policy Research, Inc.

© 2011, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

the household, and individuals if they move to a new household. The CPS, in contrast, fixes the physical address of the households that are interviewed and does not track households or household members if they migrate. The CPS identifies whether an individual resided in a different state in the year before the interview as well as the "state of origin," defined as the state in which the individual was living before the move (available in the Integrated Public Use Microdata Series). The SIPP also identifies the state of origin and the destination state (the state to which an individual has moved), but at a monthly frequency. The SIPP can also be used to identify moves between approximately 100 of the largest metropolitan areas in the nation. These features of the SIPP make it appealing as a source of migration data.

Dahmann (1989) was the first to recognize the SIPP as a source of data for geographic mobility research, but since then only a handful of studies have used SIPP to analyze migration issues. In an early study, Shumway (1993), using the 1984 SIPP panel, found that immigration influences the duration of unemployment spells. Additional studies have also used the SIPP after the survey was redesigned in 1996. De Jong, Graefe, and St. Pierre (2005) analyzed the effect of changes in welfare rules on interstate migration incentives of poor families. Graefe, De Jong, and May (2006) also analyzed the effects of welfare reform on the migration incentives of families with members limited in their ability to work due to an illness or disabling condition.

In this article, we argue that as a source of data for migration studies, the SIPP provides several advantages compared with traditional data sources. We illustrate the contents of the SIPP by reporting summary statistics of migration propensities across various demographic groups using pooled data from the 1996, 2001, and 2004 panels of the survey. We also evaluate migration propensities in relation to significant life events, such as changes in marital status and the onset of disability. We examine differences in local economic conditions in origin and destination locations in addition to differences in an individual's wages and employment outcomes surrounding the time of a move. Finally, we discuss potential research questions that could be analyzed using the rich set of SIPP household-level information pertaining to the migration decisions of individuals.

This article is organized as follows. In the next section, we review the literature on intranational migration. The following section describes the SIPP data. We then present summary statistics from the SIPP on overall migration rates, as well as migration rates for subgroups including various demographic groups and individuals who have experienced life-changing events. In the final section, we discuss potential avenues for migration research.

LITERATURE REVIEW

There are three main subtopics in the migration literature. The first describes how differences in the demographic characteristics of individuals, such as age, education, gender, and race, affect the probability of an individual moving to another state. The second subtopic explores the relationship between the migration decision and characteristics external to the individual, such as local labor market conditions, area quality-of-life measures, and national economic conditions. Finally, migration scholars commonly analyze differences in the labor market outcomes—wages and employment—of movers before and after a move. In this section, we discuss the migration literature organized in terms of these three approaches. The appendix provides a list of papers organized by approach and data source.

Individual Characteristics

The migration literature posits several contributors to the propensity to move, including age, gender, race, and, most popularly, education level. Many recent studies have analyzed the relationship between these characteristics and the decision to migrate. Differences in the likelihood of migration based on education levels are generally larger than the differences resulting from other factors. Therefore, we focus our literature discus-

sion on the relationship between education levels and geographic mobility.[1]

Using data from the 1980 U.S. Census, Malamud and Wozniak (2010) study the causal effects of college education on geographic mobility. They observe that an additional year of higher education is associated with a relatively large increase in the probability of residing outside one's birth state later in life. The authors find that the causal link between education and mobility is economically important, and this finding allows them to interpret geographic mobility as one of the returns to higher education. Using CPS data from 1980 to 2000, Basker (2003) also finds that the probability of migration increases with education. This relationship holds with the inclusion of age, state of origin, and year fixed effects.

Rosenbloom and Sundstrom (2003) analyze the long-run trends in interstate migration using microdata from the U.S. Census from 1850 to 1990. They find that the overall interstate migration rate in the United States has followed a U-shaped trend since 1850, falling until around 1900 and then rising until around 1970. Using probit models, they estimate that the rise in migration of families since 1990 is attributable to the rise in educational attainment of the parents.

Aggregate Characteristics

The human capital theory of migration (Sjaastad, 1962) states that the decision to migrate involves choosing among candidate labor markets and therefore comparing locations in terms of the differences in local economic conditions that affect the expectations for future earnings. An individual will take into account the discounted value of earnings at the potential destinations and compare them with the relocation costs and the time required for job searching at the potential destinations. Complementing this line of research, a recent strand of the literature also considers differences in indicators of quality of life or consumption amenities in addition to differences in labor market factors.

Theoretical models of migration show that, holding the cost of migration fixed, workers move between local markets (i) to arbitrage spatial differences in economic opportunities, measured by real wages or unemployment rates, or (ii) for personal reasons related to the life cycle or preferences.

Using data from the National Longitudinal Survey of Youth (NLSY) for 1979 to 1986, Borjas et al. (1992) found that interstate differences in the returns to skills are an important determinant of the migration decision. Regions that pay higher returns to skills attract more skilled workers than regions that pay lower returns. Their analysis emphasizes that individuals whose skills are more mismatched with the reward structure of their current state of residence are those most likely to leave that state and relocate to states that offer higher rewards for their skills. Other papers in the literature have similarly considered the relationship between migration and local labor market conditions as estimated by unemployment rates (see, e.g., Basker, 2003, who uses CPS data; Nakosteen et al., 2008, who use data from Statistics Sweden; and Haurin and Haurin, 1988, who use the NLSY and CPS).

Studies have used some unique approaches to control for differences in local conditions. For example, in considering the role of information on migration propensities for both first-time and repeat migrants, Herzog and Schlottman (1983) use the *Places Rated Almanac* to construct an index of the migrant stock, which captures the concentration of migrants from a potential mover's current location that already live in the destination location. The *migrant stock* is taken into consideration by individuals considering a move and may reduce move-related psychic and search costs. Using Internal Revenue Service (IRS) area-to-area migration flow data and a conditional logit approach to model domestic migration, Davies, Greenwood, and Li (2001) also include a measure of migrant stock. Both papers also control for distance between the origin and destination locations.

Herzog and Schlottman (1986, using Census data) and Whisler et al. (2008, using Census data and *Places Rated Almanac*) study the migration

[1] An alternative way of examining the relationship between education and migration considers how the ability of an individual to engage in labor-market arbitrage to exploit differences across locations varies in terms of skills. In this sense, migration is viewed as an investment in human capital.

decision motivated by considering differences in quality-of-life indicators of potential destination cities. In both studies, housing costs (a proxy for costs of living), recreation, and crime are all significant factors in the decision to migrate. The more recent study by Whisler et al., however, finds that a poor economic climate contributes to out-migration, whereas the earlier study found the opposite effect. Both studies highlight the importance of a variety of environmental factors on locational decisions.

In addition to the models in which locations are chosen in response to labor market arbitrage opportunities, recent studies have analyzed how workers with different levels of education and/or experience vary in their ability to take advantage of different economic opportunities that result from migration. For example, Wozniak (2010) analyzes whether highly educated workers are more likely to decide to relocate to areas with high labor demand compared with workers with low education levels. Using data from three decennial U.S. censuses, she estimates a conditional logit model of residential location choice in which workers respond to economic conditions in their current market and to conditions in all potential destinations. Wozniak's results show that college graduates are more responsive to distant labor market opportunities than less-educated workers.

Factors at the national level may also influence an individual's decision to move. Saks and Wozniak (2007) use historical reports of the CPS and IRS area-to-area migration flow data to analyze the response of migration flows to national economic conditions and the effect of changes in the distribution of local economic opportunities. The authors establish an empirical relationship between rates of internal migration and the business cycle in the United States. They find that migration is strongly procyclical, even after accounting for relative variations in local economic conditions over the business cycle. They argue that their results suggest that the net cost of moving changes systematically over the cycle. They also find that the procyclicality of migration is more closely related to labor markets than to housing market cycles.

Migration Outcomes

Having established the factors (both internal and external) that contribute to migration, numerous studies examine the impact of migration on several employment outcomes. Economists typically consider two types of outcomes, employment status and wages. Herzog and Schlottmann (1983), for example, use Census data to calculate post-migration unemployment rates as a function of individual and aggregate characteristics for both first-time migrants and non-return repeat migrants (i.e., people who move from a first location to a second and then on to a third location).[2] They find that non-return repeat migrants are less likely to be unemployed than first-time migrants. Education reduces the probability of unemployment for both groups, whereas migrants employed in blue-collar jobs before the move are more likely to be unemployed after the move. However, none of the destination variables in this study had a significant effect on post-migration unemployment. Basker (2003) compares employment rates several months after a move using the CPS and concludes that workers who have a job offer before they migrate are 16 percent more likely to be employed at that time than workers who migrate without a job offer. However, the author notes that the estimates are an upper bound due to unobservable characteristics that likely make migrants who move with a job offer more employable overall.

Several papers in the migration literature additionally focus on the impact of migration on wages. Rodgers and Rodgers (2000) use the Panel Study of Income Dynamics (PSID) to estimate the effect of migration on earnings by computing earnings for movers and a control group of nonmovers with similar characteristics.[3] Their results show positive returns to moving for every post-move year. In the final year of estimation, six years

[2] These characteristics include age, education, employment status, nature of work (white collar versus blue collar), distance between origin and destination, migrant stock in the destination state, the unemployment rate, growth in employment, and nonagricultural employment in the destination state.

[3] The authors control for individual characteristics, particularly those related to employment, as well as location characteristics of the control group. They also use fixed and random effects methods to control for unobserved characteristics.

after a move, migrants' hourly wages are 16 to 20 percent higher and annual earnings are 20 percent higher compared with non-movers. When the results are stratified by age, returns are slightly higher for men younger than 40 years of age, but migration has either no effect or a negative effect on the wages and earnings of men older than 40.

Along similar lines, Yankow (2003) uses the NLSY79 to estimate the returns to migration but uses local job changers as the control group. Results of this method indicate that immediately after a move, migration has positive returns for those with a high school diploma or less education but that wages decline for those with more than a high school education. In the longer term, however, the situation is reversed: Those with at least a high school education have significant increases in earnings 1 to 5 years after migration, and those with less education do not experience significant additional returns.

Wozniak (2010) poses the question of whether location choice has lasting effects on wages. She separately estimates equations for the propensity to migrate and location choice and then evaluates a specification that estimates a proxy for labor market conditions at the beginning of one's career. Her results indicate that a one-standard-deviation decline in state conditions can cause a decrease in wages ranging from 1 to 6 percent.

DATA

In addition to using a variety of conceptual and methodological approaches, the migration literature also examines a wide range of datasets. These datasets include the CPS (Basker, 2003, and Blank, 1988), the decennial Census data (Wozniak, 2010; Malamud and Wozniak, 2010; Whisler et al., 2008; Herzog and Schlottmann, 1986), the Integrated Public Use Microdata Series, based on Census data (Herzog and Schlottmann, 1983, and Rosenbloom and Sundstrom, 2003), the NLSY (Haurin and Haurin, 1988; Yankow, 2002/2003, 2003; Borjas et al., 1992), and the PSID (Rodgers and Rodgers, 2000, and Blackburn et al., 2010). Additionally, studies have used the Social Security Administration's Continuous Work History Sample (Greenwood, 1986) and data from the IRS (Davies et al., 2001).

In this article, we use the SIPP, a dataset that has been relatively underutilized in migration studies. The SIPP produces data that are nationally representative. The SIPP was specifically designed to study longitudinal or dynamic features of the population, such as changes in income, eligibility for and participation in transfer programs, household and family composition, and labor force participation. The survey selects households based on geographic location but then tracks individual household members for up to four years, even if they move from the initial location. It is precisely this feature of tracking individuals over time and across locations that allows us to analyze migration across the entire country.

SIPP interviews are retrospective: Individuals are interviewed every four months, so every month is accounted for. For example, an individual interviewed in October would be asked for information from July, August, September, and October. The sample of interviewed individuals is divided into four *rotation groups*, and each group is interviewed in a separate month. The set of interviews over the four-month period constitutes a *wave*, and the entire sample of consecutive interviews over the full interviewing period constitutes a *panel*. Starting in 1996, each panel spans between 9 and 12 waves, or 3 to 4 years. We consider three panels of data: 1996, 2001, and 2004, which cover the period from 1996 to 2008, although the panels do not overlap. The sample size of the 1996 panel included about 40,000 households; in the most recent panels, the sample sizes were close to 90,000 households.

The SIPP provides various geographic identifiers to aid in the study of migration, with some caveats. First, the survey identifies states of residence. It also identifies a set of about 100 metropolitan areas and generically classifies other areas as "nonmetropolitan" if the household is located outside a metropolitan area or the metropolitan area is small. However, only the 1996 and 2001 panels identify metropolitan areas. Therefore, for our analysis, we define *movers* as individuals who report living in a different state in the current month than they did in the previous month. Both the state of origin and the destination state are

identified in the SIPP.[4] This definition primarily accounts for long-distance moves but may identify short-distance moves as well (e.g., moving from St. Louis, MO, to St. Louis, IL). Our definition may also ignore long-distance moves that occur within state borders, such as a move from San Diego, CA, to San Francisco, CA. Nevertheless, the majority of the literature defines "moves" in this manner and we believe this approach is sufficient to identify a large number of long-distance moves.

The dataset highlighted in this paper, the SIPP, has many advantages in identifying movers within the United States. The main advantage we have identified is the longitudinal nature of the data, which allows researchers to identify changes in residence across states and, to some extent, within states. Other large, nationally representative longitudinal datasets include the PSID and the NLSY. Both of these datasets follow survey respondents for an extended time (the PSID could potentially follow an individual over an entire lifetime) and collect information that could be used to identify moves, including state and county of residence, and in the case of the NLSY, dates of residential moves.[5]

The SIPP offers several advantages over the PSID and NLSY. The first is frequency of data collection; the SIPP collects data three times per year, while the NLSY collects data once per year and the PSID currently collects data every other year. The second advantage over the NLSY is that the SIPP covers a range of individuals, whereas the NLSY samples from an age cohort spanning only four years in the most recent data collection. As mentioned above, the PSID and NLSY have previously been used to study internal migration, given the detailed migration data contained in the surveys. Depending on the scope of research and population of interest, all three datasets may provide valuable information on internal migration. In this paper, we present the advantages of working with the SIPP.

[4] In the 1996 and 2001 panels, Maine and Vermont are grouped together and North Dakota, South Dakota, and Wyoming are grouped together.

[5] After 1993, the PSID stopped including information on county of residence.

MIGRATION STATISTICS FROM THE SIPP

In this section, we use the SIPP to identify simple correlations between individual and aggregate characteristics that influence migration. It is important to emphasize that these statistics are not causal; instead, we are noting relationships that may be important to explore more formally.

We begin by presenting statistics that influence the propensity to move. These correlates are divided into two categories: (i) population groups, denoting different demographic characteristics of the movers, and (ii) life-changing events. The demographic characteristics considered include personal traits that are either permanent (such as gender and race) or categorical for a certain period of an individual's life (such as marital status, education level, or employment). The life-changing events capture important changes that an individual may experience, such as unemployment or birth of a child.

Characteristics of Movers and Stayers

Tables 1A and 1B present statistics on the propensity to move for various population groups and describes the characteristics of both movers and *stayers* (individuals who do not move during the survey period). Overall, we identified 208,472 unique individuals, among whom 7,823 (about 3.8 percent of the total sample) moved at least once during the 3- to 4-year sample period. Table 1A is organized as follows: Column 1 lists the total number of individuals in each subgroup, column 2 shows the number of movers in each subgroup, and column 3 reports propensities to migrate by subgroup.[6] Table 1A reveals significant differences across many of these demographic groups. For example, women have a higher propensity to move than men. Individuals also move less frequently as they age. This finding is in line with

[6] For the purposes of this article, we ignore the survey weights that would allow us to construct nationally representative migration probabilities and simply report propensities in terms of the individuals interviewed in the survey panels. Additionally, the migration events we identify occur on a *monthly* basis, and therefore the propensities reported cannot be compared directly with the *annual* migration rates usually reported using other datasets.

Table 1A

Propensity to Move

Characteristic	(1) *N*	(2) No. of movers	(3) % of movers
Total sample	208,472	7,823	3.8
Men	100,753	3,502	3.5
Women	107,719	4,321	4.0
Age 18-24	26,690	1,353	5.1
Age 25-54	141,294	5,634	4.0
Age 55-65	40,488	836	2.1
White	169,375	6,499	3.8
Non-white	39,097	1,324	3.4
High school dropout	27,316	699	2.6
High school graduate	69,297	1,916	2.8
Some college	63,474	2,470	3.9
Bachelor's degree or above	48,385	2,738	5.7
Never married	53,439	2,298	4.3
Married	117,321	4,244	3.6
Divorced/separated	32,202	1,147	3.6
Widowed	5,510	134	2.4
No children	119,851	4,393	3.7
1-2 children	68,762	2,618	3.8
3 or more children	19,859	812	4.1
Disabled (at any point in survey)	42,915	1,592	3.7
Never disabled	165,557	6,231	3.8
Homeowner (before the move)	138,385	3,516	2.5
Non-homeowner	70,087	4,307	6.1
Jobless	59,688	3,105	5.2
Employed	148,784	4,718	3.2

labor theory on migration, which posits that migration is a form of human capital investment and older workers have a shorter time to collect returns to such an investment. Additionally, those with a higher education level move more frequently. This fact is also explained by theory suggesting that even when the percentage difference in regional wages is equal across education levels, the absolute difference is likely higher for the highly educated, who typically command higher salaries. The propensity to move also differs by marital status: Those who have never been married move most frequently and those who are widowed move least frequently. These differences, however, may simply be detecting variations in age. The largest difference within subcategories is between homeowners and non-homeowners; the latter move about twice as often as homeowners. Differences also occur according to employment status: Jobless individuals move more frequently than people who are employed. The smallest differences in moving rates are 3.7 percent for people with disabilities and 3.8 percent for people without disabilities.

Table 1B illustrates the distribution of the overall population (column 1), movers (column 2),

Table 1B

Characteristics of Movers and Stayers

Characteristic	(1) % of group in sample	(2) % of movers	(3) % of stayers
Total sample	100.0	100.0	100.0
Men	48.3	44.8	48.5
Women	51.7	55.2	51.5
Age 18-24	12.8	17.3	12.6
Age 25-64	67.8	72.0	67.6
Age 55-65	19.4	10.7	19.8
White	81.2	83.1	81.2
Non-white	18.8	16.9	18.8
High school dropout	13.1	8.9	13.3
High school graduate	33.2	24.5	33.6
Some college	30.4	31.6	30.4
Bachelor's degree or above	23.2	35.0	22.7
Never married	25.6	29.4	25.5
Married	56.3	54.3	56.4
Divorced/separated	15.4	14.7	15.5
Widowed	2.6	1.7	2.7
No children	57.5	56.2	57.5
1-2 children	33.0	33.5	33.0
3 or more children	9.5	10.4	9.5
Disabled (at any point in survey)	20.6	20.4	20.6
Never disabled	79.4	79.6	79.4
Homeowner (before move)	66.4	44.9	67.2
Non-homeowner	33.6	55.1	32.8
Jobless	28.6	39.7	28.2
Employed	71.4	60.3	71.8

and stayers (column 3) over the various demographic subgroups considered. Because moves are relative rare events, stayers more closely resemble the overall population than movers. For example, only 8.9 percent of movers are high school dropouts, while 13.3 percent of stayers and 13.1 percent of the overall population are high school dropouts. Similarly, only 44.9 percent of movers are homeowners, while 66.4 percent of the overall population and 67.2 percent of stayers are homeowners. The largest differences between movers and stayers occur across age and education levels in terms of homeownership and employment status.

Propensity to Move and Life-Changing Events

Table 2 illustrates the propensity to move in conjunction with the following life-changing events: becoming divorced, separated, or widowed; getting married; having a first child; having an additional child; having the oldest child turn 5; becoming disabled; becoming jobless; changing employers; and becoming employed. Column 1 indicates the baseline population of individuals present in the sample for a 6-month window of observations around the month of the life-changing event, while columns 2 and 3 indicate the number

Table 2

Propensity to Move and Life-Changing Events

Life-changing event	(1) Individuals with ±6 months of observations around life-changing event*	(2) No. of people who moved within ±6 months of life-changing event†	(3) % in group who move‡	(4) No. of movers with life-changing event within ±6 months of the move§	(5) % of movers with life-changing event¶
Becoming divorced/ separated/widowed	4,087	142	3.5	248	5.1
Getting married	7,068	202	2.9	382	7.9
Having first child	6,377	134	2.1	271	5.6
Having an additional child	4,959	69	1.4	155	3.2
Oldest child turning 5	8,941	352	3.9	518	10.8
Becoming disabled	14,422	85	0.6	352	7.3
Becoming jobless	30,782	547	1.8	2,174	45.1
Changing employers	21,550	344	1.6	1,872	38.9
Becoming employed	28,609	169	5.9	1,963	40.7

NOTE: *The sample of individuals with life-changing events in column (1) includes only those individuals with ±6 months of observations around the month of the life-changing event. Similarly, column (4) lists only movers with a life-changing event within ±6 months of the move. †Column 2 notes the number who move within a ±6-month window for the narrowed sample of individuals with a life-changing event. ‡Column 3 illustrates the incidence of a life-changing event within a ±6-month window for the narrowed sample of movers. §Column 5 illustrates the propensity of a life-changing event to occur within a ±6-month window for the sample of movers. ¶The denominator for the entries in column 5 is 4,818 movers who are present in the sample for all months within a ±6-month window around the time of the move. ±6 months indicates a window 6 months before and after an event.

and percent of individuals who moved at any time within 6 months of experiencing the life-changing event.

At first glance, it appears that the majority of life-changing events are associated with low moving probabilities compared with the 3.8 percent of the total sample who move (see Table 1A). However, recall that the overall moving propensity calculated in Table 1 occurs over a three- to four-year period and the moving propensities in Table 2 represent a one-year period.

Individuals who become disabled have the lowest moving propensity over the one-year window surrounding a life-changing event: just 0.6 percent.[7] Becoming jobless (1.8 percent) and changing jobs (1.6 percent) are also associated with low moving propensities. In contrast, dissolution of a marriage through separation, divorce, or death is associated with a relatively high propensity to move: 3.5 percent. Getting married is associated with a 2.9 percent propensity to move. When a household's oldest child turns 5, there is almost a 4 percent probability of moving. Five is the age at which most children begin school, and this statistic could indicate that parents move (perhaps to a better school district) when their children are about to start school. Finally, the life-changing event with the highest moving probability is becoming employed, with almost a 6 percent probability of moving.

Table 2 also illustrates the presence of life-changing events among those who have moved. Column 4 indicates the individuals who moved and experienced a life-changing event at any time during a 6-month window around the month of the move, and column 5 shows the incidence of the life-changing event. The baseline population in this exercise is only movers who are present in the sample in all months during a 6-month window around the time of the move, and we exclude movers who are present for a shorter period. The events with higher incidence among movers are becoming jobless (45.1 percent), changing employers (38.9 percent), and becoming employed (40.7 percent).[8]

As clarification, column 2 represents the propensity for members of the general population to move when they experience a life-changing event, and column 5 represents the propensity for those who have moved to have experienced a life-changing event.

Differences in Local Economic Conditions between Origin and Destination

Since external factors may also influence migration, in Table 3A we compare state-level characteristics of the state of origin with those of the destination state.[9] Surprisingly, there seems to be little difference between the majority of the characteristics. For example, the average school rankings and average foreclosure rates are practically identical across locations. Furthermore, the differences seem to indicate that people move to slightly *worse* locations. Compared with the origin states, the destination states have about 1.5 percent lower average incomes and about 1.6 percent higher average homicide rates per 100,000 individuals. The average unemployment rates appear to be slightly lower in the destination states—4.9 percent compared with 5.0 percent in the origin states. A more careful analysis of these differences would need to consider additional amenities, such as weather. In the latter part of the twentieth century, cities with warm, dry climates dominated the list of fastest-growing areas in terms of population growth.

Tables 3B and 3C present the differences in the same measures of local economic conditions as in Table 3A further subdivided by age and education levels, respectively. The patterns are largely similar to those in Table 3A. Unemployment rates are slightly lower and homicide rates slightly higher in destination states than in origin

[7] In a recent paper that uses the SIPP, Graefe, De Jong, and May (2006) find that although families of individuals with disabilities are less likely to migrate than other families in general, they are "pushed" to migrate if they live in states where they are not exempt from welfare work activity requirements.

[8] Movers may experience multiple life-changing events surrounding a move. For example, a mover may become jobless in January and find a new job and move in March. In this case, the mover would be counted as experiencing both becoming jobless and becoming employed.

[9] The unemployment rate is calculated at the state level during the *month* of the move. Other statistics are calculated at the state level during the *year* of the move.

Table 3A

Characteristics of Origin and Destination States

Characteristics	Origin	Destination	Difference
Average unemployment rate (%)			
Mean	5.0	4.9	–0.10
SD	1.1	1.1	–0.02
Median	4.9	4.9	0
Average annual income ($US)			
Mean	31,511	31,039	–472
SD	6,347	5,984	–363
Median	30,679	30,521	–158
Average foreclosure rate (%)			
Mean	1.14	1.15	0.01
SD	0.59	0.58	–0.01
Median	1.06	1.10	0.03
Average school ranking*			
Mean	229.6	229.7	0.17
SD	9.9	9.4	–0.44
Median	230.0	230.0	0
Average homicide rate†			
Mean	4.29	4.36	0.07
SD	1.08	1.07	–0.01
Median	4.32	4.46	0.14

NOTE: *Information on average school ranking was available only for about 7,200 movers. †Homicide rates are per 100,000 individuals.

states across all age groups and education levels. Table 3B indicates that the largest differences between origin and destination states occur in the 55- to 65-year-old age group, who experience the largest absolute differences in averages between the origin and destination states in all categories. Across education levels, Table 3C illustrates that high school dropouts experience the largest differences in average income, foreclosure rates, and homicide rates between the origin and destination states. For individuals who move, we found only minor differences between average unemployment rates, income, foreclosure rates, school rankings, and homicide rates in the origin and destination states. One potential reason for this finding is that our focus on state-level data could be missing the variation in measures of location quality that exist across cities and counties within a state. Another possibility is that individuals may select their destination locations based on the reason for their move. For example, a family facing a decline in income who moves will likely relocate to a location with relatively fewer amenities, whereas a family who moves after having children will likely move to a location with relatively more amenities, such as good schools. These opposite motives would create the appearance of little to no change in location characteristics.

Labor Market Outcomes of Movers

The previous tables list conditions that may increase an individual's propensity to migrate. Here we turn to post-move outcomes. Although individuals may migrate for a variety of personal reasons, we focus on the economic outcomes surrounding a move. In particular, we look at

Table 3B

Characteristics of Origin and Destination States by Age

	Origin				Destination			
Characteristic	*N*	Mean	SD	Median	*N*	Mean	SD	Median
Average unemployment rate (%)								
Age 18-24	1,353	5.0	1.1	5.0	1,353	5.0	1.1	4.9
Age 25-54	5,634	5.0	1.1	4.9	5,634	4.9	1.0	4.9
Age 55-65	836	5.0	1.1	4.9	836	4.9	1.0	4.9
Average annual income ($US)								
Age 18-24	1,353	30,668	5,928	30,144	1,353	30,826	5,746	30,469
Age 25-54	5,634	31,543	6,402	30,667	5,634	31,062	5,993	30,557
Age 55-65	836	32,665	6,436	32,324	836	31,233	6,292	30,592
Average foreclosure rate (%)								
Age 18-24	1,353	1.16	0.60	1.08	1,353	1.12	0.56	1.06
Age 25-54	5,634	1.14	0.60	1.06	5,634	1.15	0.58	1.10
Age 55-65	836	1.10	0.58	1.05	836	1.17	0.60	1.11
Average school ranking*								
Age 18-24	1,252	229.4	9.7	230.0	1,240	229.7	9.3	230.0
Age 25-54	5,159	229.5	9.9	230.0	5,173	229.6	9.4	230.0
Age 55-65	788	230.6	9.6	231.0	793	230.9	9.5	232.0
Average homicide rate†								
Age 18-24	1,353	4,298	1,062	4,348	1,353	4,383	1,052	4,488
Age 25-54	5,634	4,288	1,086	4,293	5,634	4,352	1,064	4,439
Age 55-65	836	4,241	1,090	4,290	836	4,353	1,123	4,543

NOTE: The number of observations for average school rankings differs between origin and destination because of missing data. *Information on average school ranking was available only for about 7,200 movers. †Homicide rates are per 100,000 individuals.

wages and employment during a three-month window before and after the move.

Table 4A presents individual wages for various demographic groups in the origin and destination states. Wages in destination states increase relative to wages in origin states by about 0.4 percent on average. Wages for men decline slightly, by $0.03 per hour or about 0.1 percent, while wages for women increase by $0.24 per hour or about 1.2 percent. Wages for all age groups tend to decrease in the destination states; the exception is wages for the 55- to 65-year age group, for whom wages increase by $2.91 per hour or almost 10 percent. This group includes individuals who are near retirement age but are still employed at the time of the move. These individuals may be willing to move close to the end of their career only if the move involves a higher wage. In terms of race, wages for whites increase, while wages for non-whites decrease in the destination states. Interestingly, wages in destination states tend to increase for all education groups except those who completed high school but did not complete college.

Table 4B shows that the employment rate falls for all subgroups of movers analyzed. Employment rates decline the most for women, older people, minorities, and individuals with a high school diploma or some college. The generalized decline in employment rates is likely consistent with a short period of adjustment surrounding the move. For example, an individual may move to a location

Table 3C

Characteristics of Origin and Destination States by Education

	Origin				Destination			
Characteristic	*N*	Mean	SD	Median	*N*	Mean	SD	Median
Average unemployment rate (%)								
High school dropout	699	5.1	1.0	5.0	699	5.0	1.0	5.0
High school graduate	1,916	5.0	1.1	5.0	1,916	4.9	1.0	4.9
Some college	2,470	5.0	1.1	5.0	2,470	4.9	1.1	4.9
Bachelor's degree or above	2,738	4.9	1.1	4.9	2,738	4.9	1.1	4.9
Average annual income ($US)								
High school dropout	699	30,761	5,922	30,093	699	29,911	5,639	29,586
High school graduate	1,916	31,136	6,280	30,479	1,916	30,635	5,823	30,185
Some college	2,470	31,601	6,270	31,082	2,470	31,102	5,957	30,619
Bachelor's degree or above	2,738	31,885	6,536	31,124	2,738	31,554	6,147	30,679
Average foreclosure rate (%)								
High school dropout	699	1.17	0.58	1.12	699	1.14	0.55	1.10
High school graduate	1,916	1.15	0.59	1.09	1,916	1.17	0.59	1.13
Some college	2,470	1.13	0.60	1.05	2,470	1.15	0.58	1.09
Bachelor's degree or above	2,738	1.13	0.59	1.05	2,738	1.14	0.59	1.06
Average school ranking*								
High school dropout	637	228.2	9.9	229.0	623	228.5	9.5	229.0
High school graduate	1,762	229.3	9.8	230.0	1,762	229.8	9.4	230.0
Some college	2,285	230.0	10.0	230.0	2,292	230.4	9.2	231.0
Bachelor's degree or above	2,515	229.8	9.7	230.0	2,529	229.4	9.6	230.0
Average homicide rate†								
High school dropout	699	4.41	1.04	4.39	699	4.52	1.11	4.61
High school graduate	1,916	4.35	1.09	4.36	1,916	4.39	1.05	4.50
Some college	2,470	4.29	1.09	4.33	2,470	4.40	1.06	4.52
Bachelor's degree or above	2,738	4.20	1.08	4.16	2,738	4.26	1.07	4.28

NOTE: The number of observations for average school rankings differs between origin and destination because of missing data.
*Information on average school ranking was available only for about 7,200 movers. †Homicide rates are per 100,000 individuals.

several months before he or she begins employment. Alternatively, the statistics may reflect individuals who move to a new location in search of a job.

CONCLUSION

In this article, we contend that the Survey of Income and Program Participation (SIPP) is a useful source of data for migration research. The SIPP collects longitudinal information on demographic and employment characteristics over a period of three to four years that can be used to study the determinants of migration choices. The SIPP also identifies both the origin and destination locations, whereas other surveys identify only the destination.

We present summary statistics on the migration propensities of individuals across several demographic characteristics to illustrate some of the individual-level information available in the

Table 4A

Wages of Movers

		Before†			After†			Difference	
Wages*	*N*	Mean	SD	Median	Mean	SD	Median	Mean	Median
All movers	3,878	24.22	60.39	16.25	24.31	59.62	16.08	0.10	–0.17
Men	2,077	27.86	74.37	18.75	27.83	74.53	18.75	–0.03	0.00
Women	1,801	20.01	37.98	13.98	20.25	34.93	13.33	0.24	–0.64
Age 18-24	632	14.25	17.55	10.61	14.20	17.21	10.80	–0.05	0.19
Age 25-54	3,013	25.86	67.13	17.50	25.77	65.63	17.59	–0.09	0.09
Age 55-65	233	30.04	35.49	20.00	32.94	47.93	19.05	2.91	–0.95
White	3,254	24.59	64.66	16.42	24.96	64.48	16.24	0.37	–0.18
Non-white	624	22.27	29.35	15.63	20.93	19.95	15.16	–1.34	–0.47
High school dropout	214	13.02	15.49	8.91	13.54	15.56	9.62	0.52	0.71
High graduate	835	15.27	13.62	11.83	15.46	17.29	11.67	0.19	–0.16
Some college	1,257	20.34	43.52	15.00	19.15	24.95	14.40	–1.19	–0.60
Bachelor's degree or above	1,661	32.16	82.78	22.23	33.11	86.75	22.50	0.94	0.27

NOTE: *Hourly wage figures are reported in dollar amounts and calculated as monthly earnings divided by the product of usual hours worked per week and the number of weeks the respondent held a job. †Before and after wages are designated as the maximum wage observation within a ±3-month window of the move. Only respondents with positive earnings before and after the move were considered.

Table 4B

Employment Rates of Movers

		Before		After		Difference
Employment rate	*N*	Mean	SD	Mean	SD	Mean
All movers	6,662	0.79	0.41	0.74	0.44	–0.05
Men	2,968	0.89	0.32	0.86	0.35	–0.03
Women	3,694	0.70	0.46	0.64	0.48	–0.06
Age 18-24	992	0.81	0.39	0.80	0.40	–0.01
Age 25-54	4,942	0.81	0.39	0.77	0.42	–0.04
Age 55-65	728	0.58	0.49	0.45	0.50	–0.13
White	5,544	0.79	0.41	0.75	0.44	–0.05
Non-white	1,118	0.76	0.43	0.69	0.46	–0.06
High school dropout	554	0.59	0.49	0.55	0.50	–0.04
High school graduate	1,597	0.73	0.44	0.68	0.47	–0.05
Some college	2,099	0.80	0.40	0.74	0.44	–0.06
Bachelor's degree or above	2,412	0.86	0.35	0.82	0.38	–0.04

NOTE: Before and after employment status is determined by whether or not a person held a job within a ±3-month window of the move. Employment status is a binary variable, so the median is either 0 or 1, and we omit it.

SIPP and how it relates to the study of interstate migration. We also describe migration propensities in conjunction with significant life events surrounding the time of the move. Finally, we describe differences in local economic conditions between origin and destination states and differences in labor market outcomes before and after a move.

Among our findings we note that, in addition to previously known facts regarding the higher migration propensity of younger or more educated individuals, there are notable differences in migration propensities (i) between men and women, (ii) in terms of marital status, (iii) in terms of employment status, and (iv) between homeowners and non-homeowners. We also found differences in migration propensities after changes in the family, such as a marriage separation, divorce, widowhood, or children attaining school age. There appear to be only small differences in local economic conditions between origins and destinations; if anything, individuals seem to move to less desirable locations—those with lower average income and higher crime rates.

The SIPP provides a rich set of individual- and household-level information that can improve the study of migration. Data from the 1996, 2001, and 2004 panels suggest that future avenues for research should consider especially the migration choices of those aged 55 to 65, as the state characteristics and individual employment rates and wages between the origin and destination states exhibit the most variability for this subgroup before and after a move.

REFERENCES

Basker, Emek. "Education, Job Search and Migration." Working paper, University of Missouri–Columbia, April 2003; http://web.missouri.edu/~baskere/papers/migration.pdf.

Blackburn, McKinley. "Internal Migration and the Earnings of Married Couples in the United States." *Journal of Economic Geography*, January 2010, *10*(1), pp. 87-111.

Blank, Rebecca. "The Effect of Welfare and Wage Levels on the Location Decisions of Female-Headed Households." *Journal of Urban Economics*, September 1988, *24*(2), pp. 186-211.

Borjas, George; Bronars, Stephen and Trejo, Stephen. "Self-Selection and Internal Migration in the United States." *Journal of Urban Economics*, September 1992, *32*(2), pp. 159-85.

Chan, Sewin. "Spatial Lock-in: Do Falling House Prices Constrain Residential Mobility?" *Journal of Urban Economics*, May 2001, *49*(3), pp. 567-86.

Dahmann, Donald C. "The Survey of Income and Program Participation as a Source of Migration Data." *Journal of Economic and Social Measurement*, 1989, *15*(1), pp. 57-70.

Davies, Paul; Greenwood, Michael and Li, Haizheng. "A Conditional Logit Approach to U.S. State-to-State Migration." *Journal of Regional Science*, May 2001, *41*(2), pp. 337-60.

De Jong, Gordon F.; Graefe, Deborah Roempke and St. Pierre, Tanja. "Welfare Reform and Interstate Migration of Poor Families." *Demography*, August 2005, *42*(3), pp. 469-96.

Graefe, Deborah Roempke; De Jong, Gordon F. and May, Dee C. "Work Disability and Migration in the Early Years of Welfare Reform." *Population Research and Policy Review*, August 2006, *25*(4), pp. 353-68.

Greenwood, Michael; Hunt, Gary and McDowell, John. "Migration and Employment Change: Empirical Evidence on the Spatial and Temporal Dimensions of the Linkage." *Journal of Regional Science*, May 1986, *26*(2), pp. 223-34.

Haurin, Donald and Haurin, Jean. "Net Migration, Unemployment, and the Business Cycle." *Journal of Regional Science*, May 1988, *28*(2), pp. 239-54.

Herzog, Henry Jr. and Schlottmann, Alan. "Migrant Information, Job Search, and the Remigration Decision." *Southern Economic Journal*, July 1983, *50*(1), pp. 43-56.

Herzog, Henry Jr. and Schlottmann, Alan. "The Metro Rating Game: What Can Be Learned from the Recent Migrants?" *Growth and Change*, January 1986, *17*(1), pp. 37-50.

Malamud, Ofer and Wozniak, Abigail. "The Impact of College Education on Geographic Mobility: Evidence from the Vietnam Generation." Harris School of Public Policy Studies Working Paper No. 0811, University of Chicago, 2010; http://harrisschool.uchicago.edu/about/publications/working-papers/pdf/wp_08_11.pdf.

Nakosteen, Robert; Westerlund, Olle and Zimmer, Michael. "Migration and Self-Selection: Measured Earnings and Latent Characteristics." *Journal of Regional Science*, October 2008, *48*(4), pp. 769-88.

Rodgers, Joan and Rodgers, John. "The Effect of Geographic Mobility on Male Labor-Force Participants in the United States." *Journal of Labor Research*, January 2000, *21*(1), pp. 117-32.

Rosenbloom, Joshua and Sundstrom, William. "The Decline and Rise of Interstate Migration in the United States: Evidence from the IPUMS, 1850-1990." NBER Working Paper No. 9857, National Bureau of Economic Research, July 2003; www.nber.org/papers/w9857.pdf.

Saks, Raven and Wozniak, Abigail. "Labor Reallocation over the Business Cycle: New Evidence from Internal Migration." Finance and Economics Discussion Series Paper No. 2007-32, Board of Governors of the Federal Reserve System, April 2007; www.nd.edu/~awaggone/documents/saks_wozniak_apr07_001.pdf.

Shumway, J. Matthew. "Factors Influencing Unemployment Duration with a Special Emphasis on Migration: An Investigation Using SIPP Data and Event History Methods." *Papers in Regional Science*, April 1993, *72*(2), pp. 159-76.

Sjaastad, Larry. "The Costs and Returns to Human Migration." *Journal of Political Economy*, October 1962, *70*(5), pp. 80-93.

Valleta, Rob and Kuang, Katherine. "Is Structural Unemployment on the Rise?" Federal Reserve Bank of San Francisco *Economic Letter*, Number 2010-34, November 8, 2010; www.frbsf.org/publications/economics/letter/2010/el2010-34.html.

Whisler, Ronald; Waldorf, Brigitte; Mulligan, Gordon and Plane, David. "Quality of Life and the Migration of the College-Educated: A Life-Course Approach." *Growth and Change*, March 2008, *39*(1), pp. 58-94.

Wozniak, Abigail. "Are College Graduates More Responsive to Distant Labor Market Opportunities?" *Journal of Human Resources*, Fall 2010, *45*(4), pp. 944-70.

Yankow, Jeffrey. "Does Distance Matter? A Comparison of Boundary and Distance-Based Measures of Internal Migration." *Journal of Economics and Social Measurement*, January 2002/2003, *28*(3), pp. 161-75.

Yankow, Jeffrey. "Migration, Job Change, and Wage Growth: A New Perspective on the Pecuniary Return to Geographic Mobility." *Journal of Regional Science*, August 2003, *43*(3), pp. 483-516.

APPENDIX

A Summary of the Literature

Paper	Topic	Data source
	Mobility research using SIPP data	
Dahmann (1989)	Geographic mobility	SIPP
De Jong, Graefe, and St. Pierre (2005)	Effect of welfare reform on interstate migration for poor families	SIPP
Graefe, De Jong, and May (2006)	Effects of welfare reform on migration incentives for the work-disabled	SIPP
Shumway (1993)	Immigration and unemployment duration	SIPP
	Individual characteristics	
Basker (2003)	Causal effects of college education on geographic mobility	CPS
Malamud and Wozniak (2010)	Causal effects of college education on geographic mobility	Census
Rosenbloom and Sundstrom (2003)	Long-run trends in interstate migration with microdata	Census
	Aggregate characteristics	
Saks and Wozniak (2007)	Response of migration flows to national and local economic conditions	CPS and IRS
Basker (2003)	Relationship between migration and local labor market conditions, unemployment rates	CPS
Haurin and Haurin (1988)	Relationship between migration and local labor market conditions, unemployment rates	NLSY and CPS
Borjas, Bronars, and Trejo (1992)	Effect of interstate differences in returns to skills on migration decision	NLSY
Davies, Greenwood, and Li (2001)	Relationship between migration and local labor market conditions, unemployment rates	IRS
Herzog and Schlottman (1983)	Relationship between migration and local labor market conditions, unemployment rates	*Places Rated Almanac*
Whisler et al. (2008)	Role of quality-of-life indicators on migration decision	Census and *Places Rated Almanac*
Herzog and Schlottman (1986)	Role of quality-of-life indicators on migration decision	Census
Wozniak (2010)	Relationship between migrant educational attainment and labor market demand in new location	Census
Nakosteen et al. (2008)	Relationship between migration and local labor market conditions, unemployment rates	Statistics Sweden
	Migration outcomes	
Basker (2003)	Post-migration employment rate	CPS
Herzog and Schlottmann (1983)	Post-migration employment status and wage outcomes	Census
Wozniak (2010)	Lasting effects of location choice on wages	Census
Rodgers and Rodgers (2000)	Effect of migration on earnings	PSID
Yankow (2003)	Returns to migration	NLSY79

NOTE: CPS, Current Population Survey; IRS, Internal Revenue Service; NLSY, National Longitudinal Survey of Youth; PSID, Panel Study of Income Dynamics; SIPP, Survey of Income and Program Participation.

Dynamics of Externalities: A Second-Order Perspective

Yi Wen and Huabin Wu

First-order approximation methods are a standard technique for analyzing the local dynamics of dynamic stochastic general equilibrium (DSGE) models. Although linear methods yield quite accurate solutions for a broad class of DSGE models, some important economic issues (e.g., portfolio choice and welfare) cannot be adequately addressed by first-order methods. This paper provides yet another case when first-order methods may be inadequate for capturing the business cycle properties of a DSGE model. In particular, the authors show that increasing returns to scale (due to production externalities) may induce asymmetric business cycles and nonlinear income effects that are not fully captured by linear approximation methods. For example, hump-shaped output dynamics can emerge even when externalities are below the threshold level required for indeterminacy, and output expansion tends to be smoother and longer, whereas contraction tends to be deeper but shorter-lived, as observed in the U.S. economy. (JEL C63, E0, E32)

Federal Reserve Bank of St. Louis *Review*, May/June 2011, *93*(3), pp. 187-205.

The standard approach to studying the business cycle implications of dynamic stochastic general equilibrium (DSGE) models is to focus on the models' local dynamics near a steady state through linear (first-order) approximations (such as the log-linearization method of King, Plosser, and Rebelo, 1988). It is well known that for standard real business cycle (RBC) models with constant returns, first-order approximation methods provide quite accurate solutions and higher-order methods give almost identical predictions.

The central twist of this paper is the addition of increasing returns to scale (IRS) caused by production externalities. We show that this simple deviation from standard RBC models generates nontrivial nonlinearities that are not well captured by first-order methods. Importantly, these nonlinearities are increasing in the degree of externalities over parameter ranges that predict a unique bounded rational expectations equilibrium. The model does not rely on local indeterminacy to generate new and interesting dynamics, although one contribution of the paper is the documentation of model properties when the model gives rise to local indeterminacy.[1]

Specific results of interest are that technology shocks generate asymmetric effects on business cycles. These effects come from second-order components of the model, which are ignored by linear approximation methods. Conditional on a positive technology shock, hump-shaped impulse response functions are predicted for employment and output, consistent with much empirical

[1] "Local indeterminacy" means that there are multiple rational expectations equilibrium paths that converge to the same steady state. Note that local indeterminacy is not the same thing as multiple steady-state equilibria. Indeterminacy can arise in a model with a unique steady state.

Yi Wen is an assistant vice president and economist at the Federal Reserve Bank of St. Louis and professor of economics at Tsinghua University. Huabin Wu is a doctoral student at Shanghai University of Finance and Economics. The authors thank Adrian Peralta-Alva, Massimo Guidolin, and Bill Gavin for comments and Mingyu Chen for research assistance.

© 2011, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

work.[2] Smooth, prolonged dynamics are observed. In contrast, conditional on a negative technology shock, the model predicts sharp, less persistent dynamics. These predictions combined provide an explanation of the strong asymmetry of the business cycle in the U.S. economy.

The existing literature shows IRS are an important source of dynamics not only for endogenous growth (e.g., Romer, 1986), but also for the business cycle. Baxter and King (1991) show that incorporating production externalities into a standard RBC model generates a better overall fit of the model to U.S. data, especially under aggregate demand shocks. Benhabib and Farmer (1994), Farmer and Guo (1994), Wen (1998a), and Benhabib and Wen (2004) show that IRS can generate endogenous business cycles if externalities are large enough to make the model's steady state locally indeterminate.[3]

However, this segment of the literature has relied on first-order approximation methods to investigate the models' dynamic behaviors. Although it has been shown that first-order approximation usually yields quite accurate results for standard RBC models (such as those of Kydland and Prescott, 1982, and King, Plosser, and Rebelo, 1988), it is less clear whether such methods retain their accuracy in describing a model's local dynamics when market failures and non-convexities, such as externalities and IRS, are involved.

IRS can greatly amplify the impact of shocks (Baxter and King, 1991) and dramatically change a model's topology around its steady state so that complicated dynamics, such as bifurcations, discontinuous jumps, and complex eigenvalues, may emerge (Benhabib and Farmer, 1994, and Coury and Wen, 2009). Therefore, it is interesting to investigate whether first-order approximation methods continue to yield accurate predictions of a model's dynamics when production externalities are present.

In addition, one of the most important aspects of stochastic dynamic models—risk—cannot be captured by linear solution methods. For example, optimal portfolio decisions cannot be analyzed by linear approximation methods. For reasons such as this, second-order solution methods have been proposed and developed in the recent literature.[4]

In this paper, we apply the second-order approximation method developed in the existing literature to analyze the local dynamics of an RBC model with externalities. We show that allowing for second-order terms not only improves the accuracy of approximations, but also changes the predicted local dynamics of the model dramatically when IRS exist. In particular, the magnitude of the impulse responses to positive technology shocks is significantly smaller and smoother under the second-order approximation method than the first-order method, and hump-shaped impulse responses can emerge even with degrees of externalities that are too small to trigger indeterminacy. This finding is in sharp contrast to the results of Benhabib and Wen (2004), who find that hump-shaped dynamics emerge under the first-order approximation method only when the degree of externalities is large enough that the model becomes locally indeterminate.

The new findings on the dynamic effects of externalities are driven by the fact that externalities introduce a large nonlinear income effect, which is captured by a large negative coefficient in front of the squared term for technology shocks and a large positive coefficient in front of the cross term for capital and technology shocks. This implies that technology shocks have the following asymmetric second-order effects on aggregate output dynamics: A positive technology shock generates smooth and hump-shaped output responses because the variance of technology shocks neutralizes the shock on impact, but the covariance of capital and technology shocks enhances (and thus propagates) the effects of the positive shock in the subsequent periods. Conversely, a negative technology shock generates much larger and sharper declines in output

[2] See, for example, Cogley and Nason (1995) and Wen (1998a, 1998b).

[3] Also see Cooper and Johri (1997) and Wen (1998b) for business cycle implications of externalities in RBC models without indeterminacy.

[4] For a review of this literature, see, for example, Judd (1998); Jin and Judd (2002); Collard and Juillard (2001); Schmitt-Grohé and Uribe (2004); Swanson, Anderson, and Levin (2006); Lombardo and Sutherland (2007); and Kim et al. (2008), among others.

because the variance of technology shocks greatly amplifies the negative shock, while the covariance term reduces the shock's negative impact in the subsequent periods; this causes the impulse responses of output to monotonically increase toward the steady state from below. Such an asymmetric nonlinear effect is similar to that observed in the U.S. economy: On the one hand, economic expansion tends to be smooth and gradual while economic contraction tends to be sharp and short-lived (see, e.g., Neftçi, 1984; Sichel, 1993; Kim and Piger, 2002; McKay and Reis, 2008; and Morley and Piger, forthcoming). On the other hand, unexpected increases in oil prices (equivalent to negative total factor productivity shocks) tend to have a large adverse impact on aggregate output, whereas decreases in oil prices tend to have only a small or negligible impact on output (e.g., see Hamilton, 2003, and Mork, 1989).

Our analysis also helps to explain a puzzle in the indeterminacy literature. For example, in the Benhabib-Farmer (1994) model, positive technology shocks are extremely expansionary before indeterminacy arises but suddenly become excessively contractionary when the model becomes locally indeterminate, thereby generating sharp falls in output, investment, and hours. Using a second-order method, we find that (i) positive technology shocks have a second-order negative impact on hours worked and (ii) this nonlinear income effect increases with the degree of IRS. Hence, as externalities increase, the second-order income effect gradually dominates the first-order substitution effect, and the initial responses of hours gradually become negative around the point of indeterminacy. Once the model becomes indeterminate, the coefficients of the second-order terms for labor become zero, so the strong income effects are captured instead by the first-order terms.[5]

An important caveat is that this paper's main point is not to find frictions to improve the empirical fit of existing RBC models with the business cycle. Hence, moment matching is not our goal. Rather, we try to provide a new scenario that linear solution methods may yield inaccurate descriptions of the model's dynamics. Moreover, given that the literature has shown that IRS may be important for understanding the business cycle based on first-order approximations, we are also interested in understanding whether IRS can introduce nonlinear dynamics not captured by first-order methods.

The literature has shown that linear approximation methods may result in quantitatively significant biases even for standard RBC models without externalities. For example, Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) study the econometrics of computed dynamic models and the consequences for inference of the use of approximated likelihoods. They find that second-order approximation errors in the policy function, which are completely ignored by first-order approximation methods, have first-order effects on the likelihood function.[6] The problem of numerical simulations of dynamic economies with heterogeneous agents and economic distortions is also studied in the existing literature. For example, Feng et al. (2009) and Peralta-Alva and Santos (2010), among others, discuss problems related to the existence and computation of Markovian equilibria in economies with heterogeneous agents and market distortions, as well as convergence and accuracy properties of numerical solutions.[7]

The remainder of the paper is organized as follows: The next section presents the model with externalities and introduces the second-order perturbation method. A simple example is provided to illustrate the difference between first-order and second-order methods. We then examine the model's dynamics with and without indeterminacy. Next, we compare the accuracy of linear and second-order approximations in the presence of IRS and then offer our conclusions.

[5] See discussions in the subsection entitled "Second-Order Effects of Sunspots."

[6] However, Ackerberg, Geweke, and Hahn (2009) show by counterexample that this conclusion by Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) is false and argue that second-order approximation errors in the policy function have at most second-order effects on parameter estimates.

[7] More references to this literature can also be found in these two articles.

THE MODEL

Our model is similar to those of Baxter and King (1991) and Benhabib and Farmer (1994).[8] There exists a continuum of identical agents in the unit interval [0,1]. A typical or representative agent chooses consumption (c_t), hours worked (n_t), and capital stock (k_{t+1}) to solve

$$\max_{\{c_t, n_t, k_{t+1}\}} E_t \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t^{1-\tau}}{1-\tau} - a \frac{n_t^{1+\gamma}}{1+\gamma} \right),$$

subject to the resource constraint,

$$c_t + k_{t+1} - (1-\delta) k_t \le A_t \left(\bar{k}_t^{\alpha} \bar{n}_t^{1-\alpha} \right)^{\eta} k_t^{\alpha} n_t^{1-\alpha},$$

where $\{\bar{k}_t, \bar{n}_t\}$ denote, respectively, the average economy-wide capital stock and hours that are taken as given by individuals and A_t denotes aggregate technology shocks. The model exhibits IRS at the social level if the externality parameter $\eta > 0$.[9]

The equilibrium of the model is determined by the following necessary conditions:

(1) $$a n_t^{\gamma} = c_t^{-\tau} (1-\alpha) A_t k_t^{\alpha(1+\eta)} n_t^{(1-\alpha)(1+\eta)-1}$$

(2) $$c_t^{-\tau} = \beta E_t \left\{ c_{t+1}^{-\tau} \left(\alpha A_{t+1} k_{t+1}^{\alpha(1+\eta)-1} n_{t+1}^{(1-\alpha)(1+\eta)} + 1 - \delta \right) \right\}$$

(3) $$c_t + k_{t+1} - (1-\delta) k_t = A_t k_t^{\alpha(1+\eta)} n_t^{(1-\alpha)(1+\eta)}$$

(4) $$\log A_t = \rho \log A_{t-1} + \sigma \varepsilon_t,$$

where equation (1) is the optimal labor supply condition, equation (2) equates the marginal cost of reducing consumption in period t and the marginal gain of consumption in the next period, equation (3) is the aggregate resource constraint in equilibrium, and equation (4) specifies the dynamics of A_t. The parameter σ controls the variance of the innovation $\varepsilon_t \sim N(0,1)$ and measures the level of uncertainty in the economy. Since agents are identical, in general equilibrium the individual variables are equal to their aggregate counterparts (e.g., $k_t = \bar{k}_t$ and $n_t = \bar{n}_t$).

As shown by Benhabib and Farmer (1994), this model has a unique steady state that is a saddle if the degree of externalities is small and a sink if η is large enough. However, Coury and Wen (2009) show that this class of models may have multiple dynamic equilibria (such as stable n-period cycles) away from the steady state even if the steady state appears to be a saddle. Mindful of this fact, in the following analyses we choose the value of η sufficiently below the critical values found by Coury and Wen (2009) for n-period cycles.

Second-Order Taylor Expansion

The model's equilibrium is solved by the second-order approximation method. The variables in the above equations can be grouped into two types: the state variables and the control variables. The state variables include the capital stock and the realized exogenous shock at the beginning of each period, denoted by the vector $\mathbf{s}_t = (k_t, A_t)$. The control variables include consumption and hours, denoted by $\mathbf{z}_t = (n_t, c_t)$.[10]

Because of certainty equivalence, the equilibrium paths of economic variables determined by linear approximation methods are independent of the degree of uncertainty (σ) and the dynamic impulse responses of the model are symmetric with respect to the sign of the shocks (A_t). That is, the policy functions can be written as

(5) $$\mathbf{z}_t = g(\mathbf{s}_t),$$

and the state variables follow the law of motion

(6) $$\mathbf{s}_{t+1} = h(\mathbf{s}_t) + \zeta \sigma \varepsilon_{t+1},$$

where the 2 × 1 vector $\zeta = [01]'$ since technological innovations do not directly affect the next-period capital stock.

[8] We also studied the model of Wen (1998a) in which the degree of externalities required for indeterminacy is much smaller because of variable capacity utilization, and our conclusions remain robust. To simplify the analysis, we choose the simpler model of Benhabib and Farmer (1994) with fixed capacity utilization even though this model requires an implausibly large degree of IRS to generate a significant difference in local dynamics from those of standard RBC models.

[9] Benhabib and Farmer (1994) prove that this representative-agent model with externalities is equivalent to a decentralized market economy with increasing returns at the firm level.

[10] If indeterminacy occurs, period-t consumption can be treated as a state variable since the steady state is a sink.

However, under second-order approximations, second-order terms such as

$$E_t\hat{A}_{t+1}^2 = \frac{\sigma^2}{1-\rho^2} \text{ and } E_t\hat{A}_{t+1}\hat{k}_{t+1}$$

emerge, so the policy functions $g(\cdot)$ and equilibrium paths of state variables $h(\cdot)$ are dependent on σ. Accordingly, the general policy functions are characterized by

$$(7)\quad \mathbf{z}_t = g(\mathbf{s}_t, \sigma)$$

$$(8)\quad \mathbf{s}_{t+1} = h(\mathbf{s}_t, \sigma) + \zeta\sigma\varepsilon_{t+1}.$$

Note that when the standard deviation (SD) of the shock $\sigma = 0$, there is no uncertainty in the model and all variables remain in the non-stochastic steady state. Therefore, the impulse responses of the system to a technology shock obtained by first-order and second-order methods converge to each other as σ approaches zero.

Define $\hat{\mathbf{z}}_t \equiv \log\mathbf{z}_t - \log\mathbf{z}_0$, $\hat{\mathbf{s}}_t \equiv \log\mathbf{s}_t - \log\mathbf{s}_0$. Following the literature (see, e.g., Schmitt-Grohé and Uribe, 2004), the policy functions $\{g(\cdot), h(\cdot)\}$ can be approximated by a second-order Taylor expansion,

$$(9)\quad \begin{aligned} \hat{\mathbf{z}}_t &= \begin{bmatrix} g_s & g_\sigma \end{bmatrix} \begin{bmatrix} \hat{\mathbf{s}}_t \\ \sigma \end{bmatrix} + \\ &\frac{1}{2}\begin{bmatrix} \hat{\mathbf{s}}_t & \sigma \end{bmatrix} \begin{bmatrix} g_{ss} & 0 \\ 0 & g_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{s}}_t \\ \sigma \end{bmatrix}, \end{aligned}$$

$$(10)\quad \begin{aligned} \hat{\mathbf{s}}_{t+1} &= \begin{bmatrix} h_s & h_\sigma \end{bmatrix} \begin{bmatrix} \hat{\mathbf{s}}_t \\ \sigma \end{bmatrix} + \\ &\frac{1}{2}\begin{bmatrix} \hat{\mathbf{s}}_t & \sigma \end{bmatrix} \begin{bmatrix} h_{ss} & 0 \\ 0 & h_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{s}}_t \\ \sigma \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \sigma\varepsilon_{t+1}; \end{aligned}$$

where the first derivatives g_s and g_σ are known from the solution to the first-order system. By taking second-order derivatives of the first-order conditions with respect to s and σ, we can obtain a linear equation system with the unknown elements in cross-derivative coefficients g_{ss}, h_{ss}, $g_{\sigma\sigma}$, and $h_{\sigma\sigma}$. The cross terms are symmetric and equal to zero, $g_{s\sigma} = g_{\sigma s} = h_{s\sigma} = h_{\sigma s} = 0$, as shown by the literature (e.g., see Schmitt-Grohé and Uribe, 2004). The second-order coefficient matrices are three dimensional. For example, since we have two elements in $g(\mathbf{s}_t, \sigma)$ and two elements in $\mathbf{s}_t$, g_{ss} is thus a $2 \times 2 \times 2$ matrix.

The steady state under second-order methods differs from that under linear methods. For example, if the initial state is zero,

$$\hat{\mathbf{s}}_0 = 0, \text{ then } \hat{\mathbf{s}}_1 = \frac{1}{2}h_{\sigma\sigma}\sigma^2;$$

consequently, the system will evolve and rest at another steady state. Moreover, at the non-stochastic steady state where $\hat{\mathbf{z}}_t$ and $\hat{\mathbf{s}}_t$ are zero, the policy functions differ from those under linear solutions by a constant term proportional to σ^2.

Accordingly, to generate impulse response functions and time-series comparable to those under linear methods, we can find the stochastic steady state by numerical simulation and then introduce shocks to generate time-series or impulse responses relative to the stochastic steady-state values.[11] Furthermore, we should note that direct use of equations (9) and (10) can generate an exploding path since second-order terms result in unnecessary higher-order terms in consecutive iteration. The pruning process proposed by Kim et al. (2008) is useful for overcoming this problem. In fact, this process uses only first-order parts of the response to generate the second-order terms in the recursive computation.

A Simple Example

Suppose an economy is described by the following nonlinear equation:

$$(11)\quad y_t = E_t A_{t+1} y_{t+1}^{\theta}, \quad |\theta| < 1;$$

where y_t is an endogenous jump variable and A_{t+1} an exogenous driving process satisfying

$$(12)\quad \log A_t = \rho \log A_{t-1} + \sigma\varepsilon_t$$

and $\varepsilon_t \sim N(0,1)$, where σ denotes the SD of ε_t. The model has two steady states: $\bar{y} = 0$ and $\bar{y} = 1$. We

[11] We depart from Schmitt-Grohé and Uribe (2004) and Kim et al. (2008) by adjusting the steady state so that the impulse responses return to zero in the long run.

consider the second steady state because it is "saddle stable." To facilitate Taylor expansions around a steady state, denote $\hat{y}_t \equiv \log y_t - \log \bar{y}$ and rewrite equation (11) as

$$(13)\quad e^{\log y_t} = E_t e^{\log A_{t+1} + \theta \log y_{t+1}}.$$

First-Order Method. The first-order Taylor expansion of equation (13) around the steady state $\{\log \bar{y} = 0, \log \bar{A} = 0\}$ is given by

$$(14)\quad \hat{y}_t = E_t\left[\hat{A}_{t+1} + \theta \hat{y}_{t+1}\right].$$

Using the method of undetermined coefficients, we can guess the solution $\hat{y}_t = \gamma_A \hat{A}_t + \gamma_\sigma \sigma$ and substitute the solution into equation (14) to obtain

$$(15)\quad \gamma_A \hat{A}_t + \gamma_\sigma \sigma = E_t\left[\hat{A}_{t+1} + \theta\left(\gamma_A \hat{A}_{t+1} + \gamma_\sigma \sigma\right)\right].$$

Applying the law of motion (equation 12) and comparing the coefficients of $\{\hat{A}_t, \sigma\}$ on both sides of equation (15) gives

$$\gamma_A = \frac{\rho}{1-\theta_\rho} \text{ and } \gamma_\sigma = 0.$$

Therefore, the first-order accurate solution is given by

$$(16)\quad \hat{y}_t = \frac{\rho}{1-\theta\rho}\hat{A}_t.$$

The impulse response function of $\hat{y}_t$ to a 1-SD shock in ε_t can be generated from the state-space representation of the model,

$$(17)\quad \begin{pmatrix} \hat{y}_t \\ \hat{A}_t \end{pmatrix} = \begin{pmatrix} 0 & \rho\gamma_A \\ 0 & \rho \end{pmatrix}\begin{pmatrix} \hat{y}_{t-1} \\ \hat{A}_{t-1} \end{pmatrix} + \begin{pmatrix} \gamma_A \\ 1 \end{pmatrix}\sigma\varepsilon_t,$$

by setting $\hat{y}_0 = 0$, $\hat{A}_0 = 0$, $\sigma = 1$, $\varepsilon_1 = 1$, and $\varepsilon_t = 0$ for all $t > 1$.

Second-Order Method. The second-order Taylor expansion of equation (13) around the steady state $\{\log \bar{y} = 0, \log \bar{A} = 0\}$ is given by

$$(18)\quad \hat{y}_t + \frac{1}{2}\hat{y}_t^2 = E_t\left[\hat{A}_{t+1} + \theta\hat{y}_{t+1} + \frac{1}{2}\left(\hat{A}_{t+1} + \theta y_{t+1}\right)^2\right].$$

We can guess a second-order solution with undetermined coefficients:

$$(19)\quad \hat{y}_t = \gamma_A \hat{A}_t + \gamma_{AA}\hat{A}_t^2 + \gamma_\sigma \sigma + \gamma_{\sigma\sigma}\sigma^2.$$

Notice that using this second-order solution as a substitute in equation (18) would generate higher-order terms such as $\{\hat{A}_t^x, \hat{y}_t^x, \sigma^x\}$ with $x \geq 3$. Since these higher-order terms are irrelevant for our second-order solution, they can be ignored. That is, we can substitute the second-order terms in equation (18) with the first-order solution in equation (16) while keeping the first-order terms as they are in equation (18). Substituting the linear solution (equation (16)) into the second-order terms in equation (18) yields

$$(20)\quad \hat{y}_t = E_t\left[\hat{A}_{t+1} + \theta\hat{y}_{t+1} + \frac{1}{2}\left(\hat{A}_{t+1} + \frac{\theta\rho}{1-\theta\rho}\hat{A}_{t+1}\right)^2\right] - \frac{1}{2}\left(\frac{\rho}{1-\theta\rho}\hat{A}_t\right)^2.$$

Since $E_t\hat{A}_{t+1}^2 = \rho^2 A_t^2 + \sigma^2$, the above equation reduces to

$$(21)\quad \hat{y}_t = \theta E_t \hat{y}_{t+1} + \rho\hat{A}_t + \frac{1}{2(1-\theta\rho)^2}\sigma^2.$$

Substituting the conjectured second-order solution (19) into equation (21) and simplifying gives

$$(22)\quad \begin{aligned} &\gamma_A\hat{A}_t + \gamma_{AA}\hat{A}_t^2 + \gamma_\sigma\sigma + \gamma_{\sigma\sigma}\sigma^2 = \\ &\rho(1+\theta\gamma_A)\hat{A}_t + \theta\gamma_\sigma\sigma + \theta\rho^2\gamma_{AA}\hat{A}_t^2 + \\ &\left(\theta\gamma_{AA} + \theta\gamma_{\sigma\sigma} + \frac{1}{2(1-\theta\rho)^2}\right)\sigma^2. \end{aligned}$$

Comparing the coefficients on both sides gives

$$\gamma_A = \frac{\rho}{1-\theta_\rho},\ \gamma_\sigma = 0,\ \gamma_{AA} = 0, \text{ and}$$

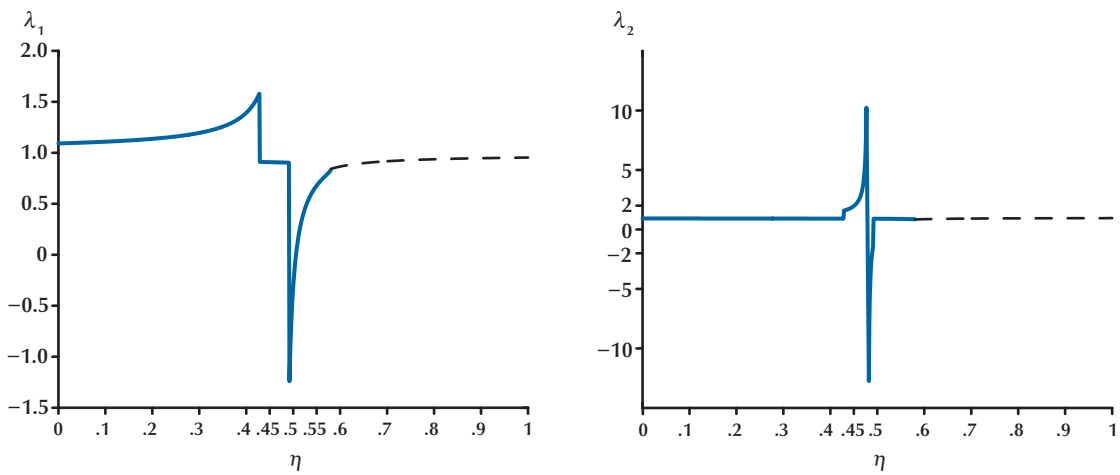
$$\gamma_{\sigma\sigma} = \frac{1}{2(1-\theta)(1-\theta\rho)^2}.$$

Hence, the second-order rational expectations equilibrium is given by

$$(23)\quad \hat{y}_t = \frac{\rho}{1-\theta\rho}\hat{A}_t + \frac{1}{2(1-\theta)(1-\theta\rho)^2}\sigma^2.$$

Figure 1

Eigenvalues as Functions of the Externality Parameter η



Note that the second-order solution conforms to the general solution in equation (9). However, in the special model the second-order effect from $\hat{A}_t^2$ does not exist and the solution differs from the first-order solution only by a constant proportional to σ^2. Thus, the corresponding impulse response function is also the same as that of the first-order solution (up to a constant term in the steady state).

Calibration and Eigenvalues

We calibrate the model based on the existing literature. In particular, we set the following: the discount factor $\beta = 0.99$, the elasticity of intertemporal substitution $\tau = 1$, the inverse elasticity of labor supply $\gamma = 0$ (indivisible labor), capital's share $\alpha = 0.3$, and the persistence of shock $\rho = 0.9$.

The eigenvalues of the model are given by the linear terms and thus are not affected by higher-order terms. Hence, the region of local indeterminacy is not influenced by the variance of technology shocks under second-order expansion. Excluding the exogenous driving process, the system has two eigenvalues. When externalities are small, only one of the eigenvalues lies inside the unit circle; thus, the steady state is a saddle. When externalities are large enough that $\eta > \eta^* = 0.4935$, both eigenvalues lie inside the unit circle and the system is indeterminate (Benhabib and Farmer, 1994). Figure 1 plots the two eigenvalues as functions of the externality parameter η. The critical value for indeterminacy is $\eta^* = 0.4935$.[12]

The dashed lines in Figure 1 correspond to the region of complex eigenvalues. The figure clearly shows that the eigenvalues undergo dramatic changes as the degree of externalities (η) increases. The explosive root jumps from positive infinity to negative infinity, while the stable root goes through similar topological changes near the critical value η^*. Both eigenvalues become stable and form a complex conjugate pair (dashed lines) for the region $\eta > \eta^*$. Such dramatic changes in eigenvalues indicate that (i) the topology of the model near the steady state is significantly altered by externalities and (ii) linear approximation may not be accurate enough to capture the curvature of the equilibrium path near the steady state.

[12] Wen (1998a) showed that introducing capacity utilization can reduce the critical value η^* to about 0.1, which is more consistent with empirical estimates. Our results apply to the model of Wen (1998a), but we choose the simpler Benhabib-Farmer (1994) model for exposition purposes in this paper.

DYNAMIC ANALYSIS

Impulse Responses

The impulse responses of the economy to a technology shock with SD $\sigma = 0.3$ are graphed in Figure 2.[13] For comparison, the left panels show the impulse responses under the first-order approximation method and the right panels show them under the second-order method. The top panels show the results when externalities are absent ($\eta = 0$), and the bottom panels show them when externalities are large enough ($\eta = 0.4$) but well below the critical value of 0.4935 needed for indeterminacy.[14]

It is clear from the top panels in Figure 2 that the first-order and second-order methods yield very similar results when externalities are not present, confirming the literature's findings that linear solution methods provide reasonably good approximations for standard RBC models. For example, the initial impulse response of output is about 0.7 under first-order approximation and about 0.68 under second-order approximation.

However, when externalities exist, the two solution methods yield dramatically different results. The bottom panels in Figure 2 indicate that the initial impulse response of output is about 1.6 under first-order approximation but only about 0.75 under second-order approximation. More importantly, while the responses of output and labor remain monotonic under the first-order method, they all become hump-shaped under the second-order method, suggesting a much richer internal propagation mechanism.

When the level of externalities increases, the hump-shaped responses become even more prominent. For η close to the critical value η^*, even a very small value of σ can give rise to hump-shaped dynamics under the second-order method. It is thus evident that production externalities dramatically change the topology of the model near the steady state. Benhabib and Wen (2004) show that under the linear approximation method, hump-shaped impulse responses and oscillating cycles emerge when the externalities are large enough to make the model indeterminate. Here, we show that such nonlinear dynamics may already exist when the degree of externalities is below the critical value required for indeterminacy, but they can be captured only by higher-order terms.[15]

The above results are obtained because the coefficients of the second-order terms in a standard RBC model without externalities are generally very small (close to zero); hence linear methods usually yield quite accurate solutions. However, once externalities or increasing returns are allowed, the second-order terms can become non-negligible and very large. For example, when $\sigma = 0.3$ and $\eta = 0.4$, we have the following decision rules for capital, consumption, and labor:

$$\begin{bmatrix} \hat{\mathbf{k}}_{t+1} \\ \hat{\mathbf{c}}_t \\ \hat{\mathbf{n}}_t \end{bmatrix} = \begin{pmatrix} 0.9124 & 0.5484 \\ 0.4325 & 0.9101 \\ -0.6272 & 4.4930 \end{pmatrix} \begin{bmatrix} \hat{\mathbf{k}}_t \\ \hat{\mathbf{A}}_t \end{bmatrix} + \tag{24}$$

$$\frac{1}{2}\begin{pmatrix} 0.0463 & -0.2652 & 0.7531 \\ 0.0143 & -0.1810 & 0.3924 \\ -0.7155 & 9.0494 & -19.6185 \end{pmatrix} \begin{bmatrix} \hat{\mathbf{k}}_t^2 \\ \hat{\mathbf{k}}_t \hat{\mathbf{A}}_t \\ \hat{\mathbf{A}}_t^2 \end{bmatrix}$$

$$+\frac{1}{2}\begin{pmatrix} 2.4241 \\ -0.4161 \\ 20.8073 \end{pmatrix} \sigma^2.$$

Note that in the labor decision rule the coefficient of $\hat{A}_t^2$ is –19.6185 and that of the cross term $\hat{k}_t\hat{A}_t$ is 9.0494, but these coefficients are close to zero when $\eta = 0$. The coefficients of $\hat{A}_t^2$ and $\hat{k}_t\hat{A}_t$ for hours worked increase in absolute value as η increases toward the critical value η^*, but both become zero again as soon as $\eta > \eta^*$. This is shown in Figure 3. Labor has no second-order

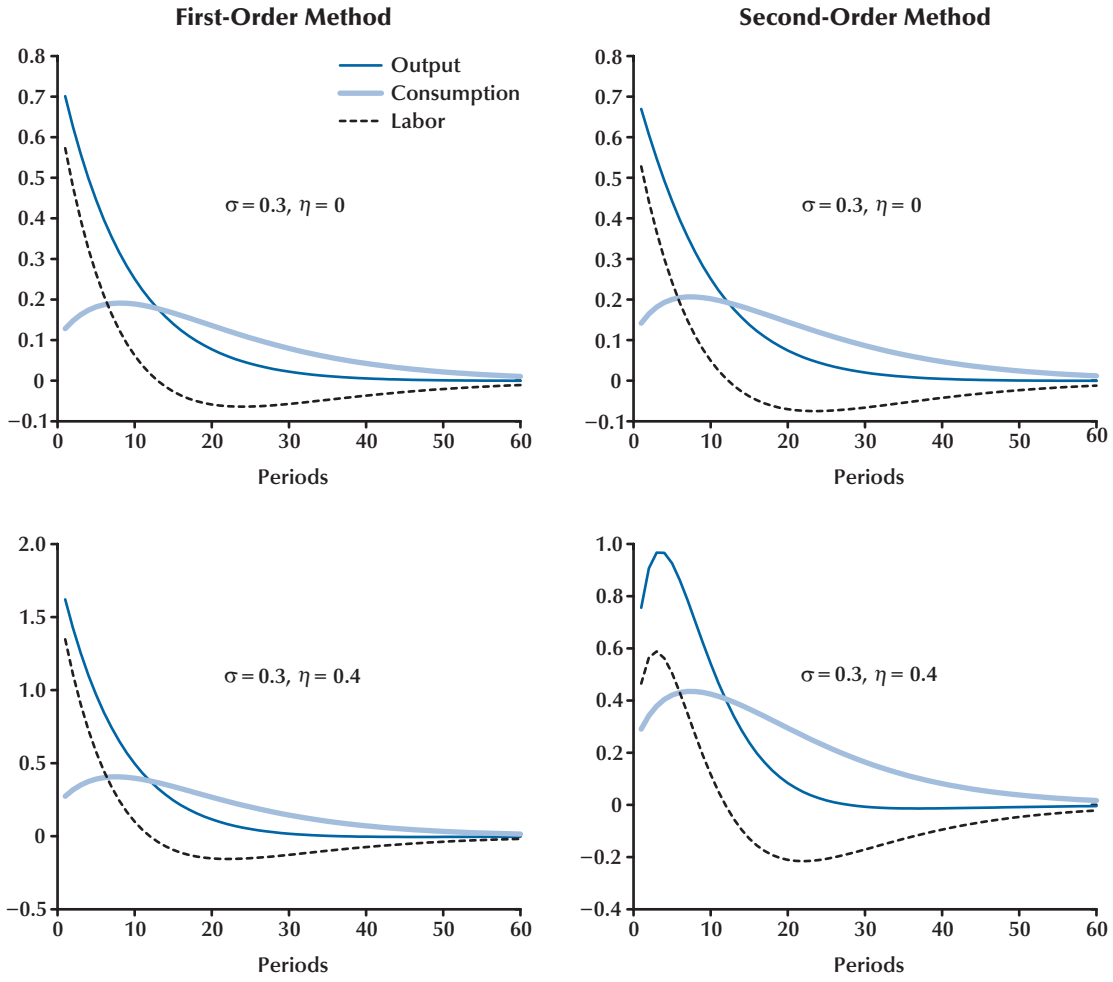
[13] If the variance of the shock σ is too small, the first-order and second-order methods yield very similar results. However, second-order methods may no longer be accurate if σ is too large. Hence, we choose a large-enough σ so that (i) the second-order terms are significant and (ii) the second-order solution is more accurate than linear methods. If we use Wen's (1998a) model with variable capacity utilization, the values of σ can be made much smaller: on the order of $\sigma = 0.03$.

[14] This value ($\eta = 0.04$) is also sufficiently below the threshold value for generating the n-period cycles found by Coury and Wen (2009).

[15] Coury and Wen (2009) show that the Benhabib-Farmer (1994) model has global indeterminacy even when the model's steady state appears to be a saddle judged by eigenvalues (i.e., for externalities below the critical level required for local indeterminacy based on the first-order approximation method).

Figure 2

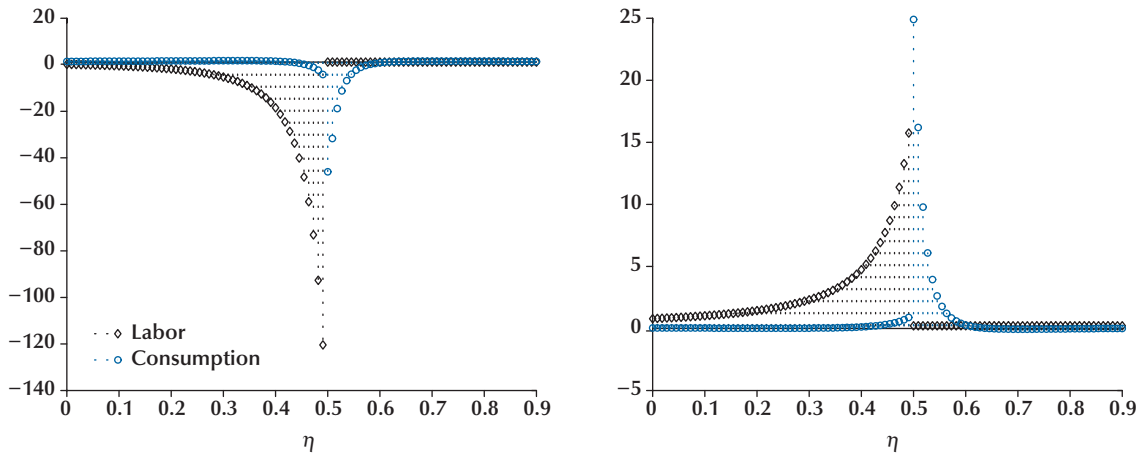
Impulse Responses of the Economy to a Positive Technology Shock



NOTE: SD, $\sigma = 0.3$. See text for details.

Figure 3

Second-Order Effects as η Increases



terms under indeterminacy because hours worked are a log-linear function of consumption and capital, both of which are state variables under indeterminacy.

Hence, the second-order method predicts more subdued and hump-shaped output responses because externalities introduce a large negative coefficient in front of the squared term for the technology shock, $\hat{A}_t^2$, and a large positive term in front of the cross term for the capital stock and technology, $\hat{k}_t\hat{A}_t$. This means that the effect of a positive technology shock on labor and output is neutralized by the squared term in the impact period but enhanced by the cross term in the subsequent periods, giving rise to the hump-shaped dynamic pattern.

Such nonlinear second-order income effects on hours worked also imply that the economy's dynamic responses to technology shocks are not symmetric. Under a negative technology shock, the squared term $\hat{A}_t^2$ reinforces the negative shock on impact, while the cross term $\hat{k}_t\hat{A}_t$ offsets the effect of the negative shock in subsequent periods, thus making the impulse response of output to a negative technology shock more contractionary but less persistent than that predicted by the linear method (Figure 4). For example, Figure 4 shows that the initial drop in output is about –1.6 by linear approximation and about –2.5 by second-order approximation. The predicted half-life is about seven quarters under the linear method and less than four quarters under the second-order method.

Such an asymmetric property is similar to what is observed in the U.S. economy. Expansions tend to be more gradual and long-lived, while contractions tend to be sharper and short-lived (see, e.g., Neftçi, 1984, and Sichel, 1993). Also, Mork (1989) and Hamilton (2003) point out that the economy's response to oil price shocks is highly asymmetric: A sudden increase in oil prices (an adverse productivity shock) tends to depress the economy, while a sudden decrease in oil prices tends to have little effect. Aguiar-Conraria and Wen (2007) argue that IRS large enough to trigger indeterminacy can explain the large negative impact of oil price increases on the U.S. economy. Here we show that smaller externalities can produce a similar effect under second-order approximation.

Given the same level of σ, as η increases, the impulse responses of output under second-order expansion become increasingly hump-shaped and the initial value can even become negative (which is then followed by a positive hump). This dynamic pattern offers one explanation of a

Figure 4

Responses to a Negative Technology Shock

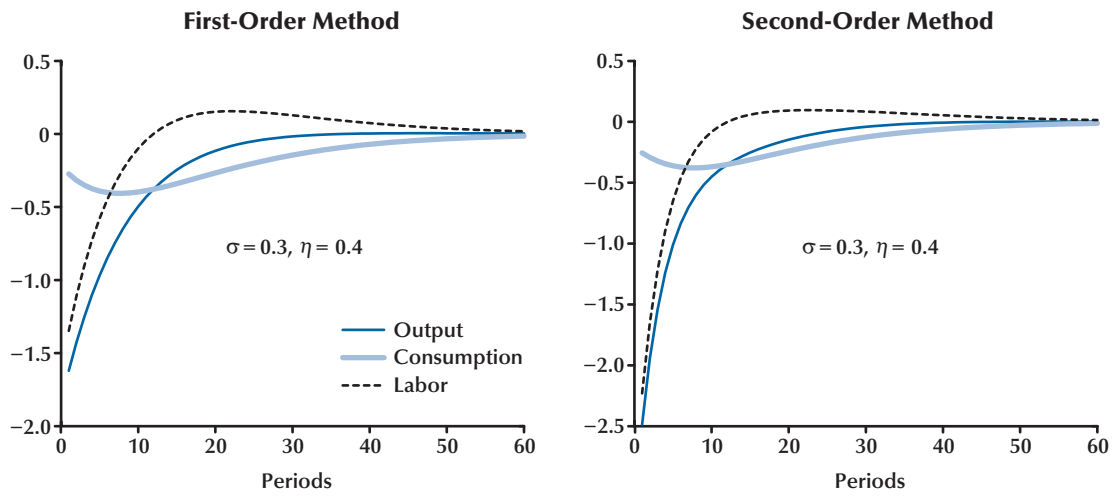
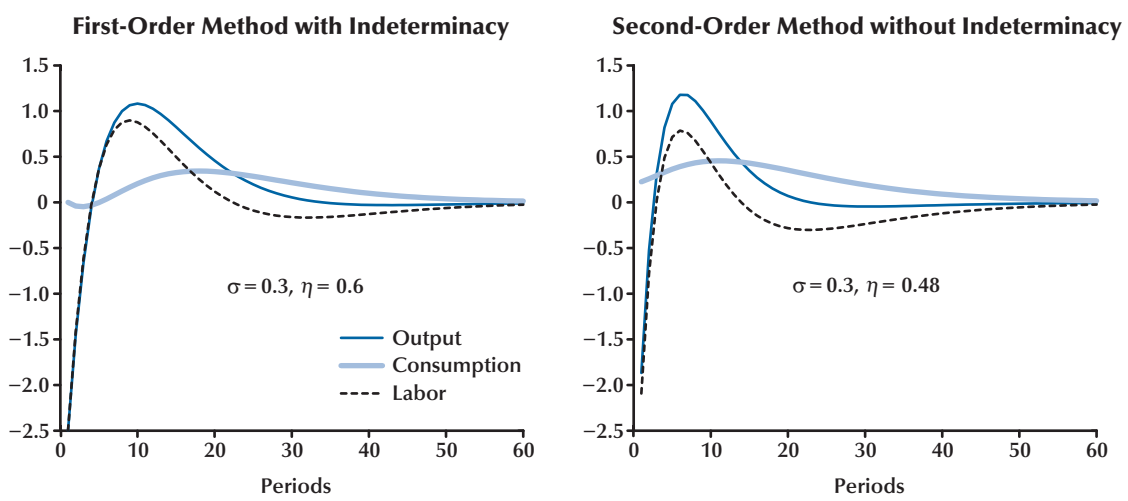


Figure 5

Responses to a Positive Technology Shock



puzzle in the indeterminacy literature: Technology shocks suddenly become contractionary once the model becomes locally indeterminate.

The left panel in Figure 5 shows the first-order approximation of the model under a positive technology shock when the steady state is indeterminate (i.e., $\eta = 0.6$),[16] and the right panel shows the second-order approximation of the model under a positive technology shock when $\eta = 0.48$, which is below the critical value for indeterminacy. The left panel shows that technology shocks are contractionary under indeterminacy, which is puzzling because the left panels in Figure 2 indicate that externalities amplify technology shocks in the positive direction. The existing indeterminacy literature has not provided an explanation for this puzzling phenomenon.

However, the right panel in Figure 5 suggests that technology shocks are contractionary from a second-order viewpoint even before the model becomes indeterminate. Therefore, we believe that the puzzle is caused by a large income effect from IRS on hours worked. When the marginal product of labor is high, it is optimal to reduce hours worked and increase leisure under the income effect but increase hours worked under the substitution effect. However, when externalities are below the critical value η^*, the income effect is captured only by second-order terms, while the substitution effect is captured by first-order terms. Therefore, the first-order method will show positive impulse responses under technology shocks. Once the model becomes indeterminate, if consumption is treated as a state variable under indeterminacy, the decision rule of labor has only first-order terms because the optimal first-order condition of labor supply (equation (1)) implies that it is a log-linear function of the model's state space. Hence, in this case there are no second- or higher-order terms in labor; consequently, the strong income effect of a technology shock on hours can be captured only by first-order terms, which explains the puzzle in the left panel of Figure 5.

[16] The initial consumption level is fixed at the steady state, $\hat{c}_{t=0}$. See Benhabib and Wen (2004) for discussions on how to generate impulse responses to fundamental shocks in an indeterminate model.

Investment Behavior

Based on linear methods, it has become well known in the RBC literature that investment is extremely volatile under technology shocks (see, e.g., Kydland and Prescott, 1982). However, this is not necessarily the case from a second-order perspective (Figure 6).

Because of the asymmetric second-order income effect, investment is very volatile only under a negative technology shock but not under a positive technology shock. In fact, if the elasticity of the labor supply is large enough or the technology shock is not highly persistent, the initial response of savings (investment) to a positive technology shock can be very mild or even negative, as shown in Figure 6. Notice that in each panel of Figure 6, second-order approximation always yields a lower level of investment than first-order approximation. This result is due to an extremely large negative coefficient of the squared term $\hat{A}_t^2$ in the investment decision rule,

$$\left[\hat{\mathbf{i}}_t\right] = \left(-2.5029 \;\; 21.9379\right)\left[\hat{\mathbf{k}}_t \;\; \hat{\mathbf{A}}_t\right]' + \frac{1}{2}\left(-10.1104 \;\; 139.2464 \;\; -439.1138\right)\left[\hat{\mathbf{k}}_t^2 \;\; \hat{\mathbf{k}}_t\hat{\mathbf{A}}_t \;\; \hat{\mathbf{A}}_t^2\right]' + \frac{96.9648}{2}\sigma^2.$$

By the same token, however, investment will always appear to be more volatile than that predicted by the linear method under a negative technology shock.

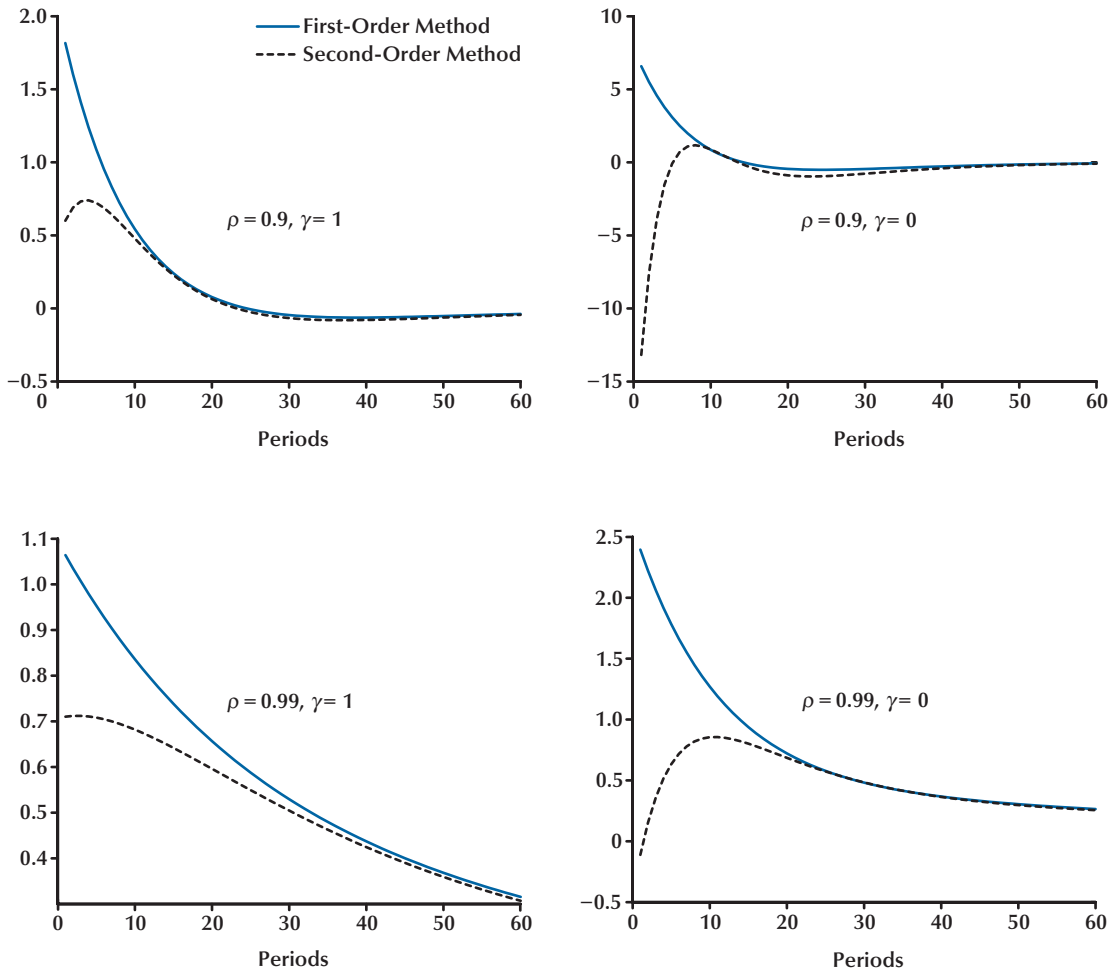
Second-Order Effects of Sunspot Shocks

Since the linear solution is used to obtain the coefficients of second-order terms (see Schmitt-Grohé and Uribe, 2004, and Lombardo and Sutherland, 2007), the zone of indeterminacy remains unchanged if indeterminacy arises in the model.[17] Farmer and Guo (1994) simulate the Benhabib-Farmer (1994) model under indetermi-

[17] When recursive equilibria fail to exist, the perturbation methods used in the literature and this paper cannot be applied (see Peralta-Alva and Santos, 2010, for the issues involved). Here we follow Benhabib and Farmer (1994) and the existing literature by taking as given the existence of recursive equilibria in the region of indeterminacy.

Figure 6

Responses of Investment to a Positive Shock

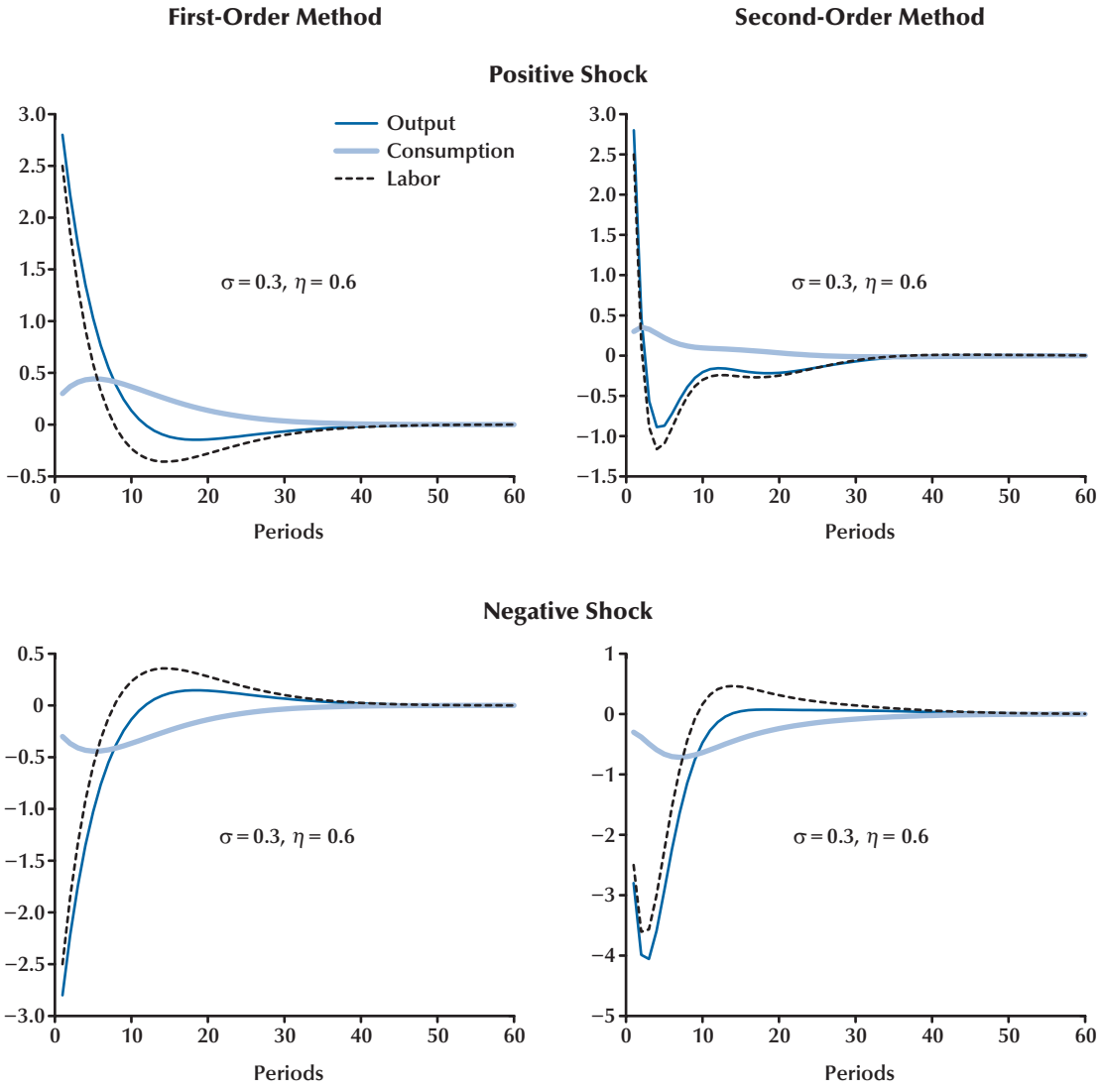


nacy by the linear method and show that sunspot shocks can generate business cycle comovements among output, consumption, investment, and hours. The effects of positive and negative sunspot shocks under the linear method are symmetric and monotonic (see the left panels in Figure 7, where the upper row pertains to a positive shock and the lower row pertains to a negative shock).

However, as shown in the right panels of Figure 7, under second-order approximation, positive and negative sunspot shocks have asymmetric effects on the economy. In particular, under a negative sunspot shock (the lower-right panel), the economy has hump-shaped impulse responses, while under a positive sunspot shock (the upper-right panel), the economy has no initial hump but tends to overshoot its steady state from above. The hump-shaped impulse responses of output to negative sunspot shocks are in contrast to the analysis of Schmitt-Grohé (2000), where the author argues (under the linear solution method) that sunspot shocks cannot generate hump-shaped output dynamics and forecastable comovements

Figure 7

Responses to Sunspot Shocks



among output, consumption, hours, and investment. Here we show that this may not necessarily be the case if the second-order effects are considered.

Notice in Figure 7 that both the first-order and second-order methods yield exactly the same initial magnitude of responses for all variables under a positive sunspot shock. The results differ only in the subsequent periods. This occurs because with indeterminacy, consumption is a state variable and the coefficients of second-order terms for labor are zero; thus, in the first period the responses of consumption and labor are due only to first-order effects while the capital stock stays unchanged. Consequently, a sunspot shock has only a first-order effect on the responses of output in the impact period. For example, the decision rules under sunspot shocks take the following form when $\sigma_s = 0.3$ and $\eta = 0.6$:

$$\begin{bmatrix} \hat{\mathbf{k}}_{t+1} \\ \hat{\mathbf{c}}_{t+1} \\ \hat{\mathbf{n}}_t \end{bmatrix} = \begin{pmatrix} 0.5070 & 1.0000 \\ -0.1304 & 1.2229 \\ -4.0000 & 8.3333 \end{pmatrix} \begin{bmatrix} \hat{\mathbf{k}}_t \\ \hat{\mathbf{c}}_t \end{bmatrix} +$$

$$(25) \quad \frac{1}{2}\begin{pmatrix} 2.5900 & -9.7502 & 9.1002 \\ 0.0327 & 0.1184 & -0.2984 \\ 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} \hat{\mathbf{k}}_t^2 \\ \hat{\mathbf{k}}_t \hat{\mathbf{c}}_t \\ \hat{\mathbf{c}}_t^2 \end{bmatrix}$$

$$+ \frac{1}{2}\begin{pmatrix} 0 \\ 8.5717 \\ 0 \end{pmatrix} \sigma_s^2 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \sigma_s^2 \varepsilon_{st+1},$$

where $\hat{A}_t = 0, \sigma_s^2$ is the variance of $\varepsilon_{st} \equiv \hat{c}_t - E_{t-1}\hat{c}_t$, which is the one-period-ahead forecasting error of consumption (sunspots). Clearly, starting from the steady state where $\hat{k}_t = \hat{k}_{t-1} = \hat{c}_{t-1} = 0$, the effects of a sunspot shock on consumption in period t are determined only by the forecasting error ε_{st} and not by any higher-order terms in the state space. In contrast, the coefficients of the second-order terms are zero in the decision rule for hours worked. In addition, output is a linear function of capital and labor. Therefore, both methods yield the same initial impulse responses for consumption, hours, and output, but in the subsequent periods the responses diverge significantly. In particular, because the cross term $\hat{k}_t\hat{c}_t$ has a large negative effect and the squared term $\hat{c}_t^2$ has a large positive effect on the capital stock, the impulse responses of output and labor become U-shaped under a negative sunspot shock. However, under a positive sunspot shock, they decrease monotonically and overshoot the steady state from above.

ACCURACY TESTING

As noted by Jin and Judd (2002), second-order approximation is not necessarily better than linear methods in terms of solution accuracy, depending on the models used and the parameter regions. Judd (1998) proposes using the Euler equation error (EEE) as a criterion for non-local accuracy testing. The EEE is expressed as the logarithm of the Euler equation residual

$$EEE = \log_{10}$$

$$(26) \quad \left| 1 - \frac{\left\{ \beta E_t \left[c_{t+1}^{-\tau} \left(\alpha A_{t+1} k_{t+1}^{\alpha(1+\eta)} n_{t+1}^{(1-\alpha)(1+\eta)} - 1 + \delta \right) \right] \right\}^{-\frac{1}{\tau}}}{c_t} \right|,$$

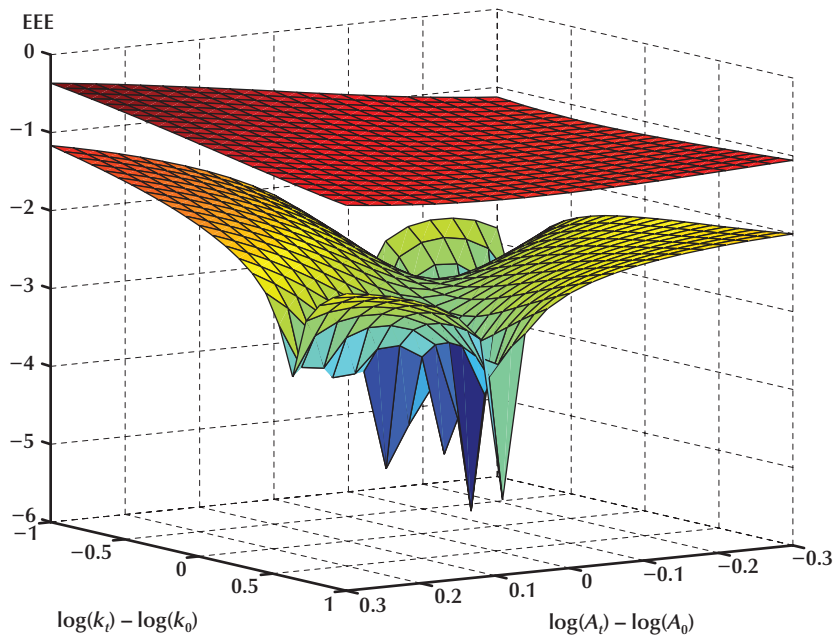
where $\{c_{t+1}, c_t, n_{t+1}, k_{t+1}\}$ are determined by the second-order policy rules discussed previously. Note that EEEs are generally negative, and a smaller (i.e., more negative) EEE implies improved accuracy. For example, the EEE is negative infinity ($-\infty$) when the solution is exact (i.e., 100 percent accurate).

Figure 8 plots the EEEs of the model under the first-order and second-order methods when $\sigma = 0.3$ and $\eta = 0.04$. The vertical axis represents the EEE, the right-front axis represents deviations of $\hat{A}_t$ from its steady state 0 in both positive and negative directions, and the left-front axis represents the deviations of capital from its steady state $\hat{k} = 0$. Since the first-order EEE lies everywhere above the second-order EEE both at the steady state and when the model is significantly away from the steady state, the second-order solution dominates the linear method in terms of accuracy under the current calibrations.

Alternatively, if we set the technology level to its steady state ($\hat{A} = 0$) or the capital stock to its steady state ($\hat{k} = 0$) and let the degree of exter-

Figure 8

Accuracy Test



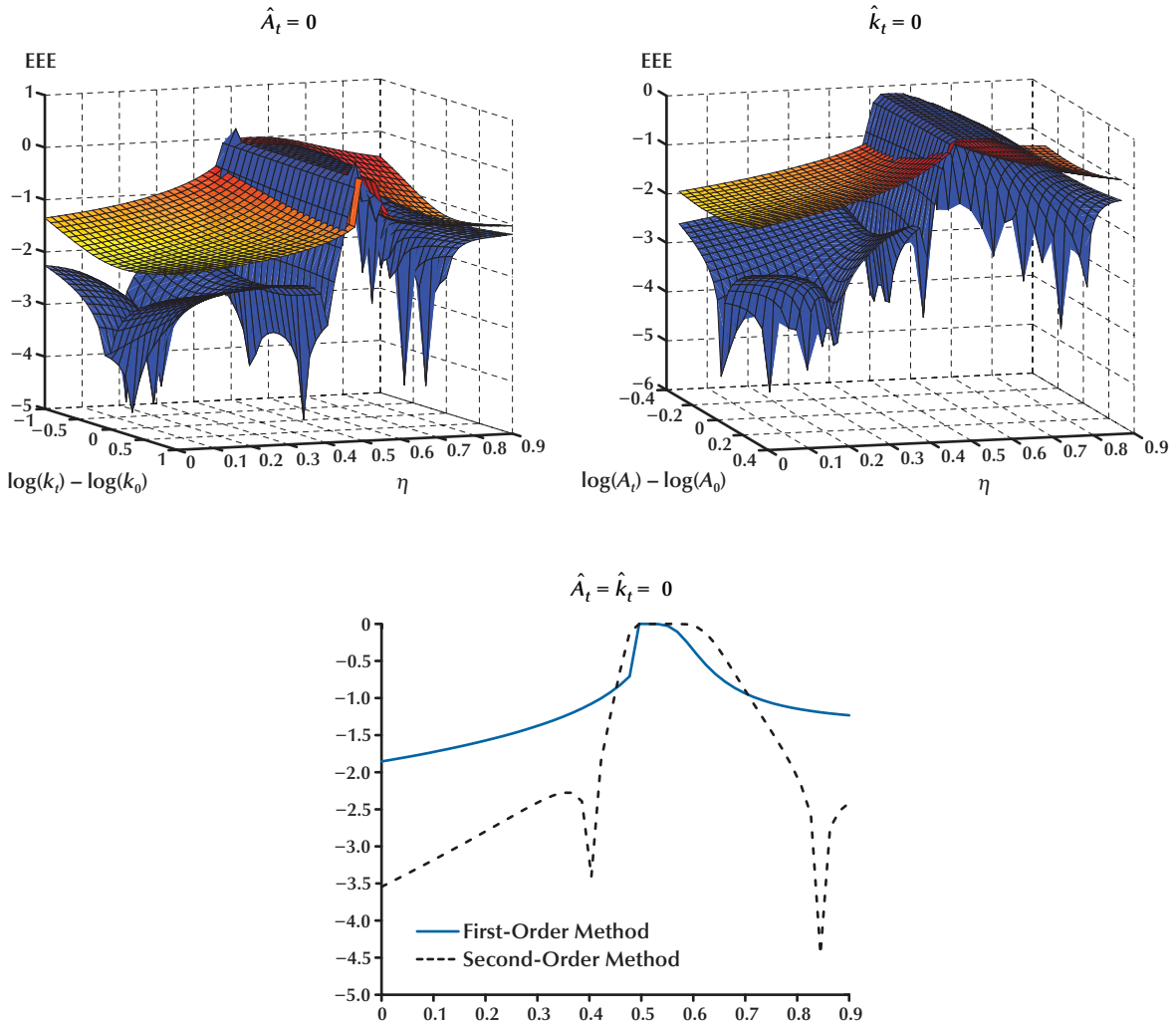
nalities vary, the top panels in Figure 9 show that the degree of accuracy deteriorates as η increases toward $\eta^* = 0.4935$. In particular, the second-order method is not necessarily better than the linear method around the critical point of indeterminacy, η^*. However, for most values of η, the second-order method dominates the first-order method. The bottom panel in Figure 9 is a two-dimensional graph of the top panels at the point where $\hat{k} = 0$ and $\hat{A} = 0$. It shows that the accuracy of both the first-order and the second-order approximation methods deteriorates as the externality η increases toward the critical point of indeterminacy, but the second-order method is more accurate than the linear method except around the point of indeterminacy.

CONCLUSION

This paper shows that externalities and increasing returns to scale can have important implications for business cycle dynamics that have not been fully appreciated by first-order approximation methods. In particular, externalities induce a strong second-order asymmetric income effect on leisure so that hump-shaped output dynamics may emerge even if externalities are sufficiently below the critical value required for local indeterminacy. Similarly, i.i.d. sunspot shocks under indeterminacy may generate forecastable comovements of output, hours, consumption, and investment, in contrast to the conclusion reached by Schmitt-Grohé (2000). Moreover, this asymmetric second-order income effect can generate business cycle dynamics that are qualitatively consistent with the asymmetric nature of the business cycle observed in the U.S. economy. Whether similar results can also be found in models with other types of frictions, such as borrowing constraints, remains to be investigated.

Figure 8

Accuracy with Respect to η



REFERENCES

Ackerberg, Daniel; Geweke, John and Hahn, Jinyong. "Comments on 'Convergence Properties of the Likelihood of Computed Dynamic Models.'" *Econometrica*, November 2009, *77*(6), pp. 2009-17.

Aguiar-Conraria, Luís and Wen, Yi. "Understanding the Large Negative Impact of Oil Shocks." *Journal of Money, Credit, and Banking*, June 2007, *39*(4), pp. 925-44.

Baxter, Marianne and King, Robert G. "Productive Externalities and Business Cycles." Discussion Paper No. 53, Federal Reserve Bank of Minneapolis, November 1991; www.minneapolisfed.org/research/DP/DP53.pdf.

Benhabib, Jess and Farmer, Roger. "Indeterminacy and Increasing Returns." *Journal of Economic Theory*, June 1994, *63*(1), pp. 19-41.

Benhabib, Jess and Wen, Yi. "Indeterminacy, Aggregate Demand, and the Real Business Cycle." *Journal of Monetary Economics*, April 2004, *51*(3), pp. 503-30.

Cogley, Tim and Nason, James. "Output Dynamics in Real-Business Cycle Models." *American Economic Review*, June 1995, *85*(3), pp. 492-511.

Collard, Fabrice and Juillard, Michel. "Accuracy of Stochastic Perturbation Methods: The Case of Asset Pricing Models." *Journal of Economic Dynamics and Control*, June 2001, *25*(6-7), pp. 979-99.

Cooper, Russell and Johri, Alok. "Dynamic Complementarities: A Quantitative Analysis." *Journal of Monetary Economics*, September 1997, *40*(1), pp. 97-119.

Coury, Tarek and Wen, Yi. "Global Indeterminacy in Locally Determinate Real Business Cycle Models." *International Journal of Economic Theory*, March 2009, *5*(1), pp. 49-60.

Farmer, Roger and Guo, Jang-Ting. "Real Business Cycles and the Animal Spirits Hypothesis." *Journal of Economic Theory*, June 2004, *63*(1), pp. 42-72.

Feng, Zhigang; Jianjun Miao, Jianjun; Peralta-Alva, Adrian and Santos, Manuel. "Numerical Simulation of Nonoptimal Dynamic Equilibrium Models." Working Paper No. 2009-018A, Federal Reserve Bank of St. Louis, April 2009; http://research.stlouisfed.org/wp/2009/2009-018.pdf.

Fernández-Villaverde, Jesús; Rubio-Ramírez, Juan F. and Santos, Manuel S. "Convergence Properties of the Likelihood of Computed Dynamic Models." *Econometrica*, January 2006, *74*(1), pp. 93-119.

Hamilton, James D. "What Is an Oil Shock?" *Journal of Econometrics*, April 2003, *113*(2), pp. 363-98.

Jin, He-hui and Judd, Kenneth L. "Perturbation Methods for General Dynamic Stochastic Models." Unpublished manuscript, Stanford University, April 2002.

Judd, Kenneth L. *Numerical Methods in Economics*. MIT Press, Cambridge, MA: MIT Press, 1998.

Kim, Change-Jin and Piger, Jeremy. "Common Stochastic Trends, Common Cycles, and Asymmetry in Economic Fluctuations." *Journal of Monetary Economics*, September 2002, *49*(6), pp. 1189-211.

Kim, Jinill; Kim, Sunghyun; Schaumburg, Ernst and Sims, Christopher A. "Calculating and Using Second-Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models." *Journal of Economic Dynamics and Control*, November 2008, *32*(11), pp. 3397-414.

King, Robert G.; Plosser, Charles I. and Rebelo, Sergio T. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics*, March-May 1988, *21*(2-3), pp. 195-232.

Kydland, Finn E. and Prescott, Edward C. "Time to Build and Aggregate Fluctuations." *Econometrica*, November 1982, *50*(6), pp. 1345-70.

Lombardo, Giovanni and Sutherland, Alan. "Computing Second-Order-Accurate Solutions for Rational Expectation Models: Using Linear Solution Methods." *Journal of Economic Dynamics and Control*, February 2007, *31*(2), pp. 515-30.

McKay, Alisdair and Reis, Ricardo. "The Brevity and Violence of Contractions and Expansions." *Journal of Monetary Economics*, May 2008, *55*(4), pp. 738-51.

Mork, Knut A. "Oil and the Macroeconomy When Prices Go Up and Down: An Extension of Hamilton's Results." *Journal of Political Economy*, June 1989, *91*(3), pp. 740-44.

Morley, James and Piger, Jeremy. "The Asymmetric Business Cycle." *Review of Economics and Statistics*, forthcoming.

Neftçi, Salih N. "Are Economic Time Series Asymmetric over the Business Cycle?" *Journal of Political Economy*, April 1984, *92*(2), pp. 307-28.

Peralta-Alva, Arian and Santos, Manuel S. "Problems in the Numerical Simulation of Models with Heterogeneous Agents and Economic Distortions." *Journal of the European Economic Association*, April/May 2010, *8*(2-3), pp. 617-25.

Romer, Paul M. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, October 1986, *94*(5), pp. 1002-37.

Schmitt-Grohé, Stephanie. "Endogenous Business Cycles and the Dynamics of Output, Hours, and Consumption." *American Economic Review*, December 2000, *90*(5), pp. 1136-59.

Schmitt-Grohé, Stephanie and Uribe, Martín. "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function." *Journal of Economic Dynamics and Control*, December 2004, *28*(4), pp. 755-75.

Sichel, Daniel E. "Business Cycle Asymmetry: A Deeper Look." *Economic Inquiry*, April 1993, *31*(2), pp. 224-36.

Swanson, Eric T.; Anderson, Gary and Levin, Andrew T. "Higher-Order Perturbation Solutions to Dynamic, Discrete-Time Rational Expectations Models." Working Paper No. 2006-01, Federal Reserve Bank of San Francisco, January 2006; www.frbsf.org/publications/economics/papers/2006/wp06-01bk.pdf.

Wen, Yi. "Capacity Utilization under Increasing Returns to Scale." *Journal of Economic Theory*, July 1998a, *81*(1), pp. 7-36.

Wen, Yi. "Can a Real Business Cycle Model Pass the Watson Test?" *Journal of Monetary Economics*, June 1998b, *42*(1), pp. 185-203.

Regional Aggregation in Forecasting: An Application to the Federal Reserve's Eighth District

Kristie M. Engemann, Rubén Hernández-Murillo, and Michael T. Owyang

Hernández-Murillo and Owyang (2006) showed that accounting for spatial correlations in regional data can improve forecasts of national employment. This paper considers whether the predictive advantage of disaggregate models remains when forecasting subnational data. The authors conduct horse races among several forecasting models in which the objective is to forecast regional- or state-level employment. For some models, the objective is to forecast using the sum of further disaggregated employment (i.e., forecasts of metropolitan statistical area [MSA]-level data are summed to yield state-level forecasts). The authors find that the spatial relationships between states have sufficient predictive content to overcome small increases in the number of estimated parameters when forecasting regional-level data; this is not always true when forecasting state- and regional-level data using the sum of MSA-level forecasts. (JEL C31, C53)

Federal Reserve Bank of St. Louis *Review*, May/June 2011, *93*(3), pp. 207-22.

Forecasting, especially as it pertains to policymaking, is typically conducted at the national level.[1] However, a few recent papers have indicated that aggregating regional forecasts may improve forecasts of national indicators. For example, Hendry and Hubrich (2006) use disaggregate models to form forecasts for aggregate variables. Similarly, Giacomini and Granger (2004) show that using a disaggregate model that accounts for spatial correlations can reduce the root mean squared error of the forecasts. Their disaggregate forecasts take advantage of cross-regional correlations yet still restrict the number of parameters estimated.[2] They argue that, under certain conditions, the sum of the forecasts from an order-p,q space-time autoregression [ST-AR(p,q)] can outperform both aggregate models and models that do not account for the spatial nature of the data. The ST-AR(p,q) model includes p temporal lags and q spatially distributed lags—that is, lags of the other regional series weighted by proximity. Thus, the ST-AR(p,q) model exploits both the

[1] There are, however, some notable exceptions of forecasting economic indicators at the subnational level (dates and regions noted in parentheses): Glickman (1971, Philadelphia MSA); Ballard and Glickman (1977, Delaware Valley); Crow (1973, Northeast Corridor); Baird (1983, Ohio); Liu and Stocks (1983, Youngstown-Warren MSA); Duobinis (1981, Chicago MSA); LeSage and Magura (1986, 1990, Ohio); and Rapach and Strauss (2005, Missouri; 2007, Eighth Federal Reserve District).

[2] Compared with a standard vector autogression (VAR), the space-time autoregression (AR) model posited in Giacomini and Granger (2004) requires the estimation of $(n^2 - n - 1)p$ fewer parameters for the same lag order p.

Kristie M. Engemann is a former senior research associate, Rubén Hernández-Murillo is a senior economist, and Michael T. Owyang is a research officer at the Federal Reserve Bank of St. Louis. This paper was prepared for the 4th Annual Business and Economics Research Group conference sponsored by the Federal Reserve Bank of St. Louis and the Center for Regional Economics—8th District on March 28, 2008, in Columbia, Missouri (as part of the Missouri Economics Conference). It appeared in Federal Reserve Bank of St. Louis *Regional Economic Development*, October 2008, *4*(1), pp. 15-29. The authors thank Dave Rapach for comments.

© 2011, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

spatial correlations and the information content in the disaggregated series.

Hernández-Murillo and Owyang (2006) take this approach to national employment data, showing that out-of-sample forecasts can be improved by modeling the spatial interactions between Bureau of Economic Analysis regions. They compare a ST-AR(p,q) model with vector autoregressions (VARs) with various levels of disaggregation. They concluded that, as predicted by Giacomini and Granger (2004), information in regional employment data is useful for forecasting national employment.

In this paper, we are interested in whether the information content of regional data can be observed at a more disaggregated level. In particular, we ask whether information for states helps forecast regional data and whether information from cities helps forecast state data. To this end, we construct horse races among four competing models with different levels of disaggregation. We then conduct out-of-sample tests to determine which model produces the best short- and long-horizon forecasts. The data used in these experiments are state- and metropolitan statistical area (MSA)-level payroll employment. In each experiment, the disaggregate data are summed to yield either state- or regional-level aggregates. In each case, we ask whether models using the disaggregate data provide lower mean squared prediction errors (MSPEs) than the aggregate alternatives. We find that the spatial relationships among states have sufficient predictive content to overcome small increases in the number of estimated parameters. The same is not always true when forecasting state- and regional-level variables using the sum of MSA-level forecasts.

The next section reviews the four models used in the horse races, followed by a section that discusses the subnational data and the construction of the "aggregate" data. The results of the out-of-sample experiments are then presented, followed by the conclusion.

MODELS

The goal of this experiment is to produce an h-period-ahead forecast of an aggregate time series—for example, employment. In this context, "aggregate" does not necessarily mean "national," although it is an obvious interpretation. Instead, here aggregate time series are data that are the sum or weighted sum of a number of (forecastable) disaggregate series. These series can be disaggregated in any manner (e.g., by regions or industries). The aggregate forecast then can be constructed directly from aggregate data or from the sum (or weighted sum) of its components. We examine four alternatives.

Suppose that period-t aggregate employment is denoted Y_t and can be written as the sum of its N disaggregate counterparts (henceforth referred to as "regions," which depending on the application may refer to either states or metro areas), y_{nt}, without error.[3] Let $\hat{Y}_{t+h}$ be the h-period-ahead forecast of Y. A forecast from the simplest model, a univariate aggregate order-p autoregression (AR(p), Model 1), has the form

$$(1)\quad \hat{Y}_{t+h} = \sum_{j=1}^{p} \Phi_j Y_{t+h-j},$$

where p is the number of lags and Φ_j are scalar coefficients.[4]

A similar univariate model can be constructed to forecast each of the individual components—in particular, region n's h-period-ahead level of employment, $\hat{y}_{n,t+h}$.[5] The aggregate forecast is the sum of the N regional forecasts (Model 2):

$$(2)\quad \hat{Y}_{t+h} = \sum_{n=1}^{N} \hat{y}_{n,t+h}^{uni} = \sum_{n=1}^{N}\sum_{j=1}^{p} \phi_{nj} y_{n,t+h-j},$$

where $\hat{y}_{n,t+h}^{uni}$ is region n's employment forecast from the univariate AR(p) model and ϕ_{nj} are scalar coefficients.

---

[3] The implicit assumption made here is that the aggregate is exactly the sum of its component parts. That is,

$$Y_t = \textstyle\sum_{n=1}^{N} y_{nt}$$

holds identically. Of course, the validity of this assumption depends greatly on the choice of data.

[4] Potential constants and time trends are suppressed in this section for notational convenience.

[5] Henceforth, we refer to the disaggregate components as "regions," although they can, in principle, be of any type (e.g., industry, state, MSA).

An alternative to Model (2) that accounts for the comovement between the regions is a VAR forecast (Model 3). The aggregate forecast obtained from such a model can be written as

$$(3)\quad \hat{Y}_{t+h} = \sum_{n=1}^{N} \hat{y}_{n,t+h}^{var} = \sum_{n=1}^{N}\sum_{k=1}^{N}\sum_{j=1}^{p} \Gamma_{nkj} y_{k,t+h-j},$$

where $\hat{y}_{n,t+h}^{var}$ is region n's employment forecast and Γ_{nkj} is the (scalar) lag-j effect of region k on region n's employment taken from the VAR coefficient matrices.

Finally, we consider a ST-AR(p,q) model (Model 4), which accounts explicitly for the spatial correlations among regions by imposing a relationship that depends on the proximity to a region's neighbors. The spatial weights w_{nk} are chosen a priori and are intended to reflect proximity between pairs of regions, for example, in terms of geographic characteristics such as contiguity or distance. Interaction between regions is governed by a weighting matrix $W = \{w_{nk}\}$ satisfying

$$w_{nk} \geq 0, \quad w_{nn} = 0, \quad \text{and} \quad \sum_{k \neq n} w_{nk} = 1$$

$$(4)\quad \begin{aligned} \hat{Y}_{t+h} &= \sum_{n=1}^{N} \hat{y}_{n,t+h}^{star} \\ &= \sum_{n=1}^{N}\left[\sum_{j=1}^{p} \phi_j y_{n,t+h-j} + \sum_{k=1}^{N}\sum_{l=1}^{q} \psi_l w_{nk} y_{k,t+h-l}\right], \end{aligned}$$

where ϕ_j and ψ_l are scalar autoregressive and scalar spatial lag coefficients, respectively. The weighting matrices used in the empirical applications are discussed below.

The primary differences among the four models involve a tension between modeling the (in-sample) cross-spatial correlations and parameter proliferation. Clearly, Models (1) and (2) are the most parsimonious models. However, these models neglect potentially predictive information in the comovement between the variables. On the other hand, the VAR depicted in Model (3) may overfit the in-sample data. Under parameter certainty, the VAR forecast in Model (3) weakly dominates the three alternative Models (1), (2), and (4). However, Giacomini and Granger (2004) show that forecasting from an estimated VAR (Model 3) is less efficient than forecasting from the ST-AR model (Model 4).[6] Because the ST-AR model is a restricted form of the VAR, the error associated with parameter uncertainty decreases. Giancomini and Granger, however, are unable to determine whether the ST-AR model or the univariate model is more theoretically efficient (i.e., whether interaction between regions yields significant information for forecasting). In the following section, we investigate whether accounting for spatial interaction in regional employment data is sufficiently elucidative to warrant the use of disaggregate data in forecasting.

EMPIRICAL DETAILS

Hernández-Murillo and Owyang (2006) tested the forecasting efficacy of the spatially disaggregated model for national employment. Here, we consider further disaggregation by examining the model's ability to forecast state- and Federal Reserve District-level employment. We conduct three experiments. First, we forecast Eighth District employment using the sum of state-level employment.[7] Second, we forecast District employment using the sum of Eighth District MSA-level employment.[8] Finally, we forecast state-level employment for each of the seven District states using MSA-level employment.

Data

Although a number of aggregate business cycle indicators exist, relatively few series are available at the disaggregate level. Two series available at a state level with both a reasonable

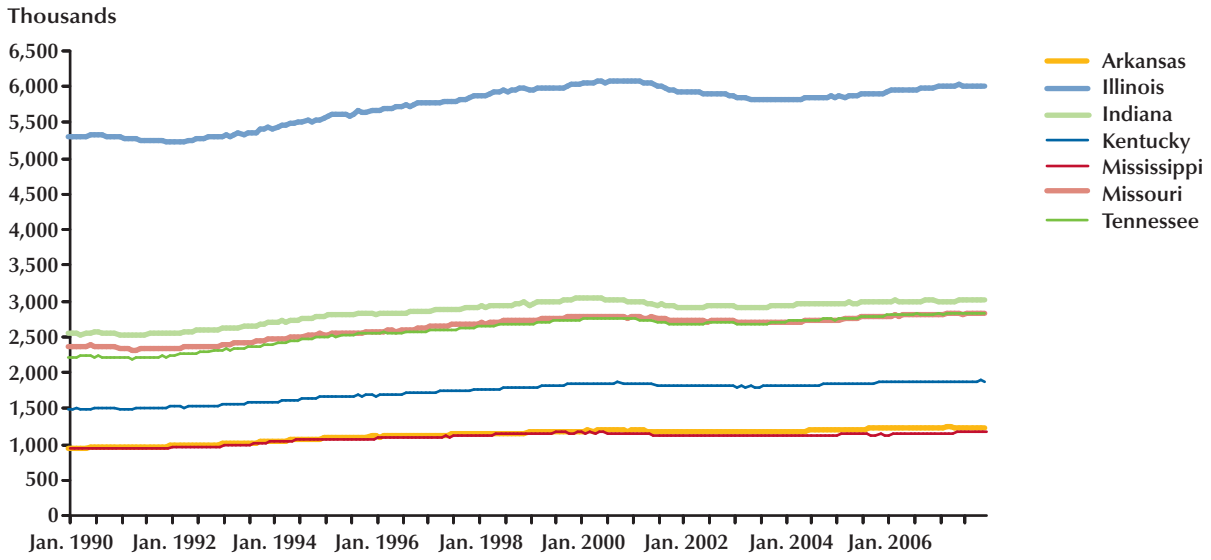
[6] Under certain conditions, the univariate aggregate model yields a lower mean squared error. For a discussion of these conditions, see Giacomini and Granger (2004).

[7] The Federal Reserve's Eighth District contains portions of seven states: Missouri, Illinois, Tennessee, Arkansas, Kentucky, Indiana, and Mississippi. Only Arkansas lies entirely in the Eighth District. However, for purposes of this experiment, we make the simplifying assumption that the District consists of the entirety of all seven states.

[8] In constructing District-level employment for this experiment and state-level employment for the next experiment, we use the sum of MSA-level employment. For the former, we include only MSAs located in the Eighth District, and for the latter, we include all MSAs in the states. Rural employment is omitted in each case.

Figure 1

Eighth District States' Payroll Employment by State



NOTE: The employment series for each state is seasonally adjusted.

frequency and sufficiently large sample are personal income (quarterly) and employment (monthly).[9] At an MSA-level, only employment is readily available. We, therefore, concentrate our efforts on the appropriate employment forecasts.

For our forecasting experiments, we use state- and MSA-level employment data from the Bureau of Labor Statistics' payroll employment survey. For the first experiment, state-level employment is summed to yield an approximation of the Eighth District employment level. In the same manner, the appropriate aggregates are constructed from MSA-level data in the following two experiments for forecasting District- and state-level data. For each exercise, the full sample is January 1990 to December 2007. For convenience, the state- and MSA-level data are plotted in Figures 1 and 2, respectively. Summary statistics for the data are provided in Tables 1 and 2.

[9] Gross state product, which is the state-level equivalent to national gross domestic product, is annual and only available at a one-year lag.

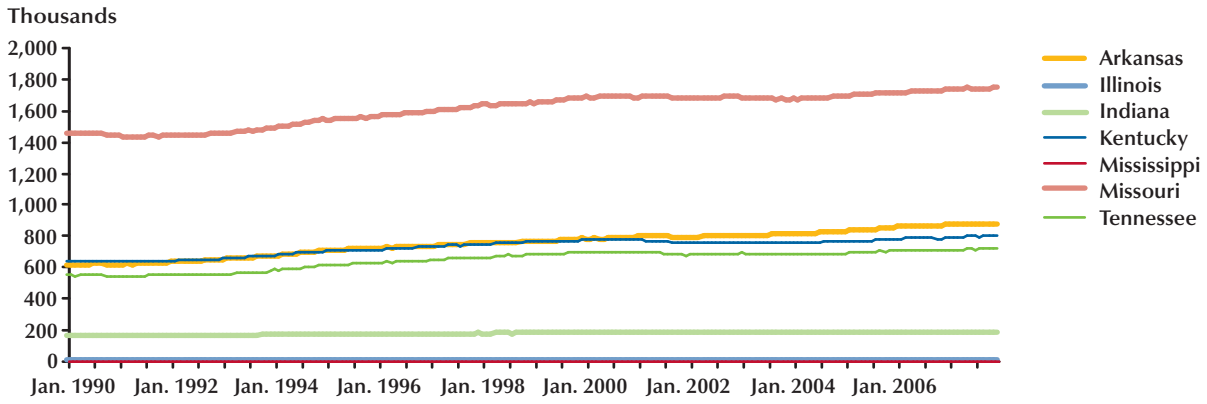
For each of the last two experiments, we construct the District- and state-level aggregates by omitting rural employment. Table 3 shows that the rural component of employment for each state in the Federal Reserve's Eighth District is significant. The difficulty, however, of adding rural employment to the forecasting regressions (at least those that account for cross-regional correlations) lies in modeling the comovements between rural and urban employment. In particular, for the spatial model (4), modeling the distance between the rural and MSA centroids is problematic.

Forecasting Scheme

We could use one of two forecasting schemes—recursive or rolling window. A recursive forecasting scheme fixes the initial period for the in-sample data. Each additional period is added to the sample and the model is reestimated. Thus, the estimation window expands as the sample expands. Conversely, the rolling window scheme fixes the size of the dataset used to make the fore-

Figure 2A

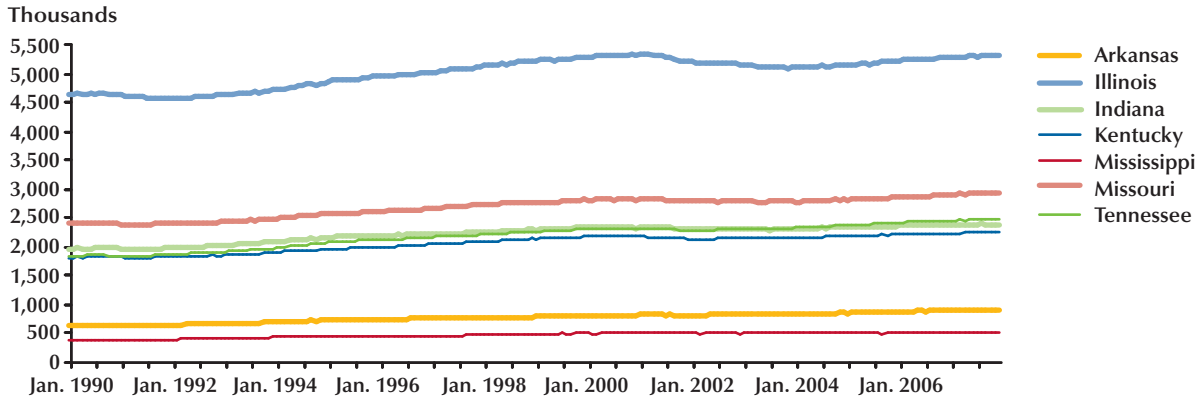
Eighth District MSAs' Payroll Employment by State



NOTE: The employment series for each state is seasonally adjusted and consists of the sum of all MSAs in that state.

Figure 2B

Total State MSAs' Payroll Employment by State



NOTE: The employment series for each state is seasonally adjusted and consists of the sum of all MSAs in that state.

Table 1

State-Level Summary Statistics

	Level (thousands)							Growth rate (percent)					
			Maximum		Minimum					Maximum		Minimum	
State	Mean	Variance	Level	Date	Level	Date	Skewness	Mean	Variance	Growth	Date	Growth	Date
Arkansas	1,095.8	7,670.6	1,209.4	Mar. 2007	912.2	Feb. 1990	–0.73	1.6	7.8	9.0	Jan. 1995	–5.2	Jan. 1991
Illinois	5,711.0	75,994.0	6,059.8	Jun. 2000	5,201.4	Mar. 1992	–0.64	0.8	6.8	11.7	Sep. 1995	–6.8	Jul. 2001
Indiana	2,824.3	26,891.1	3,015.2	May 2000	2,492.5	Mar. 1991	–0.85	1.0	12.0	10.6	Feb. 1999	–10.4	Jan. 1999
Kentucky	1,705.5	17,009.6	1,859.8	Nov. 2007	1,461.7	Apr. 1990	–0.64	1.4	16.4	29.6	May 1992	–20.1	Apr. 1992
Mississippi	1,085.0	5,442.7	1,171.2	Dec. 2007	928.8	Jan. 1990	–1.02	1.4	15.9	15.9	Oct. 1993	–18.6	Sep. 2005
Missouri	2,603.0	26,105.7	2,805.4	May 2007	2,294.9	Apr. 1991	–0.66	1.0	10.6	11.0	Apr. 1993	–14.6	Jan. 1991
Tennessee	2,560.9	39,700.8	2,816.8	Dec. 2007	2,171.0	Apr. 1991	–0.74	1.5	12.8	17.1	Nov. 1994	–13.2	Apr. 1996

NOTE: Monthly growth rates are annualized.

Table 2

MSA-Level Summary Statistics

	Level (thousands)							Growth rate (percent)					
			Maximum		Minimum					Maximum		Minimum	
MSA	Mean	Variance	Level	Date	Level	Date	Skewness	Mean	Variance	Growth	Date	Growth	Date
Fayetteville-Springdale-Rogers, AR*	156.2	993.8	208.6	Mar. 2007	105.2	Jan. 1990	0.09	4.0	25.1	38.0	Jan. 2001	–8.1	Apr. 2007
Fort Smith, AR-OK*	110.0	102.3	126.0	Aug. 2007	89.9	Jan. 1990	–0.53	2.1	37.0	22.9	Apr. 1994	–13.3	Nov. 2006
Hot Springs, AR*	34.0	11.4	39.6	Jul. 2007	27.0	Jan. 1990	–0.48	2.4	62.7	37.1	May 2003	–26.2	Feb. 2006
Jonesboro, AR*	44.3	17.4	49.8	Dec. 2007	36.6	Jan. 1990	–0.65	2.0	58.2	34.0	Jul. 1993	–15.8	Jan. 2000
Little Rock-North Little Rock, AR*	307.2	698.0	348.4	Dec. 2007	255.1	Feb. 1990	–0.53	1.8	13.8	24.5	Jan. 2001	–5.5	Sep. 2001
Pine Bluff, AR*	39.3	1.1	41.0	Nov. 2003	36.5	Feb. 1991	–0.70	0.7	75.3	35.3	Apr. 1991	–37.0	Jul. 2006
Texarkana, TX-AR*	51.9	7.2	56.6	Oct. 2007	47.0	May 1991	–0.15	1.0	34.9	22.3	Jan. 1996	–16.2	Sep. 1993
Bloomington-Normal, IL	81.9	87.6	92.6	Feb. 2002	65.3	Jan. 1990	–0.43	2.5	120.5	48.4	Jun. 1992	–38.4	Jan. 1995
Champaign-Urbana, IL	107.1	30.9	114.9	Sep. 2007	98.0	Apr. 1990	–0.28	1.2	102.3	52.9	Sep. 2007	–23.4	Nov. 1999
Chicago-Naperville-Joliet, IL	3,707.6	27,386.5	3,925.5	Jan. 2001	3,393.5	Dec. 1991	–0.62	0.6	8.3	12.4	Apr. 1993	–8.2	Jul. 2001
Danville, IL	33.1	0.9	35.2	Nov. 1996	31.2	Dec. 2006	–0.31	0.3	137.9	48.0	Apr. 1991	–42.4	Jan. 1999
Davenport-Moline-Rock Island IA-IL	179.6	69.7	190.0	Nov. 1999	164.5	Apr. 1990	–0.45	0.9	28.0	23.9	Jul. 1998	–18.8	Jan. 1999
Decatur, IL	55.6	3.7	60.7	Mar. 2000	52.5	Apr. 1992	1.05	0.5	78.2	54.2	May 1992	–29.9	Nov. 1991
Kankakee-Bradley, IL	42.0	5.3	44.6	Jan. 2007	35.8	Jan. 1990	–1.10	1.6	85.1	31.5	Jan. 1998	–28.5	Apr. 1992
Lake County-Kenosha County, IL-WI	345.8	1,945.4	401.5	Jun. 2007	265.1	Jan. 1990	–0.38	2.4	20.8	26.4	Oct. 1993	–8.7	Jun. 1993
Peoria, IL	171.4	103.0	188.2	Jun. 2007	149.1	Apr. 1992	–0.40	1.4	81.4	69.9	May 1992	–38.4	Jul. 1994
Rockford, IL	153.3	80.6	166.1	Jul. 2000	133.2	Aug. 1991	–0.77	1.4	81.9	35.7	Feb. 1990	–28.4	Aug. 2000
Springfield, IL	111.6	5.3	117.3	Aug. 2000	106.1	Feb. 1990	0.03	0.6	66.7	44.3	Aug. 1993	–30.0	Sep. 1992

NOTE: *Indicates an MSA located in the Eighth District and used in the second experiment; monthly growth rates are annualized.

Table 2, cont'd

MSA-Level Summary Statistics

	Level (thousands)							Growth rate (percent)					
			Maximum		Minimum					Maximum		Minimum	
MSA	Mean	Variance	Level	Date	Level	Date	Skewness	Mean	Variance	Growth	Date	Growth	Date
Anderson, IN	47.4	6.3	51.1	Mar. 1990	41.4	Jun. 2007	–0.70	–0.7	70.7	30.0	Aug. 1994	–24.0	Jul. 1994
Bloomington, IN	75.7	38.5	85.2	Aug. 2006	62.3	Jul. 1990	–0.80	2.4	202.9	80.4	Jan. 1993	–39.9	Jan. 2003
Columbus, IN	41.5	9.7	46.0	Oct. 2007	34.2	Feb. 1990	–0.88	2.0	85.2	47.0	Aug. 1992	–32.3	Jul. 1992
Elkhart-Goshen, IN	117.9	104.8	133.7	Feb. 2006	95.1	Mar. 1991	–0.58	1.6	52.8	22.9	Apr. 1992	–20.9	Apr. 1995
Evansville, IN-KY*	170.2	110.5	182.0	Nov. 2002	150.2	Feb. 1990	–0.70	1.1	28.9	26.6	Jan. 2001	–18.0	Oct. 1994
Fort Wayne, IN	208.5	80.2	220.0	Dec. 1998	190.6	Mar. 1991	–0.74	0.7	27.3	21.5	Jul. 2005	–17.6	Jan. 1999
Gary, IN	273.5	66.0	285.6	Jun. 1998	257.9	Apr. 1990	–0.34	0.5	23.2	21.6	Jan. 2001	–17.5	Sep. 2007
Indianapolis, IN	802.3	6,428.6	921.8	Aug. 2007	666.4	Apr. 1990	–0.32	1.9	15.7	18.4	Oct. 1998	–13.1	Jan. 1999
Kokomo, IN	50.9	7.5	55.7	Dec. 1999	44.1	Jul. 2003	–0.01	2.7	857.6	312.8	Aug. 2003	–74.9	Jul. 2003
Lafayette, IN	87.8	40.3	97.0	Aug. 2007	74.8	May 1991	–0.59	1.7	104.8	37.1	Jan. 1995	–51.1	Jan. 2003
Michigan City-La Porte, IN	47.5	1.6	50.3	Apr. 2000	45.1	Sep. 1991	0.04	0.4	36.8	21.8	Sep. 1999	–22.1	Oct. 1997
Muncie, IN	57.0	6.4	62.9	Jul. 1995	52.0	Jul. 2005	0.03	0.6	129.1	68.3	Jun. 2003	–31.3	Jun. 1997
South Bend-Mishawaka, IN	140.9	62.7	150.9	May 2000	123.6	Apr. 1992	–1.02	0.9	35.9	23.8	Oct. 2004	–20.6	Apr. 1992
Terre Haute, IN	74.4	6.5	78.2	Mar. 2000	68.2	Jul. 1990	–0.85	0.6	43.8	24.5	Jan. 1993	–34.3	Jan. 2003
Bowling Green, KY*	51.2	39.8	62.8	Dec. 2007	39.4	Jan. 1990	–0.15	3.0	71.1	46.7	Sep. 1991	–32.4	Jan. 1992
Cincinnati-Middletown, OH-KY-IN	965.5	4,351.7	1,047.4	Oct. 2007	859.0	Jan. 1990	–0.47	1.2	9.4	15.3	Oct. 1993	–6.2	Jan. 2005
Elizabethtown, KY*	41.9	16.1	48.8	Nov. 2007	35.2	Jan. 1990	0.00	2.0	36.6	22.1	Jan. 2006	–15.6	Apr. 2000
Huntington-Ashland, WV-KY-OH	110.0	35.6	121.2	Sep. 2007	99.8	Sep. 1991	0.07	1.1	34.3	34.0	Feb. 1994	–17.6	Jan. 1994
Lexington-Fayette, KY	231.3	407.3	257.1	Dec. 2007	194.2	Apr. 1990	–0.62	1.7	22.9	20.5	Jul. 2000	–11.2	Jul. 1995
Louisville, KY-IN*	579.1	1,495.5	629.4	Jul. 2007	504.2	Apr. 1991	–0.70	1.3	17.3	16.2	Dec. 2006	–17.4	Jan. 1991
Owensboro, KY*	47.8	12.0	51.6	Dec. 2007	40.8	Jul. 1991	–0.92	1.5	38.6	24.5	Apr. 1999	–16.3	Feb. 2003

NOTE: *Indicates an MSA located in the Eighth District and used in the second experiment; monthly growth rates are annualized.

Table 2, cont'd

MSA-Level Summary Statistics

	Level (thousands)							Growth rate (percent)					
			Maximum		Minimum					Maximum		Minimum	
MSA	Mean	Variance	Level	Date	Level	Date	Skewness	Mean	Variance	Growth	Date	Growth	Date
Gulfport-Biloxi, MS	99.3	202.7	116.2	Jan. 2005	71.0	Jan. 1990	–0.75	3.7	176.1	75.1	Aug. 1992	–78.9	Sep. 2005
Hattiesburg, MS	51.7	26.4	62.1	Aug. 2007	43.7	May 1991	–0.01	2.0	35.9	52.5	Oct. 2005	–10.5	Sep. 2005
Jackson, MS	232.9	471.5	263.4	Oct. 2007	194.6	Jan. 1990	–0.40	1.7	10.6	18.0	Jul. 2003	–7.3	Aug. 1990
Pascagoula, MS	54.2	9.2	60.8	Jul. 1999	46.1	Jan. 1990	–0.03	2.9	392.5	191.7	Apr. 2007	–71.6	Sep. 2005
Columbia, MO*	77.9	99.2	93.3	Oct. 2007	61.2	Jun. 1990	–0.24	2.5	26.9	22.7	Oct. 1993	–13.2	Apr. 1990
Jefferson City, MO*	72.2	44.2	80.3	Dec. 2007	59.0	Apr. 1990	–0.72	1.9	32.0	22.1	Jul. 1994	–16.3	Apr. 1995
Joplin, MO	71.7	35.2	80.1	Dec. 2007	60.2	Feb. 1991	–0.70	1.7	27.7	20.0	Jul. 2004	–14.7	May 2002
Kansas City, MO-KS	932.0	3,513.1	1,024.2	Aug. 2007	824.1	Jun. 1991	–0.60	1.2	12.3	14.0	Jul. 2007	–14.8	Jul. 2002
St. Joseph, MO-KS	49.4	20.7	59.6	Aug. 2007	42.8	Jul. 1991	0.50	1.9	59.6	35.2	Oct. 1991	–19.8	Sep. 1996
St. Louis, MO-IL*	1,280.5	4,295.7	1,362.6	May 2007	1,165.0	Jun. 1991	–0.58	0.8	8.1	10.0	Jan. 1998	–7.5	Dec. 2000
Springfield, MO*	167.5	425.1	202.9	Dec. 2007	131.2	Jul. 1990	–0.31	2.5	17.0	16.8	Aug. 1990	–11.2	Jan. 2003
Chattanooga, TN-GA	226.9	193.9	248.9	Jul. 2007	200.6	May 1991	–0.38	1.2	24.7	19.7	Jul. 1999	–12.1	Jan. 1996
Clarksville, TN-KY	70.8	100.4	85.7	Oct. 2007	53.4	Dec. 1990	–0.26	2.7	45.4	25.7	Sep. 1993	–18.8	Jul. 1990
Cleveland, TN	38.7	10.5	42.9	Jan. 2007	32.2	Aug. 1991	–0.74	2.1	169.0	109.3	Jan. 1994	–42.9	Mar. 1993
Jackson, TN*	55.8	30.5	62.3	Jun. 2007	45.7	Jan. 1991	–0.59	1.9	57.8	31.6	Jul. 1999	–28.4	Aug. 1999
Johnson City, TN	73.2	22.7	81.7	Nov. 2007	64.2	Feb. 1990	–0.10	1.6	52.7	31.7	Sep. 2003	–28.0	Jul. 1997
Kingsport-Bristol-Bristol, TN-VA	119.1	21.5	125.7	Apr. 1997	107.4	Jan. 1990	–1.11	1.0	52.2	25.1	Jul. 1990	–24.0	Sep. 1990
Knoxville, TN	293.9	777.1	338.9	Sep. 2007	240.8	Apr. 1990	–0.23	2.0	18.2	14.4	Apr. 1993	–11.9	Jan. 2003
Memphis, TN-AR-MS*	581.0	2,773.8	647.1	Dec. 2007	487.5	Mar. 1991	–0.66	1.6	22.9	23.4	Jan. 1994	–12.2	Aug. 1993
Morristown, TN	46.9	18.1	51.9	May 2006	38.1	Jan. 1990	–0.75	2.1	87.2	46.4	Jul. 1993	–22.5	Oct. 2002
Nashville-Davidson-Murfreesboro, TN	651.1	5,679.1	765.6	Dec. 2007	522.9	Feb. 1991	–0.38	2.2	18.3	16.9	Nov. 1994	–7.4	Oct. 2001

NOTE: *Indicates an MSA located in the Eighth District and used in the second experiment; monthly growth rates are annualized.

Table 3

Rural Employment, by State in 2006

State	Rural employment (percent)
Arkansas	36.3
Illinois	11.6
Indiana	20.0
Kentucky	36.0
Mississippi	52.5
Missouri	23.9
Tennessee	22.1
Average	28.9

SOURCE: USDA, Economic Research Service, State Fact Sheets.

cast. With each new period, recent data are added and data at the beginning of the sample are dropped. The rolling window scheme is particularly useful for cases in which the data-generating process experiences structural breaks. This has been shown to be the case for both state- and MSA-level employment (see Owyang, Piger, and Wall, 2005, 2008, and Owyang et al., 2008). Therefore, we choose to use a rolling window forecasting scheme with a 13-year sampling period. The number of lags for each model is chosen using the Bayesian information criterion (BIC) on the initial subsample and remains fixed for the entire forecasting experiment.

Spatial Weighting

Two sets of weights are considered for the first forecasting experiment. The first set of weights takes into account the distance between the centroids of economic regions, and the second considers geographic contiguity as a categorical qualification. Under the first definition,

$$w_{nk} = \left(1/d_{nk}\right)\Big/\left(\sum_{k\neq n} 1/d_{nk}\right),$$

where d_{nk} is the distance between the geographic centroids of regions n and k. Under the second definition,

$$w_{nk} = \left(\eta_{nk}\right)\Big/\left(\sum_{k\neq n} \eta_{nk}\right),$$

where η_{nk} =1 if regions n and k are geographically adjacent, and η_{nk} = 0 otherwise. Both of the final two experiments use only the distance between centroids because contiguity cannot be established for most MSAs.

RESULTS

A few broadly consistent features are notable for the three forecasting experiments. In particular, for the District forecasts the aggregate AR exhibits greater MSPEs at every horizon than the ST-AR model. The difference in MSPEs for the ST-AR model and a more parsimoniously parameterized VAR is often small, especially for short horizons; and the disaggregate AR can provide some (small) forecasting advantages over the more heavily parameterized ST-AR model at short horizons but is inferior at long horizons.

Forecasting District Employment with State-Level Data

The first set of results considers forecasting Eighth Federal Reserve District employment using state-level data. As mentioned previously, state-level data support two possible spatial weighting matrices for the ST-AR model: distance and contiguity. We present results for both weighting matrices.

Figure 3 shows the relative decline in MSPEs for the ST-AR model using centroid distance as the spatial metric relative to each of the forecasting models. Obvious from these results is that weighting state-level interactions by distance provides some advantage to aggregate forecasting over weighting by contiguity. The advantage may result because a contiguity weighting scheme would suppress potentially important interactions between noncontinuous states.[10]

For both weighting schemes, the informational advantage in modeling the regional interactions

[10] As alluded to above, the weighting matrix in spatial econometrics is determined exogenously. Conley and Molinari (2007) propose a test of the spatial weighting matrix. However, their test is conducted in-sample and is a joint test of model and spatial weighting misspecification.

Figure 3

Efficiency Gain for ST-AR Model, Using Eighth District States

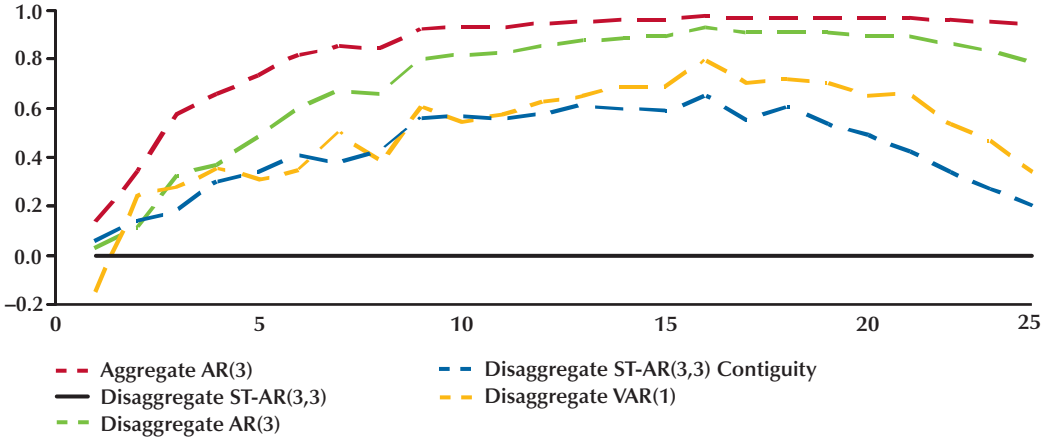


Figure 4

Efficiency Gain for ST-AR Model, Using Eighth District States (setting equal lag lengths)

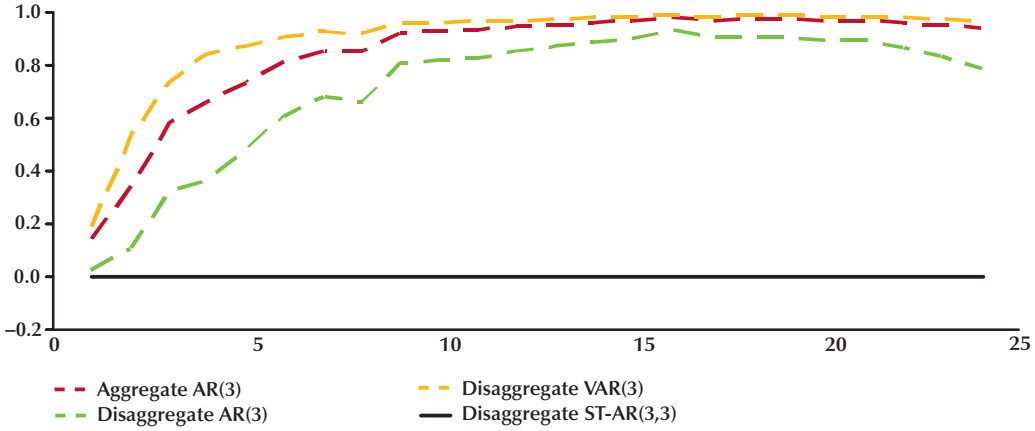
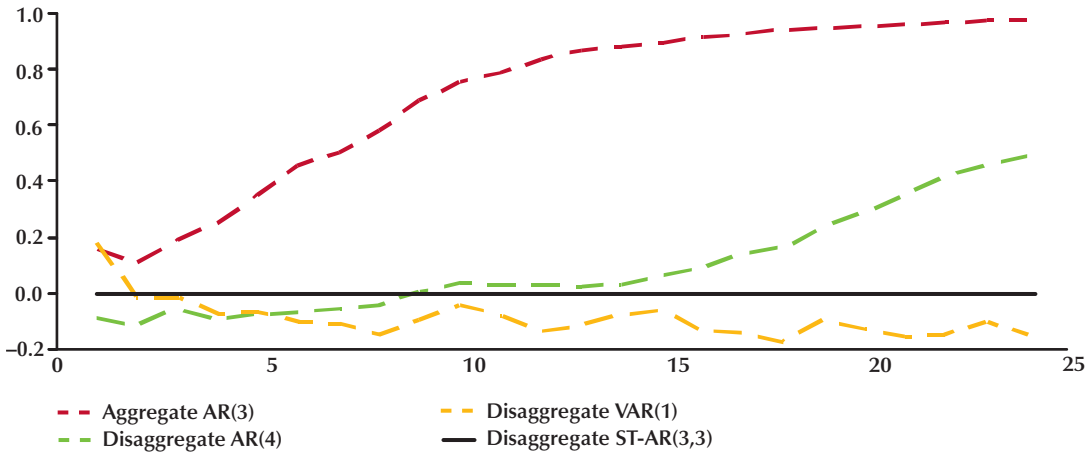


Figure 5

Efficiency Gain for ST-AR Model, Using Eighth District MSAs



is obvious. The VAR and the ST-AR models yield lower MSPEs for almost every horizon. At very short horizons, the disaggregate AR has predictive ability similar to that of the VAR and the ST-AR models. However, at longer horizons, neglecting the regional interactions can increase the MSPE by up to 90 percent.

The regional VAR and the ST-AR models produce an interesting comparison. First, it is important to note that the lag order chosen by the BIC for the VAR is much shorter than that for the ST-AR. This negates, to some extent, the reduction in the MSPEs gained by reducing parameter uncertainty in the more parsimoniously parameterized ST-AR model. Figure 4 demonstrates the informational advantage for a ST-AR model versus a VAR with equal lag length. This finding is consistent with the theoretical findings in Giacomini and Granger (2004): Increasing the number of estimated parameters in the VAR with equal lags leads to potential overfitting and an increase in the MSPEs.

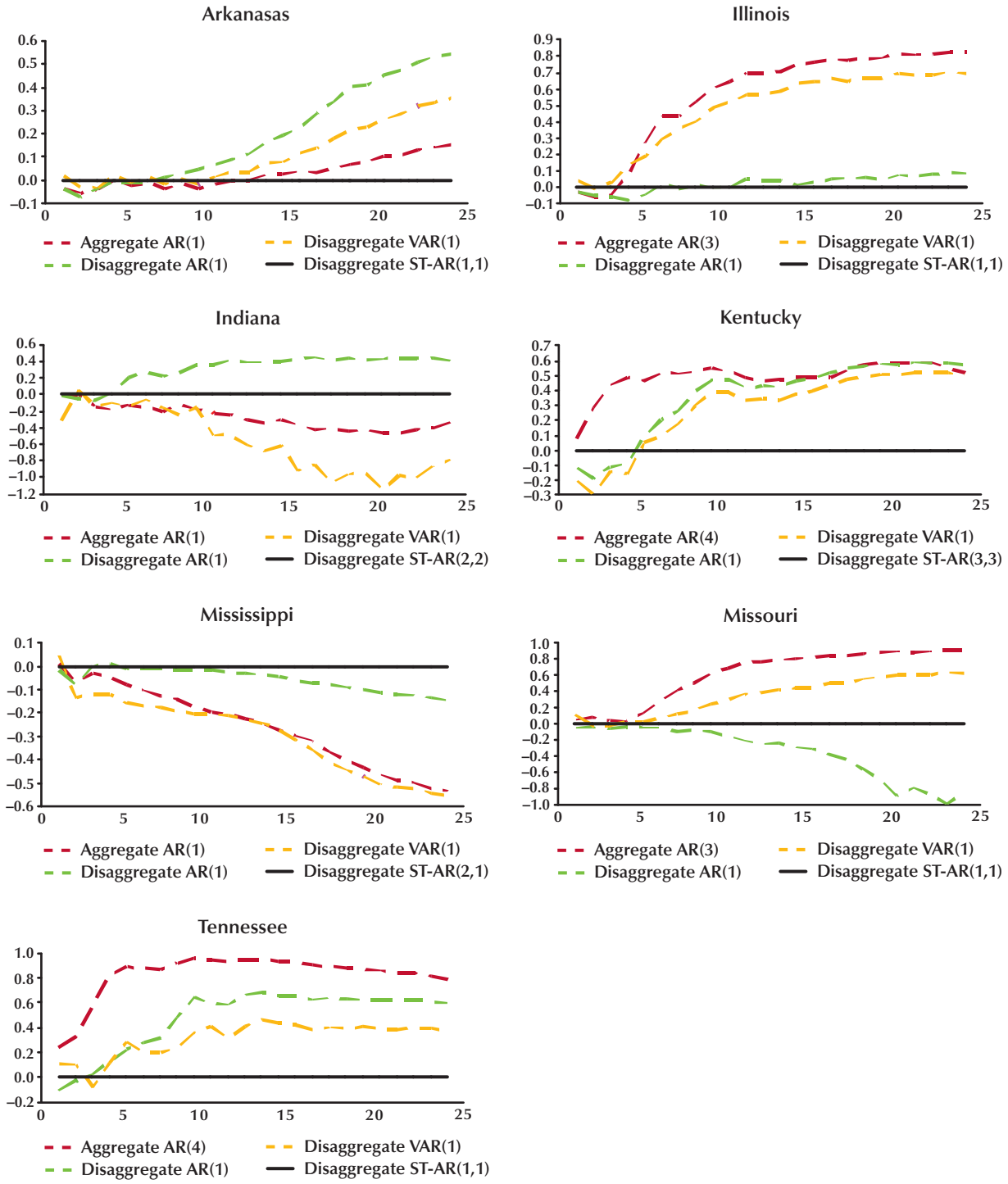
Forecasting District Employment with MSA-Level Data

As Figure 5 shows, the results for disaggregating at the MSA level are broadly consistent with those for the state data. The disaggregate models perform better out of sample than the aggregate AR model. The ST-AR model is more efficient than the disaggregate AR at long horizons. At shorter horizons, this information advantage is eroded and sometimes negative. Moreover, the VAR performs better in this case than the ST-AR model for most horizons.

These results suggest several possible explanations. In the previous case, District data were disaggregated into seven states; here, the District is disaggregated into 18 MSAs. Although the increase in the number of disaggregate units may not seem significant, it leads to a substantial increase in the number of estimated parameters for the ST-AR model. This increase may erode the model's forecasting advantage because of the increased uncertainty from estimating the extra parameters. Second, the MSA may be an improper level of disaggregation. A third possibility is that the spatial weighting matrix used in this exercise does not properly model the interactions. This could potentially explain why the VAR model performs better than the ST-AR model despite estimating a comparable number of parameters.

Figure 6

Efficiency Gain for ST-AR Model, Forecasting State Employment with MSAs



Forecasting State Employment with MSA-Level Data

We conducted similar experiments using the level of employment in the seven states in the Eighth District as the aggregate and the MSAs in those states as the disaggregate components. Our motivation is to determine the optimal level of disaggregation in forecasting employment. Unfortunately, few results are consistent across states (Figure 6). For example, most states yield lower MSPEs for the disaggregate forecasting models versus the aggregate AR model. Mississippi is an exception: The aggregate AR gives roughly similar MSPEs as the VAR and much lower MSPEs than either the ST-AR or disaggregate AR model. Overall, the model with the lowest MSPE for each state differs. The ST-AR model provides the lowest MSPE for about half of the states but performs considerably worse than even the aggregate AR for Mississippi and Indiana.

One notable fact in these results is that, for a given model, the lag order called for by the (in-sample) BIC varies substantially across the states. Not surprisingly, the ST-AR model tends to perform worse in states in which the in-sample criterion calls for longer lags. This can lead to an increase in parameter uncertainty or overfitting.[11] Similarly, for states in which long lags are called for in the AR model, this model performs poorly.

[11] The tension between in-sample and out-of-sample fit is not surprising (see Hansen, 2008).

We therefore conclude that, although some information may be gleaned from modeling spatial relationships, disaggregation to the MSA level should be done with some caution.

CONCLUSION

Recent studies have shown that, at times, aggregate variables can be more accurately forecasted by summing disaggregate forecasts. In particular, using models that take into account the spatial interactions of the disaggregate series can improve forecasting performance. This occurs at the expense of estimating additional parameters. This tension naturally leads to the question of how much disaggregation is "optimal."

We conducted a number of forecasting experiments along these lines. In general, we find that disaggregation can produce better forecasts. For example, by disaggregating a regional variable (the Eighth Federal Reserve District's employment level) into states, we achieved a significant reduction in the MSPE versus the aggregate AR. Using the state level as the aggregate, however, yields less consistent results, which suggests that the exploitable regional interactions at the MSA level may not be sufficiently informative to overcome the increase in estimated parameters. We imagine that further disaggregation—perhaps to the county level—might increase this tension between exploitable spatial interactions and increased parameter uncertainty.

REFERENCES

Baird, Catherine A. "A Multiregional Econometric Model of Ohio." *Journal of Regional Science*, November 1983, *23*(4), pp. 501-15.

Ballard, Kenneth and Glickman, Norman J. "A Multiregional Econometric Forecasting System: A Model for the Delaware Valley." *Journal of Regional Science*, August 1977, *17*(2), pp. 161-77.

Conley, Timothy G. and Molinari, Francesca. "Spatial Correlation Robust Inference with Errors in Location or Distance." *Journal of Econometrics*, September 2007, *140*(1), pp. 76-96.

Crow, Robert T. "A Nationally-Linked Regional Econometric Model." *Journal of Regional Science*, August 1973, *13*(2), pp. 187-204.

Duobinis, Stanley F. "An Econometric Model of the Chicago Standard Metropolitan Statistical Area." *Journal of Regional Science*, August 1981, *21*(3), pp. 293-319.

Giacomini, Raffaella and Granger, Clive W.J. "Aggregation of Space-Time Processes." *Journal of Econometrics*, January 2004, *118*(1/2), pp. 7-26.

Glickman, Norman J. "An Econometric Forecasting Model for the Philadelphia Region." *Journal of Regional Science*, April 1971, *11*(1), pp. 15-32.

Hansen, Peter R. "In-Sample Out-of-Sample Fit: Their Joint Distribution and Its Implications for Model Selection." Unpublished manuscript, 2008.

Hendry, David F. and Hubrich, Kirstin. "Forecasting Economic Aggregates by Disaggregates." Working Paper No. 589, European Central Bank, February 2006; www.ecb.eu/pub/pdf/scpwps/ecbwp589.pdf.

Hernández-Murillo, Rubén and Owyang, Michael T. "The Information Content of Regional Employment Data for Forecasting Aggregate Conditions." *Economics Letters*, March 2006, *90*(3), pp. 335-39.

LeSage, James P. and Magura, Michael. "Econometric Modeling of Interregional Labor Market Linkages." *Journal of Regional Science*, August 1986, *26*(3), pp. 567-77.

LeSage, James P. and Magura, Michael. "Using Bayesian Techniques for Data Pooling in Regional Payroll Forecasting." *Journal of Business and Economic Statistics*, January 1990, *8*(1), pp. 127-35.

Liu, Yih-wu and Stocks, Anthony H. "A Labor-Oriented Quarterly Econometric Forecasting Model of the Youngstown-Warren SMSA." *Regional Science and Urban Economics*, August 1983, *13*(3), pp. 317-40.

Owyang, Michael T.; Piger, Jeremy and Wall, Howard J. "Business Cycle Phases in U.S. States." *Review of Economics and Statistics*, November 2005, *87*(4), pp. 604-16.

Owyang, Michael T.; Piger, Jeremy and Wall, Howard J. "A State-Level Analysis of the Great Moderation." *Regional Science and Urban Economics*, November 2008, *38*(6), pp. 578-89.

Owyang, Michael T.; Piger, Jeremy M.; Wall, Howard J. and Wheeler, Christopher H. "The Economic Performance of Cities: A Markov-Switching Approach." *Journal of Urban Economics*, November 2008, *64*(3), pp. 538-50.

Rapach, David E. and Strauss, Jack K. "Forecasting Employment Growth in Missouri with Many Potentially Relevant Predictors: An Analysis of Forecast Combining Methods." Federal Reserve Bank of St. Louis *Regional Economic Development*, 2005, *1*(1), pp. 97-112; research.stlouisfed.org/publications/red/2005/01/RapachStrauss.pdf.

Rapach, David E. and Strauss, Jack K. "Forecasting Real Housing Price Growth in the Eighth District States." Federal Reserve Bank of St. Louis *Regional Economic Development*, November 2007, *3*(2), pp. 33-42; research.stlouisfed.org/publications/red/2007/02/ Rapach.pdf.

Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442