Federal Reserve Bank of St. Louis

REVIEW

JANUARY/FEBRUARY 2011 VOLUME 93, NUMBER 1

REVIEW

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**.

ISSN 0014-9187

Economic Freedom and Employment Growth in U.S. States

Thomas A. Garrett and Russell M. Rhine

The authors extend earlier models of economic growth and development by exploring the effect of economic freedom on U.S. state employment growth. They find that states with greater economic freedom—defined as the protection of private property and private markets operating with minimal government interference—experienced greater rates of employment growth. In addition, they find that less-restrictive state and national government labor market policies have the greatest impact on employment growth in U.S. states. Beyond labor market policies, state employment growth is influenced by state and local government policies, but not the policies of all levels of government, including the national government. Their results suggest that policymakers concerned with employment should seriously consider the degree to which their own labor market policies and those of the national government may be limiting economic growth and development in their respective states. (JEL H70, O20, O51, R58)

Federal Reserve Bank of St. Louis *Review*, January/February 2011, *93*(1), pp. 1-18.

Large differences exist in the economic growth and development of countries around the world. An extensive literature finds numerous factors that, taken together, explain why certain countries experience greater rates of income and employment growth than others. The most-cited factors contributing to economic growth include the stock of human capital, investment in technology, trade specialization and foreign direct investment, and low levels of political corruption.[1] In addition to these factors, a more recent literature has explored the role of economic and political institutions in the economic growth of countries. Studies have shown that countries with greater economic freedom—meaning the protection of private property and private markets operating with minimal government interference—have greater rates of economic growth than countries with lower levels of economic freedom (Cole, 2003; Sturm and De Haan, 2001; Powell, 2003; Gwartney, 2009).[2]

Differences in economic growth (as measured by income and employment) also exist across subnational jurisdictions (e.g., states, provinces). For example, average annual employment growth in the United States from 1960 to 2008 was nearly 2 percent, but employment growth in individual states was much different—ranging from 0.8 percent in New York to nearly 5.5 percent in Nevada.

[1] See Barro (1997, 2001) and Barro and Sala-i-Martin (2004) for a review of the literature on cross-country economic growth. See also Billger and Goel (2009), Chatterjee and Turnovsky (2007), and Blankenau and Simpson (2004).

[2] In *Economic Freedom of the World*, Gwartney and Lawson (2009) derive a single economic freedom index number for each country that places each country on a continuum from 0 to 10, where 10 represents the highest degree of reliance on free-market capitalism. The index considers five categories: the size of government, property rights and the legal system, trade freedom, sound money, and minimal regulation.

Thomas A. Garrett is an assistant vice president and economist at the Federal Reserve Bank of St. Louis. Russell M. Rhine is an associate professor of economics at St. Mary's College of Maryland.

In addition, the average annual per capita income growth for the 10 Canadian provinces from 1981 to 2008 was 4.3 percent, but the growth rates for individual provinces ranged from a low of 3.8 percent in British Columbia to a high of 5.3 percent in Newfoundland and Labrador.[3]

Many factors that explain differences in cross-country growth also explain differences in state economic growth. Crain and Lee (1999) and Garrett, Wagner, and Wheelock (2007) have shown that income growth is higher in U.S. states with greater industrial diversity, a greater percentage of the population with a college degree, a greater percentage of the population in the labor force, and state government as a smaller share of gross state product (GSP). Tomljanovich (2004) demonstrated that higher state tax rates reduce state economic growth (measured by per capita GSP) for several years following a tax increase.[4] Similarly, Nickell, Nunziata, and Ochel (2005) and Daveri and Tabellini (2000) found that higher labor taxes reduced employment. Finally, Quan and Beck (1987) and Nistor (2009) found that states and counties with greater human capital investment (i.e., education) had lower unemployment rates and greater employment growth.

It is reasonable that differences in economic freedom across states may explain variation in the growth of U.S. states as well. Economic and political institutions, such as business regulation, taxation, and government spending, differ across state governments just as they do across national governments. To date, however, empirical models of state economic growth have essentially ignored the potential role of state economic and political institutions in state-level economic growth.

In this paper, we augment prior models of state economic growth by examining the effect of economic freedom on U.S. state employment growth. We use the state-level economic freedom indices in Karabegovic and McMahon's (2008) *Economic Freedom of North America 2008*.[5] The overall index, described in more detail later, considers three areas of state-level economic freedom—the size of government, taxation, and labor market freedom. In essence, the economic freedom indices measure the size of governments, defined very broadly. The paper's testable hypothesis is that states with greater economic freedom (i.e., smaller government, less taxation, and more labor market freedom) have higher rates of employment growth. More economic freedom in a state can lead to greater employment growth through two channels: (i) by encouraging higher levels of entrepreneurial activity and small-business creation (Kreft and Sobel, 2005) and (ii) by reducing the costs, both financial and regulatory, on existing businesses in the state (Karabegovic and McMahon, 2008).

We conduct several empirical exercises using different measures of economic freedom. It is reasonable to believe that the three areas of economic freedom do not exert equal influences on state employment growth. This belief is motivated by the fact that each area of economic freedom has a different impact on other state-level economic variables, such as entrepreneurship and income inequality (Kreft and Sobel, 2005, and Ashby and Sobel, 2008). Thus, we not only test for the effect of aggregate economic freedom in our employment growth models, but we also consider how each of the three areas of aggregate economic freedom influences state employment growth. This provides an opportunity to determine which economic and political factors (the size of government, taxation, or labor market freedom) have the greatest impact on state employment growth.

Because federal economic policies (e.g., minimum wage legislation, federal personal and corporate income taxes, federal government transfers to states) influence the economic and political climate in individual states, our empirical models consider state-level economic freedom indices based on state and local government policies, as well as economic freedom indices for national, state, and local government policies. This allows us to determine which level(s) of government policy have the greatest impact on state employment growth.

---

3 U.S. employment data are from the Bureau of Labor Statistics. Canadian data are from Statistics Canada.

4 See also Dye (1980) and Wasylenko and McGuire (1985).

5 *Economic Freedom of North America 2008* can be accessed at www.freetheworld.com/efna.html.

Our results indicate that economic freedom is a significant factor in state employment growth in addition to the more traditional determinants of growth, such as industrial diversity and human capital. We find that the effect of economic freedom on state employment growth varies depending on the period studied and which economic freedom index is used. Differences are found when we consider (i) each of the three areas of economic freedom individually and (ii) economic freedom based on state government policies versus state and national government policies. The results have important policy implications for all those concerned with subnational economic growth and development.

ECONOMIC FREEDOM IN U.S. STATES

The state-level economic freedom indices used here are from *Economic Freedom of North America 2008* (Karabegovic and McMahon, 2008). The indices are "an attempt to gauge the extent of the restrictions on economic freedom imposed by governments in North America" (Karabegovic and McMahon, 2008, p. 3). The underlying intuition for the indices is that once state governments reach a certain size in terms of taxation, spending, and regulation, additional government intervention in the private sector reduces economic growth. One conjecture in the literature is that the optimal size of each state government in terms of maximizing private sector economic growth (through government spending on infrastructure, eliminating externalities, and so on) is less than the current size of state and local governments (Mitchell, 2005). Thus, if the conjecture is true, it is expected that, on the margin, states with relatively greater government intrusion in the private sector (i.e., those with lower economic freedom indices) will experience lower economic growth.

The economic freedom indices are constructed on a 10-point scale, with a higher value denoting greater economic freedom. Economic freedom is evaluated using two levels of government—the subnational level (state and local governments) and the "total" government level (national, state, and local governments).[6] Overall freedom indices for the two levels of government are each based on three areas of government intervention: the size of government (Area 1), takings and discriminatory taxation (Area 2), and labor market freedom (Area 3).[7] A higher index for each of the three areas implies a smaller state government, less taxation, and greater labor market freedom, respectively. Each area has its own economic freedom index, and the overall index is an equally weighted average of the three areas. The indices are constructed using data on each of the components (Table 1), and each economic freedom index for a particular state is relative to that of all other states by construction.[8]

Figure 1 illustrates the variation in overall economic freedom (subnational level) across the continental U.S. states for 2005. Economic freedom ranged from a low of 5.5 in West Virginia to a high of 8.3 in Delaware, with an average value of 6.9. States in the Southeast and the Midwest tend to have a higher level of economic freedom than states on the West Coast and in the Northeast. Although not shown here, the level of economic freedom in each state is similar in proximal years, but large differences do exist in the level of economic freedom in a state over time.[9]

The primary advantage of the economic freedom index is that it provides a concise, summary measure of government restrictions on free-market activity.[10] As a result, not only have dozens of

---

6 Although not used in this paper, provincial economic freedom indices are available for the Canadian provinces; see Karabegovic and McMahon (2008).

7 "Discriminatory taxation" means the taxing of only those individuals engaging in a particular activity (e.g., sales taxes are paid only by those who make taxable retail purchases). The term "takings" refers to the revenue to governments acquired through taxation. The average correlation among the three areas of state-level economic freedom is 0.51.

8 See Karabegovic and McMahon (2008, pp. 77-80) for a discussion of how the economic freedom indices are calculated.

9 Annual state-level economic freedom indices from 1980 to 2005 are available at the website of the Economic Freedom Network (www.freetheworld.com/efna.html).

10 An econometric advantage of using the economic freedom index in empirical modeling rather than the set of variables in Table 1 is that many of the latter are highly correlated, thus likely decreasing the precision of the coefficient estimates. The use of a single measure of economic freedom eliminates this potential problem.

Figure 1

Economic Freedom in U.S. States (2000)

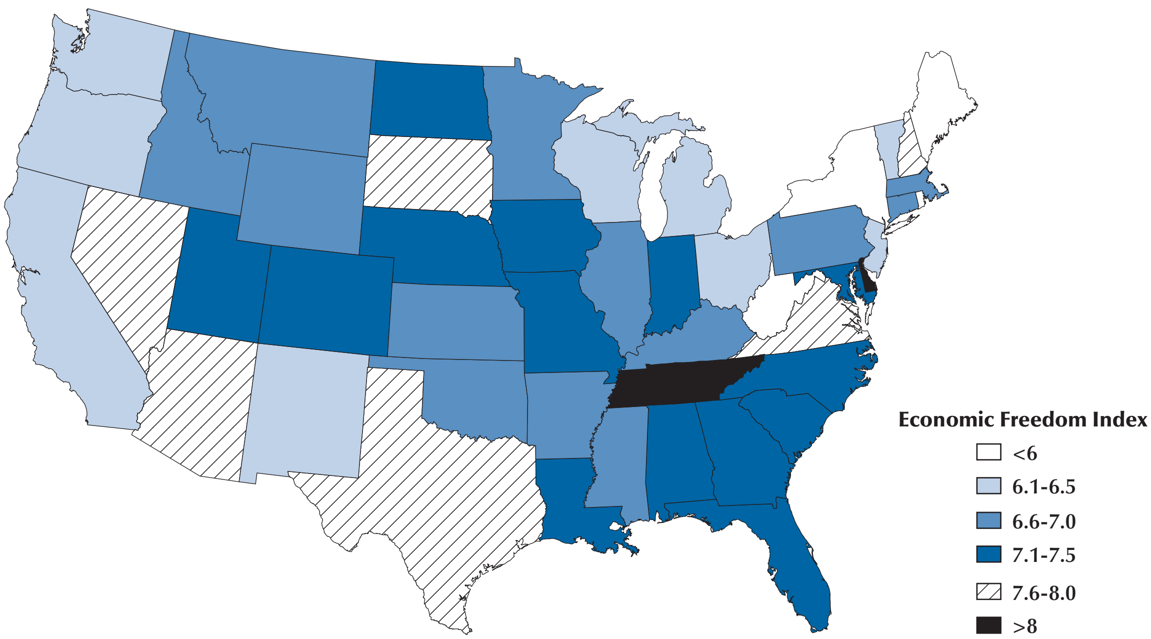


Table 1

Areas and Components of State-Level Economic Freedom

Area	Component
Area 1: Size of government	
1A	General consumption expenditures by government as a percentage of GDP
1B	Transfers and subsidies as a percentage of GDP
1C	Social Security payments as a percentage of GDP
Area 2: Takings and discriminatory taxation	
2A	Total tax revenue as a percentage of GDP
2B	Top marginal income tax rate and the income threshold at which it applies
2C	Indirect tax revenue as a percentage of GDP
2D	Sales taxes collected as a percentage of GDP
Area 3: Labor market freedom	
3A	Minimum wage legislation
3B	Government employment as a percentage of total employment
3C	Union density

SOURCE: Karabegovic and McMahon (2008).

studies explored the impact of economic freedom on various measures of economic growth, but additional studies also have explored how the economic freedom index correlates with other variables, such as health (Norton, 1998, and Esposto and Zaleski, 1999), migration (Melkumian, 2004), income inequality (Ashby and Sobel, 2008), the productivity of investment (Dawson, 1998), and entrepreneurship (Ovaska and Sobel, 2005, and Kreft and Sobel, 2005).

Although the economic freedom index has been used in many studies, it is not without critics (Hanson, 2003). One criticism is that the index, because it is a summary measure, is less precise in measuring economic freedom than many of the component variables used to create the index, thus generating bias in empirical models. One way to mitigate this problem (as we do here) is to estimate regression models using the economic freedom index for each area (see Table 1) in addition to the overall freedom index (Heckelman, 2005). A second criticism is the simultaneity (both in levels and in growth rates) between the economic freedom index and economic outcomes such as gross domestic product (GDP) and income.[11] Studies have regressed future growth on the contemporaneous economic freedom index to minimize this problem, which is the methodology we follow here. A final criticism of the economic freedom index is that it entails ideological bias because the index is created by a free-market organization—the Fraser Institute. Ashby and Sobel (2008) argue, however, that even if an ideological bias exists, this bias actually ensures that the index does capture the desired measurements.

Despite some controversy surrounding the economic freedom index, we assume that the index approach is valid. We leave it to future research to determine whether the index is an appropriate gauge of economic freedom. As discussed in the following section, we design our empirical models to ensure that the potential econometric problems of the economic freedom index—specification bias and simultaneity—are taken into account.

[11] If economic freedom is a normal good, then wealthier countries would demand more economic freedom.

DATA AND EMPIRICAL METHODOLOGY

We estimate our models of state employment growth for three separate periods (1980-90, 1990-2000, and 2000-05) using data for the 50 U.S. states.[12] We perform the analysis for the three periods to assess the temporal robustness of the relationship, if any, between economic freedom and state employment growth.[13] We run several empirical specifications, each of which considers one of the several economic freedom indices discussed earlier—the overall index, the index for each of the three areas (see Table 1), and the indices for subnational government (state and local) policies and total government (national, state, and local) policies.

Our empirical models are designed to examine the degree to which differences in economic freedom across states in the initial year of each 10- and 5-year period can explain differences in state employment growth over the period.[14] Two reasons exist for choosing this framework. First, regressing future employment growth on an initial value of the economic freedom index minimizes any simultaneity and endogeneity between the economic freedom index and state employment growth that exists over time (Heckelman, 2005). Second, a time lag exists between when government policies are implemented and when their effects are known or realized, so it is reasonable

[12] We end the analysis in 2005, the latest year for which economic freedom data were available at the time of writing.

[13] The economic freedom index is not available before 1981. Thus, our models of employment growth over the 1980-90 period use the economic freedom index for 1981.

[14] Our empirical specification is similar to that used in the convergence literature. Implicit in the empirical specification is the idea that each economy has a steady-state growth path that follows a time trend. Quah (1993) provides cross-country evidence on income growth that refutes this assumption. Durlauf (2001) raises other potential problems, such as nonlinearities, a disconnect between growth theory and empirical modeling (i.e., which variables should be included in growth models and the potential problem of simultaneity), and, finally, heterogeneous parameters. We argue that differences across states in terms of heterogeneous parameters and growth paths are likely to be significantly less than differences across countries because political systems and components of government revenue and spending are much more similar across states than across countries.

to model future employment growth as a function of past government policies.[15]

Although there is little disagreement that fiscal policy and government regulation work with lags, we have no a priori hypothesis as to the exact lag to consider in our empirical models. Previous studies have considered lags ranging from several years to several decades. To ensure consistency with many previous studies, we chose to explore the effect of economic freedom on state employment growth over two 10-year periods and one 5-year period. Our results are, of course, specific to the starting years chosen and the length of time for which we specify employment growth.

Previous studies on state economic growth serve as a guide for the variables to include in our models. Of the dozen or so variables we could have included, we chose those that were significant determinants of economic growth in earlier studies. To alleviate likely simultaneity between state employment growth and each of the independent variables, some variables in our models (described below) are specified as growth rates, whereas the levels of other variables are included for the initial year of the study period (1980, 1990, or 2000).[16]

We account for the human capital of a state by including the percentage change in the share of the state's population older than 25 years of age that has obtained a bachelor's degree or higher (Quan and Beck, 1987, and Nistor, 2009).[17] Our expectation is that states with greater growth in the share of the population with a college degree will have higher rates of employment growth.

State population density (persons per square mile) for the initial year is included to capture the effects of agglomeration on state economic growth. Haughwout (1999), Blumenthal, Wolman, and Hill (2009), and Puga (2010) have demonstrated that areas with greater agglomeration experience higher growth rates.[18] Assuming a concave path for state economic growth as suggested by the convergence literature (Carlino and Mills, 1996, and Webber, White, and Allen, 2005), we expected that states with greater population density in the initial year would have lower rates of future employment growth.

For the initial year of each study period we include employment for the manufacturing and service sectors to control for industry mix. Each type of employment is expressed as a percentage of total employment (Nichol, 2009).[19] The expected sign of each variable is unclear. Generally, the manufacturing sector's share of total employment in U.S. states has been declining, whereas the service sector's share has increased. States with a greater share of employment in manufacturing in the initial year may have experienced slower total employment growth if employment growth in other sectors, including service sectors, was insufficient to offset any decline in manufacturing. Similarly, states with a greater share of employment in services in the initial year may have experienced greater employment growth if service sector growth offset declining growth in other sectors. In short, the sign of each variable depends on the relative size of each sector in the initial year and the employment dynamics in all other sectors (Elhorst, 2003).

Descriptive statistics for the variables used in the analysis are shown in Table 2. A few comments regarding the data are noteworthy. Employment growth across the states averaged 21 to 23 percent for the 1980s and 1990s and 2.5 percent in the early 2000s. Overall economic freedom averaged slightly above 7.0 in each of the three years. Economic freedom in Areas 1 and 2 (except for the 1990-2000 period) decreased over time, whereas economic freedom for Area 3 increased over time. The standard deviations (SDs) of the

[15] See Auerbach and Gale (2009).

[16] We considered several variables in addition to the variables used in the final empirical models. Specifically, we considered human capital spending, the share of a state's population between 18 and 64 years of age, and the measure of industrial diversity suggested by Crain and Lee (1999). We considered these variables in levels and in percent changes. The coefficients of these variables were statistically insignificant in most regression specifications, and there was little change in the size and sign of the remaining coefficients. We thus chose to drop these variables from our final specifications.

[17] Data are from the U.S. Census.

[18] State population and area data were obtained from the U.S. Census.

[19] Employment share data were calculated using industry employment data from the Bureau of Economic Analysis.

Table 2

Descriptive Statistics

Variable	Mean	SD
Percent change in employment$_{1980\text{-}1990}$	21.53	12.62
Percent change in employment$_{1990\text{-}2000}$	23.20	12.09
Percent change in employment$_{2000\text{-}2005}$	2.56	4.63
Economic freedom$_{1981}$	7.052	0.942
Economic freedom$_{1990}$	7.060	0.688
Economic freedom$_{2000}$	7.010	0.692
Economic freedom$_{1981}$ (Area 1)	7.702	0.988
Economic freedom$_{1990}$ (Area 1)	7.616	0.800
Economic freedom$_{2000}$ (Area 1)	7.330	0.938
Economic freedom$_{1981}$ (Area 2)	7.228	1.036
Economic freedom$_{1990}$ (Area 2)	6.938	0.739
Economic freedom$_{2000}$ (Area 2)	6.988	0.786
Economic freedom$_{1981}$ (Area 3)	6.220	1.161
Economic freedom$_{1990}$ (Area 3)	6.638	0.907
Economic freedom$_{2000}$ (Area 3)	6.742	0.811
Percent change in bachelor's degree$_{1980\text{-}1990}$	22.87	7.52
Percent change in bachelor's degree$_{1990\text{-}2000}$	20.65	4.35
Percent change in bachelor's degree$_{2000\text{-}2005}$	10.91	2.73
Population density$_{1980}$	154.87	222.60
Population density$_{1990}$	166.19	235.35
Population density$_{2000}$	181.90	250.15
Percent in services$_{1980}$	27.30	4.47
Percent in services$_{1990}$	34.48	4.66
Percent in services$_{2000}$	38.98	4.56
Percent in manufacturing$_{1980}$	21.24	8.27
Percent in manufacturing$_{1990}$	15.68	5.85
Percent in manufacturing$_{2000}$	13.06	4.83

NOTE: The sample size is 50 for 1980, 1990, and 2000. The economic freedom index is for the state and local government level. Area 1, size of government; Area 2, takings and discriminatory taxation; Area 3, labor market freedom. See text for further description of the economic freedom indices.

Table 3

State and Local Economic Freedom and State Employment Growth (1980-90)

Variable	Dependent variable: Percent change in state payroll employment (1980-90) (1)	(2)	(3)	(4)
Economic freedom$_{1981}$	3.768* (1.72)			
Economic freedom$_{1981}$ (Area 1)		5.684** (3.23)		
Economic freedom$_{1981}$ (Area 2)			0.953 (0.53)	
Economic freedom$_{1981}$ (Area 3)				2.754 (1.30)
Percent Δ in bachelor's degree	0.529 (0.95)	0.536 (1.10)	0.471 (0.80)	0.648 (1.11)
Population density$_{1980}$	–0.018** (2.27)	–0.016** (2.12)	–0.019** (2.24)	–0.020** (2.41)
Percent in services$_{1980}$	1.482** (3.04)	1.638** (3.67)	1.446** (2.74)	1.399** (2.79)
Percent in manufacturing$_{1980}$	0.406 (1.10)	0.589* (1.78)	0.373 (0.98)	0.275 (0.70)
Constant	–50.149** (1.99)	–75.600** (3.23)	–29.471 (1.10)	–37.726 (1.39)
Regional dummy variables	Yes	Yes	Yes	Yes
Adjusted R^2	0.274	0.339	0.225	0.263
Adjusted R^2 (omit Freedom Index)	0.240	0.240	0.240	0.240
Observations	50	50	50	50

NOTE: * Denotes significance at the 10 percent level, ** at 5 percent. Absolute *t*-statistics are listed in parentheses and are based on White's heteroskedasticity-consistent standard errors. Area 1, size of government; Area 2, takings and discriminatory taxation; Area 3, labor market freedom. See text for further description of the economic freedom indices.

economic freedom indices suggest that variation in economic freedom across the states generally decreased over time.

RESULTS

The empirical results for each period are shown in Tables 3 through 5. All regressions included the economic freedom index at the state and local government level. In addition, all regressions included a set of eight regional dummy variables based on Census divisions to control for heterogeneity in growth rates across regions.

A brief summary of the findings for the other independent variables is warranted before we focus on the economic freedom results.[20] The coefficients of the percentage change in the share of the population with a bachelor's degree are positive, but they are statistically significant only for the 2000-05 period (and in one specification for the 1990-2000 period). The coefficient of the percentage change in population density in the

[20] We also pooled the three periods to estimate a panel data model. The coefficient estimates from this model were roughly the average of the coefficients estimates from the three separate models. The results from the panel estimation are available on request.

Table 4

State and Local Economic Freedom and State Employment Growth (1990-2000)

Variable	Dependent variable: Percent change in state payroll employment (1990-2000) (1)	(2)	(3)	(4)
Economic freedom$_{1990}$	4.459**			
	(2.57)			
Economic freedom$_{1990}$ (Area 1)		3.270**		
		(2.10)		
Economic freedom$_{1990}$ (Area 2)			2.187	
			(1.26)	
Economic freedom$_{1990}$ (Area 3)				4.421**
				(2.75)
Percent Δ in bachelor's degree	0.652	0.456	0.524	0.768*
	(1.43)	(1.08)	(1.11)	(1.71)
Population density$_{1990}$	–0.016**	–0.016**	–0.016**	–0.016**
	(2.60)	(2.66)	(2.52)	(2.39)
Percent in services$_{1990}$	1.383**	1.541**	1.532**	1.205**
	(2.15)	(2.41)	(2.30)	(1.71)
Percent in manufacturing$_{1990}$	0.392	0.517	0.479	0.334
	(0.83)	(1.14)	(1.03)	(0.73)
Constant	–71.041**	–70.004**	–60.517**	–61.834**
	(2.57)	(2.55)	(2.11)	(2.26)
Regional dummy variables	Yes	Yes	Yes	Yes
Adjusted R^2	0.619	0.611	0.580	0.627
Adjusted R^2 (omit Freedom Index)	0.572	0.572	0.572	0.572
Observations	50	50	50	50

NOTE: * Denotes significance at the 10 percent level, ** at 5 percent. Absolute *t*-statistics are listed in parentheses and are based on White's heteroskedasticity-consistent standard errors. Area 1, size of government; Area 2, takings and discriminatory taxation; Area 3, labor market freedom. See text for further description of the economic freedom indices.

initial year is negative and significant for the 1980-90 and 1990-2000 periods. This finding corresponds to our prior hypothesis that states with higher agglomeration have lower future employment growth rates. The coefficients of the share of total employment in manufacturing are negative and significant for the 2000-05 period but are generally not significant for the two earlier periods. The coefficients of the share of total employment in services are positive and significant for the 1980-90 and 1990-2000 periods but not for 2000-05.

Our key variables of interest are the economic freedom indices. We first discuss the results regarding the effect of overall economic freedom on state employment growth (column 1 of Tables 3 through 5). In accordance with our hypothesis, the coefficient of the overall economic freedom index is positive and significant for all three periods. The results indicate that a one-unit increase in the economic freedom index (roughly equal to 1 SD) in the initial year of a period resulted in increased employment growth of 3.8 percentage points from 1980 to 1990, 4.5 percentage points from 1990 to 2000, and 1.4 percentage points from 2000 to 2005. In terms of explaining the variation in state employment growth, a comparison of the adjusted R^2 from each of the

Table 5

State and Local Economic Freedom and State Employment Growth (2000-05)

	Dependent variable: Percent change in state payroll employment (2000-05)			
Variable	(1)	(2)	(3)	(4)
Economic freedom$_{2000}$	1.351*			
	(1.90)			
Economic freedom$_{2000}$ (Area 1)		0.937*		
		(1.72)		
Economic freedom$_{2000}$ (Area 2)			0.546	
			(1.03)	
Economic freedom$_{2000}$ (Area 3)				1.797**
				(2.35)
Percent Δ in bachelor's degree	0.543**	0.500*	0.467*	0.591**
	(2.17)	(1.96)	(1.90)	(2.53)
Population density$_{2000}$	–0.003	–0.003	–0.003	–0.003
	(1.29)	(1.38)	(1.23)	(1.45)
Percent in services$_{2000}$	–0.064	–0.085	0.022	–0.095
	(0.36)	(0.46)	(0.11)	(0.52)
Percent in manufacturing$_{2000}$	–0.514**	–0.533**	–0.473**	–0.501**
	(3.77)	(3.96)	(3.59)	(3.46)
Constant	0.282	4.403	1.987	–0.949
	(0.03)	(0.50)	(0.19)	(0.11)
Regional dummy variables	Yes	Yes	Yes	Yes
Adjusted R^2	0.573	0.570	0.544	0.595
Adjusted R^2 (omit Freedom Index)	0.547	0.547	0.547	0.547
Observations	50	50	50	50

NOTE: * Denotes significance at the 10 percent level, ** at 5 percent. Absolute *t*-statistics are listed in parentheses and are based on White's heteroskedasticity-consistent standard errors. Area 1, size of government; Area 2, takings and discriminatory taxation; Area 3, labor market freedom. See text for further description of the economic freedom indices.

reported models with the adjusted R^2 from unreported models that omit the economic freedom index (the last three rows of Tables 3 through 5) shows that the overall economic freedom index explains roughly 3 to 5 percent of the total variation in state employment growth in each period.

The results for the economic freedom indices for Areas 1, 2, and 3 are shown in columns 2 through 4 of Tables 3 through 5, respectively.[21] First, consider the economic freedom index for the size of government (Area 1). The coefficient of this index is positive and significant for all three periods, revealing that employment growth is higher in states with smaller state and local governments as a share of total output. The Area 1 freedom index coefficient is largest for the 1980-90 period, revealing that a one-unit change in the index resulted in a 5.7-percentage-point increase in state employment growth. A one-unit change in the index for the two remaining periods resulted in a 3.3-percentage-point increase (1990-2000)

[21] We initially included the three area economic freedom indices in a single regression equation. However, the high correlation among the area freedom indices (average $\rho \approx 0.65$) dramatically decreased the precision of the coefficients estimates and, in some cases, produced improbable results. Thus, despite the recognized potential for omitted variable bias, we chose to estimate separate regressions for each of the area economic freedom indices.

and a 0.9-percentage-point increase (2000-05) in employment growth. Again comparing adjusted R^2s, the inclusion of the Area 1 economic freedom index explains roughly 2 to 10 percent of the total variation in state employment growth.

The coefficient estimates for the Area 2 economic freedom indices (takings and discriminatory taxation), although positive, are not statistically significant in any period. Thus, relative differences in taxation across the states in the initial years do not influence future state employment growth. One reason for this finding may be that the majority of taxes considered by the economic freedom index (see Table 1) are consumer-based taxes, and the taxes levied on businesses may be ultimately borne by consumers. Another explanation may be that the growth periods of 5 and 10 years considered in this paper are longer than the impact of tax changes on employment growth, as in Tomljanovich (2004), which showed that tax changes have only short-run impacts on economic growth (within 5 years).

The coefficient estimates on the economic freedom index for Area 3 (labor market freedom) are positive and statistically significant for the 1990-2000 and 2000-05 periods. A one-unit increase in labor market economic freedom increases employment growth by 4.4 percentage points and 1.8 percentage points for the 1990-2000 and 2000-05 periods, respectively, and the inclusion of the labor market freedom index explains about 5 percent of the total variation in state employment growth based on each adjusted R^2.

The magnitude of the coefficients is larger for Area 3 than for Areas 1 and 2 in the 1990-2000 and 2000-05 periods, thereby suggesting that in more recent years labor market freedom has had a greater impact on state employment growth than the size of government and taxation. This is an intuitive result since business formation and expansion is directly influenced by labor costs, which constitute a significant portion of a firm's total costs. Our finding agrees with that of Kreft and Sobel (2005), who find that, of the three area economic freedom indices, labor market freedom has the largest impact on the number of sole proprietorships across the states.

At this point, a summary of our empirical results regarding the impact of economic freedom on state employment growth is worthwhile. We find that overall economic freedom has a positive and statistically significant effect on future state employment growth for the three periods. In addition, state labor market policies have the greatest impact on state employment growth, and the size of state governments has some impact as well. There is no evidence that taxation has a significant impact on future state employment growth for the periods considered in this study. Overall, the various economic freedom indices explain roughly 3 to 5 percent of the total variation in state employment growth, on average. This finding, in addition to the significant coefficient estimates, suggests that, at least for our sample periods, economic freedom is a significant factor in state employment growth, but economic freedom explains a relatively small percentage of the across-state variation in state employment growth.

The Economic Significance of Economic Freedom

In this section we highlight the economic significance of the economic freedom coefficient estimates by examining employment growth for the 10 states with the lowest economic freedom rankings for the initial year of each study period. Specifically, for these 10 states, we ask what state employment growth would have been if each state had an economic freedom index equal to the average U.S. state freedom index (see Table 2).

To answer this question, we first use the previous regression estimates (column 1 of Tables 3 through 5) to predict employment growth using each state's actual level of economic freedom. Next, we predict the state's employment growth using the mean level of economic freedom across the states and then we compare the two predictions of employment growth. Finally, we use the predicted level of employment in the initial year for each period to assess the increase in state employment for each period. One caveat of this prediction exercise is that we assume the effect of economic freedom on employment growth is the same for each state—that is, the estimated coefficients in column 1 of Tables 3 through 5 reflect the freedom-employment relationship in each state.

Let us consider some findings shown in Table 6.[22] Employment growth from 1980 to 1990 in New York would have been over 7 percentage points higher (column 3 vs. column 2) if its level of economic freedom (5.00) had equaled the U.S. state average (7.05). This translates into over 550,000 jobs (column 4). For Montana, the state with the lowest level of economic freedom in 1990, employment growth would have been 6.5 percentage points higher (about 19,300 jobs) if its level of economic freedom had equaled the U.S. state average. Finally, for the 2000-05 period, employment growth in Vermont would have been positive (0.7 percent, 2,900 jobs) if its economic freedom had equaled the U.S. state average.

In sum, comparing predicted employment growth based on the actual economic freedom index with the average U.S. state freedom index reveals that, on average across the 10 states, the states with the lowest level of economic freedom would have experienced employment growth roughly 5 percentage points higher over each 10-year period if the states had economic freedom equal to the U.S. state average. Employment growth would have been about 1.3 percentage points higher, on average, over the 2000-05 period.

National and State Economic Freedom Indices

The previous analysis considered economic freedom at the state and local government levels. In this section, we consider how economic freedom at the national, state, and local levels of government (i.e., "total government") influences state employment growth. The three areas and subcomponents for the "total government" economic freedom index are identical to those used for the state and local government freedom index (see Table 1), except the total government area indices also incorporate federal expenditures, tax collections, employment, and minimum wage legislation.[23] We reestimated all regressions in Tables 3 through 5 using the total government economic freedom indices in Karabegovic and McMahon (2008). The coefficients of the total government economic freedom indices can be compared with the state and local government economic freedom coefficients to assess the marginal effect of national government policies on state employment growth. For the sake of brevity, Table 7 presents only the coefficient estimates for the economic freedom indices.[24]

The results in Table 7 indicate that economic freedom at the total government level generally has no impact on state employment growth. The majority of the coefficients of economic freedom, although positive, are statistically insignificant. In only 3 of the 12 specifications is the effect of the freedom index statistically significant. These results, compared with the earlier results, generally suggest that relative differences in state and local government policy influence state employment growth, whereas relative differences in total government policies do not have a significant influence on state employment growth.

The one clear exception is labor market policies. The coefficient for Area 3 (labor market freedom) is significant for the 1990-2000 and 2000-05 periods, as is the overall index for 1990 to 2000. The coefficients of labor market freedom are greater than those obtained with state-level labor market freedom indices, thus indicating the cumulative increase in employment growth as a result of considering national-level labor market policies in addition to state-level policies (6.7 vs. 4.4 for 1990-2000 and 2.4 vs. 1.8 for 2000-05). In addition, the relative increase in the size of the coefficients for total government labor market policies compared with state-level policies is less than the size of the coefficients when state-level labor market policies are considered. This indicates that state-level labor market policies influence state employment growth more than national-level labor market policies.

[22] Another caveat is that, because the economic freedom index is a relative index, in reality the economic freedom index for one state cannot change without changing the index for all other states. Thus, a state cannot simply move to the mean economic freedom level because most likely the mean level of economic freedom will change. Nevertheless, the exercise does reveal how much higher employment growth would have been if the low-freedom states had an economic freedom index equal to the mean of all U.S. states.

[23] The average correlation between the state and local government freedom indices (the overall index and that of each of the three areas) and the total government indices is about 0.50.

[24] Our complete results are available on request.

Table 6

Forecasted Employment Gains from Greater Economic Freedom: The 10 States with the Lowest Economic Freedom

	(1)	(2)	(3)	(4)
State	Freedom index	Predicted employment growth (%)	Predicted employment growth at freedom mean (%)	Increase in employment at freedom mean
1980-90: Freedom score 1980				
New York	5.00	23.17	30.90	557,240
Michigan	5.20	9.15	16.13	240,264
Rhode Island	5.50	12.70	18.55	23,275
Maine	5.70	23.75	28.84	21,294
West Virginia	5.70	8.31	13.41	32,909
Oregon	5.80	28.16	32.88	49,298
Vermont	5.80	26.55	31.27	9,435
Hawaii	6.00	27.80	31.76	16,014
Minnesota	6.00	13.24	17.20	76,162
California	6.10	36.17	39.76	353,297
1990-2000: Freedom score 1990				
Montana	5.60	37.29	43.80	19,335
New York	5.70	9.00	15.06	498,117
West Virginia	5.80	12.48	18.09	35,396
Michigan	5.90	17.01	22.18	204,104
Maine	6.10	12.65	16.93	22,901
North Dakota	6.20	16.74	20.58	10,200
Minnesota	6.30	23.40	26.79	72,386
Oregon	6.30	25.85	29.24	42,564
Rhode Island	6.30	1.97	5.35	15,385
Washington	6.30	24.62	28.01	72,623
2000-05: Freedom score 2000				
West Virginia	5.50	1.33	3.37	15,014
Alaska	5.80	7.90	9.53	4,643
Maine	5.80	1.92	3.55	9,857
Rhode Island	5.90	–0.02	1.48	7,153
New York	6.00	2.23	3.59	117,866
Hawaii	6.10	10.31	11.54	6,774
Montana	6.10	9.68	10.91	4,807
New Mexico	6.20	6.90	8.00	8,153
Vermont	6.30	–0.27	0.69	2,868
California	6.40	3.44	4.26	119,397

NOTE: Column 2 contains the state-specific predicted values from the first regression specification in Tables 3 through 5. Column 3 lists the state-specific predicted values from the first regression specification in Tables 3 through 5 using the mean value of economic freedom (state and local government only): 7.05 for 1980, 7.06 for 1990, and 7.01 for 2000. The data in column 4 were computed using 1980, 1990, and 2000 employment levels.

Table 7

Total Government Economic Freedom and State Employment Growth

Variable	Dependent variable: Percent change in state payroll employment (1)	(2)	(3)	(4)
1980-90				
Economic freedom	1.963			
	(0.44)			
Economic freedom (Area 1)		2.850		
		(1.01)		
Economic freedom (Area 2)			–0.315	
			(0.05)	
Economic freedom (Area 3)				0.985
				(0.29)
1990-2000				
Economic freedom	4.123*			
	(1.80)			
Economic freedom (Area 1)		2.411		
		(1.42)		
Economic freedom (Area 2)			1.933	
			(0.87)	
Economic freedom (Area 3)				6.728**
				(3.10)
2000-05				
Economic freedom	1.241			
	(1.51)			
Economic freedom (Area 1)		0.783		
		(1.24)		
Economic freedom (Area 2)			0.263	
			(0.38)	
Economic freedom (Area 3)				2.413**
				(2.50)

NOTE: *Denotes significance at the 10 percent level, ** at 5 percent. Absolute *t*-statistics are listed in parentheses and are based on White's heteroskedasticity-consistent standard errors. Area 1, size of government; Area 2, takings and discriminatory taxation; Area 3, labor market freedom. See text for further description of the economic freedom indices. Each regression contains the same variables as the state and local economic freedom regressions shown in Tables 3 through 5. The full set of estimates is available on request.

We explored the interesting finding that economic freedom at the total government level does not explain state employment growth except in the case of labor market policies. A look at the raw data shows that, on average, the economic freedom indices at the total government level are generally smaller than those at the state and local levels and, more importantly, the indices have significantly smaller SDs. For example, the average 1990 state and local government freedom index is 7.06 and has an SD of 0.69, whereas the average 1990 total government freedom index is 7.02 with an SD of 0.52. Thus, across states there is much less variation in total government economic freedom than in state and local economic freedom. This does not imply, however, that total government economic freedom does not necessarily influence state employment growth in a single state, but rather that differences in state and local government policy, and not total government policy, explain a portion of the variation in employment growth across U.S. states.

SUMMARY AND CONCLUSION

Explaining differences in the economic growth and development of countries and regions around the world has been the focus of a wide body of research. Human capital, technology, trade specialization, and economic freedom—meaning the protection of private property and private markets operating with minimal government interference—are generally considered the principal determinants of economic growth and development. A more recent line of research has attempted to explain economic growth and development across subnational jurisdictions as well. To date, however, empirical models of subnational economic growth have ignored the importance of economic freedom in explaining differences in the economic growth of subnational jurisdictions.

In this paper, we augmented previous models of subnational economic growth by considering the role of economic freedom in explaining differences in employment growth in U.S. states. We considered employment growth over three periods: 1980-90, 1990-2000, and 2000-05. For each period, we find that states with greater overall economic freedom have higher rates of employment growth. This finding supports the conjecture in earlier literature that the current size of state and local governments, defined broadly, is larger than optimal. Generally, we find that a one-unit increase in the economic freedom index (roughly equal to 1 SD) increases employment growth by 1 to 4 percentage points for our sample periods, depending on specification. In addition, roughly 2 to 5 percent of the variation in employment growth across the states is explained by economic freedom.

Further results suggest that labor market freedom and a smaller state government, which are two components of overall economic freedom, are important determinants of employment growth across U.S. states, with the former factor the more important. Different tax policies across states do not have a significant effect on state employment growth, however. We also provide the interesting result that, in most cases, differences in employment growth across states can be partly explained by state and local government policies, but not policies of all levels of government. We do find, however, that labor market freedom at the state and national levels is a significant determinant of state employment growth, and state-level labor market policies appear to be more influential than national-level policies. This finding serves as an important policy implication for officials concerned with increasing economic growth.

Of note, the limitations of our study also serve as areas for future research. First, our results regarding the impact of economic freedom on employment growth are specific to the three periods we studied. Although we would generally expect a positive relationship between employment growth and economic freedom, there is no reason to assume that the magnitudes of our coefficient estimates are not period specific. Future research could extend our work by considering different periods, as well as shorter and longer durations, such as 3 years or 20 years. Second, because economic freedom is measured as an index, it is somewhat difficult to precisely implement policy based on our results given that the index is an aggregate of 10 government policy variables. The specific effect of each policy variable on the economic freedom index is unclear. Rather than considering only the overall freedom index and the index for each of the three components (size of government, taxation, labor market freedom), future research might implement a freedom index for each of the 10 government policy variables.

REFERENCES

Ashby, Nathan J. and Sobel, Russell S. "Income Inequality and Economic Freedom in the U.S. States." *Public Choice*, March 2008, *134*(3-4), pp. 329-46.

Auerbach, Alan J. and Gale, William G. "Activist Fiscal Policy to Stabilize Economic Activity." NBER Working Paper No. 15407, National Bureau of Economic Research, October 2009.

Barro, Robert J. *Determinants of Economic Growth: A Cross-Country Empirical Study* (Lionel Robbins Lectures). Cambridge, MA: MIT Press, 1997.

Barro, Robert J. "Human Capital and Growth." *American Economic Review*, May 2001, *91*(2), pp. 12-17.

Barro, Robert J. and Sala-i-Martin, Xavier. *Economic Growth*. Second Edition. Cambridge, MA: MIT Press, 2004.

Billger, Sherrilyn M. and Goel, Rajeev K. "Do Existing Corruption Levels Matter in Controlling Corruption?" *Journal of Development Economics*, November 2009, *90*(2), pp. 299-305.

Blankenau, William F. and Simpson, Nicole B. "Public Education Expenditures and Growth." *Journal of Development Economics*, April 2004, *73*(2), pp. 583-605.

Blumenthal, Pamela; Wolman, Harold L. and Hill, Edward. "Understanding the Economic Performance of Metropolitan Areas in the United States." *Urban Studies*, March 2009, *46*(3), pp. 605-27.

Carlino, Gerald and Mills, Leonard. "Convergence and the U.S. States: A Time-Series Analysis." *Journal of Regional Science*, November 1996, *36*(4), pp. 597-616.

Chatterjee, Santanu and Turnovsky, Stephen J. "Foreign Aid and Economic Growth: The Role of Flexible Labor Supply." *Journal of Development Economics*, September 2007, *84*(1), pp. 507-33.

Cole, Julio H. "The Contribution of Economic Freedom to World Economic Growth, 1980-99." *Cato Journal*, Fall 2003, *23*(2), pp. 189-98; www.cato.org/pubs/journal/cj23n2/cj23n2-3.pdf.

Crain, W. Mark and Lee, Katherine J. "Economic Growth Regressions for the American States: A Sensitivity Analysis." *Economic Inquiry*, April 1999, *37*(2), pp. 242-57.

Daveri, Francesco and Tabellini, Guido. "Unemployment, Growth, and Taxation in Industrial Countries." *Economic Policy: A European Forum*, April 2000, *15*(30), pp. 47-88.

Dawson, John W. "Institutions, Investment, and Growth: New Cross-Country and Panel Data Evidence." *Economic Inquiry*, October 1998, *36*(4), pp. 603-19.

Durlauf, Steven N. "Manifesto for a Growth Econometrics." *Journal of Econometrics*, January 2001, *100*(1), pp. 65-69.

Dye, Thomas R. "Taxing, Spending, and Economic Growth in the American States." *Journal of Politics*, November 1980, *42*(4), pp. 1085-107.

Elhorst, Paul J. "The Mystery of Regional Unemployment Differentials: Theoretical and Empirical Explanations." *Journal of Economic Surveys*, December 2003, *17*(5), pp. 709-48.

Esposto, Alfredo G. and Zaleski, Peter A. "Economic Freedom and the Quality of Life: An Empirical Analysis." *Constitutional Political Economy*, June 1999, *10*(2), pp. 185-97.

Garrett, Thomas A.; Wagner, Gary A. and Wheelock, David C. "Regional Disparities in the Spatial Correlation of State Income Growth, 1977-2002." *Annals of Regional Science*, September 2007, *41*(3), pp. 601-18.

Gwartney, James D. "Institutions, Economic Freedom, and Cross-Country Differences in Performance." *Southern Economic Journal*, April 2009, *75*(4), pp. 937-56.

Gwartney, James D. and Lawson, Robert. *Economic Freedom of the World: 2009 Annual Report*. Vancouver, BC: Economic Freedom Network (Fraser Institute), 2009.

Hanson, John R. II. "Proxies in the New Political Economy: Caveat Emptor." *Economic Inquiry*, October 2003, *41*(4), pp. 639-46.

Haughwout, Andrew F. "State Infrastructure and the Geography of Employment." *Growth and Change*, September 1999, *30*(4), pp. 549-67.

Heckelman, Jac C. "Proxies for Economic Freedom: A Critique of the Hanson Critique." *Southern Economic Journal*, October 2005, *72*(2), pp. 492-501.

Karabegovic, Amela and McMahon, Fred. *Economic Freedom of North America: 2008 Annual Report* (Canadian Edition). Vancouver, BC: Fraser Institute, 2008.

Kreft, Steven F. and Sobel, Russell S. "Public Policy, Entrepreneurship, and Economic Freedom." *Cato Journal*, Fall 2005, *25*(3), pp. 595-616; www.cato.org/pubs/journal/cj25n3/cj25n3-15.pdf.

Melkumian, Arsen V. "A Gravity Model of Legal Migration to the United States." Working paper, Western Illinois University, 2004.

Mitchell, Daniel J. "The Impact of Government Spending on Economic Growth." *Executive Summary Backgrounder*, No. 1831, March 2005; www.heritage.org/research/reports/2005/03/the-impact-of-government-spending-on-economic-growth.

Nickell, Stephen; Nunziata, Luca and Ochel, Wolfgang. "Unemployment in the OECD Since the 1960s: What Do We Know?" *Economic Journal*, January 2005, *115*(500), pp. 1-27; www.res.org.uk/economic/freearticles/january05.pdf.

Nistor, Adela. "Assessing the Effectiveness of Human Capital Investments on the Regional Unemployment Rate in the United States: 1990 and 2000." *International Regional Science Review*, January 2009, *32*(1), pp. 65-91.

Norton, Seth W. "Poverty, Property Rights and Human Well-Being: A Cross-National Study." *Cato Journal*, Fall 1998, *18*(2), pp. 233-45.

Ovaska, Tomi and Sobel, Russell S. "Entrepreneurship in Post-Socialist Economies." *Journal of Private Enterprise*, Fall 2005, *21*(1), pp. 8-28.

Powell, Benjamin. "Economic Freedom and Growth: The Case of the Celtic Tiger." *Cato Journal*, Winter 2003, *22*(3), pp. 431-48; www.cato.org/pubs/journal/cj22n3/cj22n3-3.pdf.

Puga, Diego. "The Magnitude and Causes of Agglomeration Economies." *Journal of Regional Science*, February 2010, *50*(1), pp. 203-19.

Quah, Danny. "Empirical Cross-Section Dynamics in Economic Growth." *European Economic Review*, April 1993, *37*(2-3), pp. 426-34.

Quan, Nguyen T. and Beck, John H. "Public Education Expenditures and State Economic Growth: Northeast and Sunbelt Regions." *Southern Economic Journal*, October 1987, *54*(2), pp. 361-76.

Sturm, Jan-Egbert and De Haan, Jakob. "How Robust Is the Relationship between Economic Freedom and Economic Growth?" *Applied Economics*, June 2001, *33*(7), pp. 839-44.

Tomljanovich, Marc. "The Role of State Fiscal Policy in State Economic Growth." *Contemporary Economic Policy*, July 2004, *22*(3), pp. 318-30.

Wasylenko, Michael and McGuire, Therese. "Jobs and Taxes: The Effect of Business Climate on States' Employment Growth Rates." *National Tax Journal*, December 1985, *38*(4), pp. 497-511.

Webber, Don J.; White, Paul and Allen, David O. "Income Convergence across U.S. States: An Analysis Using Measures of Concordance and Discordance." *Journal of Regional Science*, August 2005, *45*(3), pp. 565-89.

A Primer on Social Security Systems and Reforms

Craig P. Aubuchon, Juan C. Conesa, and Carlos Garriga

This article reviews the characteristics of different social security systems. Many configurations arise depending on the nature of a system's funding and determination of benefits. Many reforms propose changing the social security systems. The authors focus their analysis of the transition from a pay-as-you-go to a fully funded system. They argue that the key component of any reform is the treatment of the implicit liabilities of a country's social security system. The welfare gains accruing to some cohorts as a result of such reforms usually stem from either a partial or complete default on the implicit debt of the system, and in that sense the gains imply only a redistribution of welfare across agents. In contrast, the elimination of existing distortions in social security financing can generate efficiency gains, allowing for welfare improvements for all agents. This result shifts the focus from the nature of the system itself and centers the debate on the distortions associated with social security. (JEL H2, E62, D31)

Federal Reserve Bank of St. Louis *Review*, January/February 2011, *93*(1), pp. 19-35.

"Social security," by its simplest definition, is a contract between a government and its constituents. Under this contract, citizens provide funding to a social security system, and in exchange they receive benefits from the system during their nonworking years, generally during old age or prolonged illness (disability). The U.S. Social Security system was implemented in 1935 under President Franklin Roosevelt. This system, formally known as Old-Age, Survivors, and Disability Insurance (OASDI), is a pay-as-you-go (PAYG) system in which workers provide financing through a Social Security tax; these contributions provide benefits to the currently retired or disabled. The system requires an implicit guarantee that future generations will then provide the same support for them. In contrast, fully funded (FF) social security systems require that benefits accrue based on individual contributions paid over time. In such systems, future obligations are fully funded by earlier contributions.

Financial sustainability of the U.S. Social Security system is an important policy concern because of the aging population, particularly the baby boom cohorts. The recent recession, combined with the renewed political focus on health care and long-term health costs, has led to increased interest in the long-run financial stability of the U.S. Social Security system. While the Board of Trustees of OASDI has expressed the need for long-term reform for several years, their 2009 report described how lower gross domestic product (GDP) and fewer covered workers affect the long-term outlook (see Board of Trustees, 2009). The Board moved forward its projections for the year in which outlays will exceed revenues (2016) and the year in which

Craig P. Aubuchon was a research analyst at the Federal Reserve Bank of St. Louis. Juan C. Conesa is an associate professor in the department of economics at the Universitat Autònoma de Barcelona, Spain. Carlos Garriga is an economist at the Federal Reserve Bank of St. Louis. The authors thank Ricardo DiCecio, William Gavin, and Christopher Neely for their reviews and comments.

current trust funds will be exhausted (2037). The 2010 Trustees Summary Report concludes that these imbalances "demonstrate the need for timely and effective action. The sooner the solutions are adopted, the more varied and gradual they can be" (see Social Security and Medicare Boards of Trustees, 2010). These solutions include higher taxes, lower benefits, or a combination of both to replenish the trust fund. However, some analysts advocate a transition to an FF Social Security system.

Building on the seminal work of Auerbach and Kotlikoff (1987), several quantitative analyses simulate the transition from a PAYG to an FF system and find substantial efficiency and welfare gains in the long run. However, the gains often come at the expense of the transition generation. For example, Huang, Selahattin, and Sargent (1997) show that partial or full privatization implies large short-run welfare losses that cannot be compensated by the long-run gains. Conesa and Krueger (1999) show that in the presence of uninsurable labor income uncertainty the welfare losses of the initial cohorts are large and constitute a political barrier to potential reforms. In contrast, a large empirical literature argues that the macroeconomic effects of privatizations have been small, particularly with regard to aggregate saving rates.[1]

The objective of this paper is to provide a theoretical framework to illustrate the effects of reforming social security. We use a simple overlapping generations model to demonstrate when a transition from a PAYG system to an FF system can (and cannot) be welfare improving for all households in the economy. The model includes the initial generations alive at the beginning of the reform as well as future generations. We first show that a PAYG system and an FF system can be equivalent by using a simple government transfer mechanism that makes explicit the implicit debt of the PAYG system. This is nothing more than the recognition of the implicit liability of the social security system with future generations. An immediate application of the equivalence result is that it would be straightforward to engineer a Pareto-neutral social security transition. The equivalent FF system produces the same level of welfare, household decisions (i.e., labor supply and consumption), and output. If the reform implies only the recognition of the implicit debt with no welfare effects, how can this be reconciled with the sizable welfare gains noted in the literature? In most simulations of reforms, pensions and social security contributions are eliminated over time in some arbitrary way. Generally, most such exercises imply some partial or complete default on promises (equivalent to a default on the implicit debt of social security), which was the root of the large welfare losses of the transition cohorts. When the government is allowed to default on the implicit liabilities, the welfare of either existing or future cohorts—but not both—can be improved. Therefore, the post-reform welfare gains by some individuals are the result of not honoring these liabilities and not of the reform itself. The equivalence result shows that when the implicit debt is honored, there is no room for welfare improvements in the event of a privatization.

Because intergenerational redistribution alone cannot generate Pareto improvements in a dynamically efficient economy, these improvements are possible if and only if there are distortions in social security financing or in the tax system. In such circumstances, the presence of additional distortions, not the reform itself, is the key to any welfare improvement. Therefore, sizable welfare gains are possible when substantial economic distortions occur, but welfare gains are negligible if the social security contributions and the tax code are close to optimal. These results might explain why the empirical literature has found small macroeconomic effects resulting from many reforms in actual economies. Our exercise suggests that no macroeconomic and welfare effects should be expected unless distortions are minimized or eliminated. Successful reform requires some substance, not just relabeling of government debt.

[1] Coronado (2002) finds that after the 1981 Chilean social security reform (covered in the next section), wealthy families increased aggregate savings by approximately 7 percent. In contrast, Butelman and Gallego (2000) find that low-income Chilean households increased their level of debt. Disney, Emmerson, and Smith (2003) and Granville and Mallick (2002) find no effect on aggregate household saving rates after the 1986 U.K. reforms.

Table 1

Configurations of Social Security Systems

Type of plan	Funded	Unfunded
Defined benefit	Traditional employer pension (example: Switzerland)	United States Australia United Kingdom
Defined contribution	U.S. Roth IRA or 401(k) Chile Latin America Australia United Kingdom	
Notional defined contribution		Italy Sweden

OVERVIEW OF SOCIAL SECURITY SYSTEMS

Not all social security systems are designed the same. Obviously, different countries have defined their social security contracts according to the principles that shape their cultures and economies. In general, there are four broad types of social security systems. These systems combine two elements. First, a system can either be unfunded (such as a PAYG system) or FF (based on accumulated assets). Second, a system can provide payments based on either defined benefits or defined contributions. Note that an FF system is not the same as a privatized system. An FF system is simply a model for savings and usually represents a switch from a defined-benefit to a defined-contribution system.[2] Diamond (1996) states, "I think that the distinction between contribution and benefit base is more illuminating than the distinction between privatization and government-run systems, for various pieces of either type of system can be privatized" (p. 75). Table 1 summarizes the different configurations of social security systems.

In 1994, the World Bank released a comprehensive review, *Averting the Old Age Crisis: Policies to Protect the Old and Promote Growth*, and advocated a three-pillar model to social security. The three pillars have been broadly interpreted as the World Bank model and consist of (i) a publicly managed, unfunded, defined-benefit pillar; (ii) a privately managed, funded, defined-contribution pillar; and (iii) a voluntary savings pillar. Orszag and Stiglitz (1999) point out that the World Bank did not explicitly argue for a privately managed second pillar but that many scholars have interpreted it as such.[3] This three-pillar approach should sound familiar to most Americans: Those who participate in Social Security through payroll taxes, save through a 401(k) plan or individual retirement account (IRA), and manage their own private assets engage in the three pillars advocated by the World Bank.

Many workers are inadequately prepared for retirement because they do not participate in the voluntary second and third pillars. According to the 2007 *Survey of Consumer Finances* (*SCF*), only 60.9 percent of U.S. households 55 to 64 years of age have a retirement account outside Social Security and only 41.6 percent of households 35 years of age and younger have a retirement account in place (see Federal Reserve Board,

[2] Diamond (2004) gives an alternative definition of "defined contribution" and notes that the heart of a defined-contribution system is the fact that the risk to future outcomes is on the side of benefits, which are a function of the realized returns on funded contributions. This is in contrast to the risk of an unfunded system, which generally falls on future taxes.

[3] For a recent example, see the February 2009 testimony of Alicia Munnell and Peter Drucker before the U.S. House of Representatives; they articulate that a strong second pillar of an FF defined-contribution plan will help diversify risk and improve retirement portfolios.

2007).[4] Across all ages, the *SCF* finds that 57.7 percent of families have some rights to a defined-benefit pension or account plan through a current or past employer. Thus, for any combination of reasons, a large majority of families have chosen not to participate in retirement savings outside Social Security, even though Social Security is designed to replace only 40 percent of pre-retirement income for the average worker retiring in 2007 (Social Security Administration, 2008). The differences across systems (see Table 1) are explained next.

Funded Defined-Contribution System

Current private pensions, such as 401(k) plans and Roth IRAs, are a type of funded defined-contribution system. Workers contribute a percentage of their salaries during working years, often with a matching employer contribution up to an established limit. Workers are free to choose the investment of their funds and are eligible to withdraw their savings during their retirement years, with a total sum equal to their defined contribution plus investment earnings. Chile's social security reform in 1981 remains the best-known international example of an FF defined-contribution system; Diamond (1996) presents a survey of the literature examining the pros and cons of this model. Under the Chilean system, all workers are required to contribute 10 percent of their salary into a savings plan of their choice, which is administered and regulated by the Administradora de Fondos de Pensiones. As in the United States, eligibility for retirement is based on age and early retirement is available to those with sufficient accumulated savings. At retirement, workers can choose monthly withdrawals or purchase an annuity. Furthermore, workers are guaranteed a minimum pension paid from the general revenue fund. In Chile, the benefits of such a system include reduced exposure to political and demographic risk since retirement benefits are funded and cannot be reduced through taxes.

Several other countries, including most of Latin America, have a funded defined-contribution pillar that follows Chile's example. Valdés-Prieto (1998) presents a summary of the reforms in Peru, Colombia, Argentina, Bolivia, Mexico, El Salvador, and Uruguay. He offers five reasons why Chile's model is so successful, including low levels of private-sector corruption, little political pressure on investment options, and successful implementation of a redistributive means-tested benefit to workers not covered by the Administradora de Fondos de Pensiones.

Bateman and Piggot (1998) summarize the pension reforms in Australia and the success of its funded defined-contribution pillar, the superannuation guarantee. The Australian system operates under a model similar to Chile's with the exceptions that fund choices for each pension are governed by a board of trustees and assets and allocation are unrestricted. Australia's system is also unique in that taxes are levied at all three possible points—contributions, investment earnings, and benefits. Finally, Australia offers the choice of either a lump-sum payment or an annuity at retirement. The lump-sum payment is unique among funded defined-contribution plans; the authors point out that from a societal perspective, the policy is inefficient because individuals who spend their lump sums must then rely on the state.

The U.K. system also offers a privatized, funded defined-contribution system but a unique one, in that it allows workers to opt out of their public, unfunded defined-benefit system. Indeed, as Johnson (1998) reports, between 1988 and 1992, the United Kingdom offered an additional 2 percent "incentive" rebate to workers switching to a private pension. Thus, an unfunded future benefit is replaced with a currently funded contribution. Johnson notes that younger workers benefit more from switching and, as expected, have done so in large numbers. However, studies have also shown that many workers have opted out of an occupational pension program to join the private pension, in essence giving up any company con-

[4] The 2007 *SCF* does not include annuities in this measure and notes that some "families may have used funds from previous employment to purchase an annuity at retirement" (p. A23). The survey also notes that among older age brackets (55 to 64 years), workers can withdraw funds from some retirement accounts as early as age 59½.

tribution to their plans. Thus, the issue of choice has been effective for most, but some workers have made second-best choices.

Funded Defined-Benefit System

More traditional pensions, similar to those awarded to older U.S. workers during previous decades, are good examples of funded defined-benefit systems. Workers pay into the pension system, and the corporation manages how these contributions are invested. Workers then receive a defined benefit at retirement, which is usually based on years of service or some other related measure.

Switzerland currently offers a hybrid system: a funded defined-contribution system with a guaranteed return. The burden of the Swiss compulsory occupational pension scheme (overseen by a trustee board) is placed directly on individual corporations. The board chooses the pension insurance and the amount and is responsible for enrolling workers. Workers contribute 17 percent of their salary, half of which is matched by the corporation. Under these requirements, the system is purely a defined-contribution plan. However, each pension plan is required to return a minimum of 4 percent nominal interest each year.[5] If a pension plan is ever underfunded, the firm or corporation must make up the difference. This legal, and explicit, guarantee on returns makes the Swiss system more of a defined-benefit plan since workers know with some certainty the value of their future annuity.

Unfunded Defined-Benefit System

A publicly operated, unfunded defined-benefit plan constitutes the first pillar of social security among most countries. These systems are often described as PAYG because current workers pay taxes to provide a benefit to the current retired generation. An individual's benefits are offered either as means-tested, such that a worker receives benefits only if they are below an income threshold or as a universal benefit given to all workers, often calculated as a percentage of the earnings average over a set number of working years. Benefits are also linked to either wage growth or price growth so that benefits stay roughly in line with the cost of living. The U.S. Social Security system is generally considered an unfunded defined-benefit system. Workers pay into the system through a tax, which is then transferred to the current retired generation in the form of a defined benefit. In the United States, the universal benefit is based on a worker's average earnings over a 35-year period, up to a certain income level. Furthermore, under the U.S. system, benefits are also provided to spouses and dependents. Diamond (2004) asserts that the unfunded nature of U.S. Social Security makes sense given the early decision to redistribute wealth and provide full benefits to the initial generation—which did not pay into the system—because the incomplete-funding risk is shared across future cohorts. An upcoming section outlines the model for a PAYG system that generates this type of redistribution.

Unfunded defined-benefit plans are generally useful for a redistribution of wealth—that is, to help guarantee a minimum level of income for any worker who participates in the labor force for an agreed-upon number of years. Unfunded defined-benefit plans often favor workers with lower incomes or with noncontinuous work histories. Often, the benefit is set at a basic subsistence level and is intended to be supplemented with other retirement savings. For example, Switzerland provides a guaranteed minimum pension for its entire population, with the pension paid from general revenues. Hence, the minimum pension is independent of a worker's salary or time in the labor force. This universal benefit redistributes income to poorer workers during their retirement years.

Australia and the United Kingdom, among other countries, have a public, unfunded, defined-benefit system as a first pillar for social security. In particular, the United Kingdom has a two-tier state pension scheme. First, workers are provided a basic state pension, paid at a single flat rate that

[5] The 4 percent rate was chosen to be slightly below the long-run return on Swiss government bonds of 4.5 percent. Hepp (1998) summarizes, "The return guarantee was deemed unambitious enough to avoid frequent funding shortfalls...but explicit enough to enhance the credibility of the system" (p. 536).

offers a subsistence-level income.[6] Most workers are covered by occupational pension schemes, which, in one form or another, predate the state scheme; for workers not covered by the occupational program, the state also offers the State Earning Related Pension Scheme (SERPS). A unique feature of SERPS is that workers can opt out of the plan and the associated National Insurance contribution if their private pension or occupational scheme provides a guaranteed minimum pension equivalent to SERPS. In contrast, Australia offers a means-tested benefit, such that the full-rate pension is equal to 25 percent of male average earnings (40 percent for couples), but this payment is phased out as retirement income and other assets accumulated under the other two pillars increase. Similarly, Sweden offers a means-tested pension for workers with no or low income. This guaranteed pension is financed through general revenue taxes and is independent of workers' notional defined-contribution (NDC) plans.

Unfunded Defined-Contribution System

Sweden and Italy are concrete examples of countries with an unfunded defined-contribution social security system. In recent years, both countries have switched to an NDC plan. The government credits each worker for the taxes he or she and the employer contribute, and then pays a benefit equal to the worker's contributions plus a notional interest rate. Första AP-fonden (AP1, one of the pension funds managing the Swedish system) describes the Swedish income pension system as follows: "The income pension system is of the defined contribution type, meaning that the size of future pension benefits depends on the amount of contributions made and return on the invested capital. The income pension system is also a so-called PAYG system, which means that the pension contributions paid in every month are used to pay current income pension benefits to those who have already retired."[7] However, because of the notional interest rate, Sundén (1998) states that the Swedish plan is "more similar to a defined benefit plan...since the government has to cover its pension liability through annual contributions" (p. 582).

CURRENT STATE OF U.S. SOCIAL SECURITY

The unfunded defined-benefit plan of the United States, known as OASDI, contains two separate parts. The first, and the focus of this paper, is the Old-Age and Survivors Insurance (OASI), which pays monthly benefits to retired workers and their families. The second component, Disability Insurance (DI), pays monthly benefits to disabled workers and their families. The 2009 annual report of the Board of Trustees presents the most-current picture of Social Security in the United States and notes that, in 2008, almost 35 million retired workers and their dependents and another 6 million survivors of deceased workers received benefits. An estimated 162 million workers paid Social Security and payroll taxes. Total benefits paid in 2008 were $615 billion, and the Social Security Trust Fund collected $805 billion, prompting the Trustees to note that "the combined OASI and DI Trust Funds are adequately financed over the next ten years" (see Board of Trustees, 2009, p. 2). However, the Trustees also state very clearly: "The financial condition of the Social Security and Medicare programs remains challenging" and that the current PAYG system "does not satisfy the short range test of financial adequacy."

These long-run problems arise as the baby boom generation begins to retire and reduces the number of covered workers per beneficiary from

[6] Johnson and Stears (1996) point out that the basic state pension is technically a defined-contribution scheme, since the pension is paid only to workers who contribute taxes during 90 percent of their working life and have Class 1 contributions on earnings of at least 52 times the weekly lower earnings limit during each working year. Thus, to qualify for the pension at age 65, a worker must have contributed for 44 years (male) or 39 years (female). However, the authors note that "there are so many special provisions that virtually everyone becomes eligible, no matter what their employment and contribution history..." (p. 1111). These exceptions include a deduction on working years for home responsibility protection, which helps guarantee coverage for stay-at-home caregivers.

[7] More information on the Swedish system is available at the Första AP-fonden website (www.ap1.se/en/Our-mission/The-Swedish-pension-system/).

a historical average of three workers per beneficiary to just two workers per beneficiary. As is well documented elsewhere, the Board of Trustees (2009) currently predicts that income received will exceed benefits payments in 2016. Because of the surplus accrued and interest generated on these savings, the payment of benefits will not be reduced until 2037, when the surplus is predicted to run out. At this point, the Trustees estimate that under their intermediate assumptions, payable benefits will be 76 percent of scheduled benefits and by 2038, tax income will cover 74 percent of scheduled benefits.[8]

The Trustees conclude that for the trust fund to remain solvent throughout the 75-year projection period and pay scheduled benefits at 100 percent, one of three things must happen under business as usual.[9] First, the combined payroll tax could be immediately and permanently increased from 12.4 percent to 14.4 percent. Second, scheduled benefits could be reduced by an amount equal to an immediate and permanent reduction of 13 percent. Or third, a general revenue transfer equivalent to $5.3 trillion in present value terms could be made to the trust fund.

These measures all imply a welfare loss of some type for workers and retirees during the 75-year projection period. An important economic consideration to the Trustees' conclusions is the deadweight loss associated with an increase in the payroll tax. With a higher tax rate the government would collect more revenue unless the higher taxes distort an individual's decision on how much to work. The revenue lost because of a decrease in the number of hours worked (say, from an individual declining overtime hours) is the deadweight loss of the tax. Feldstein and Liebman (2002) note that under standard theory, the deadweight loss of a tax system increases with the square of the marginal tax rate. Over time, the increased deadweight loss makes continued tax increases to fund a demographic shift less desirable since each subsequent tax increase results in a larger deadweight loss.

In the early 1980s, the United States faced a similar Social Security dilemma and pursued a strategy of tax increases and delayed benefits in the form of an increase in the normal retirement age (NRA). The Social Security Amendments of 1983 extended the NRA from age 65 to age 67 for the cohort of workers that turns 62 in 2022. Diamond (1996) provides a summary of the changes, noting that the law did not change the minimum age (62 years) to claim retirement benefits, nor did it extend the age to obtain benefits independent of earnings (70 years). Rather, it simply changed the level of benefits as a function of the age at which they are first claimed. With an NRA of 65, workers can receive 80 percent of their benefits starting at the minimum age, 62. Under the new NRA of 67, workers receive only 70 percent of their scheduled benefit starting at age 62. Thus, extending the NRA by two years is the equivalent of cutting benefits by one-eighth. Diamond also notes that extending the NRA might have unintended consequences since there were no corresponding benefit cuts for early withdrawals for DI. This provides an incentive to apply for DI benefits at the earliest date. The loss from future income of working years and revenue savings from the reduced benefit represents another source of potential welfare loss.

Gramlich (1998) provides a brief overview of the recommendations from the 1994-96 Social Security Advisory Council. This group, which Gramlich chaired, offered three options to address the long-run actuarial soundness of Social Security.[10] Options included

[8] Under intermediate assumptions, the Board estimates a total fertility rate of 2 children per woman, an annual percentage change in productivity for the total U.S. economy of 1.7 percent, an unemployment rate of 5.5 percent, and an inflation rate (measured by the consumer price index) of 2.8 percent. Jeske (2003) notes that slight changes to the Board's assumptions can lead to drastic changes in the long-term forecast of Social Security. It is important to note that these changes run both ways: With slightly higher growth Social Security will face no funding problems, but with slightly lower growth Social Security will be even less likely to meet existing obligations. Jeske (2003) concludes that a PAYG system "therefore implies a substantial amount of risk, contrary to the amount that proponents of social security would admit" (p. 16).

[9] The Trustees also note the impact of the 2007-09 recession on the Social Security system and the "business as usual" scenario, including raising estimates for the projected deficit and modeling lower GDP growth in the upcoming years.

[10] See Pecchenino and Pollard (1998) for a formal theoretical treatment of the three proposals put forth by Gramlich (1998).

- a Maintenance of Benefits plan, with minimal changes in benefit schedules or tax rates but a large portion of trust fund assets invested in equities, with the goal of a higher rate of return to restore actuarial balance;
- a Publicly Held Individual Accounts (IAs) plan to replace the defined-benefit system with a large-scale defined-contribution system, with OASDI as a weaker first pillar that provides a poverty-line flat benefit;
- a Two-Tiered System with Privately Held Individual Accounts plan, which Gramlich termed the "kind and gentle" benefit cut plan. With IAs, high-wage workers would experience slight benefit cuts and workers would be required to contribute to centrally managed investment accounts that convert to real annuities upon retirement.

The next subsection reviews some of the theoretical contributions addressing the issue of a transition from a PAYG to an FF system, hence providing a quantitative evaluation of many of the reforms discussed so far.

Effects of a Transition from a Pay-as-You-Go to a Fully Funded System

The differences between a PAYG and an FF social security system have been studied extensively in the economic literature. Here we review several recent papers that study the welfare implications of the transition. For a more comprehensive and nuanced review of the existing literature on social security reform, see Feldstein and Liebman (2002) and Diamond (2004).

Kotlikoff (1998) considers intergenerational welfare and efficiency in U.S. Social Security reform and advocates a consumption tax to finance the transition from a PAYG system to a defined-contribution personal security system. Under his model, an uncompensated welfare transition results in significant increases to capital stock (36.7 percent), aggregate labor supply (3.7 percent), output (11.2 percent), and real wages (7.1 percent) compared with the baseline model. However, this scenario leads to short-run decreases in aggregate welfare. Using a lump-sum redistribution authority to compensate the initial generation, Kotlikoff finds long-run efficiency and welfare gains above the baseline scenario but below the uncompensated transition. Kotlikoff concludes: "[T]he extent to which privatization results in efficiency gains depends on the ability of future generations to compensate current workers for the loss of consumption as a result of financing the transition" (p. 37). Conesa and Krueger (1999) propose an environment augmented to include uninsurable labor income risk, hence giving a PAYG social security system an additional role as a partial insurance device, and show that the transition offers similar conclusions.

Birkeland and Prescott (2007) consider an overlapping generations model with no population growth in both a PAYG system and an FF system. They compare the models under four demographic scenarios calibrated to match the current United States and a future United States with lower population growth and longer retirements. They note that the PAYG system has little or no explicit debt but does not fully maximize welfare for the society. Early in their paper, the authors challenge the notion of debt, stating "The government debt that a country owes to its citizens is not debt in the usual sense…[it] is a mechanism that facilitates intergenerational borrowing and lending, and is an integral part of a welfare-improving saving-for-retirement system" (p. 2).

Birkeland and Prescott (2007) also note that the Congressional Budget Office estimates that the current implicit guarantees of the PAYG system represent liabilities four times gross national income. These implicit guarantees represent a form of debt for the unborn generation, which lowers lifetime welfare. The authors find that with current U.S. demographic assumptions, welfare under an FF system is 9.2 percent higher than a PAYG system. The FF system has a higher explicit debt-to-gross national income ratio, and individuals work more under the FF system because workers have a higher take-home wage without a Social Security tax.[11] Lifetime con-

[11] Their model considers a population growth rate of 1 percent, an NRA of age 65, and a 20-year retirement. Their future U.S. model considers a country with no population growth, an NRA of age 65, and a 30-year retirement.

sumption also remains higher for a future United States with an FF system, by 5.5 percent. While the authors find that aggregate welfare is better than under an FF system and note that an increase in explicit government debt is not a burden to future generations, they do not consider the welfare of the transition generation.

Similarly, Jeske (2003) considers the transition between social security systems and also finds that in the long run every generation benefits more from an FF system. In that system, individual savings are higher, which in turn increase the aggregate capital stock. The higher capital stock acts as a buffer to the economy in the event of large aggregate economic shocks, which disrupt the PAYG social security system. Jeske (2003) argues that "private savings are more desirable and affordable if both benefits and contributions [to PAYG] are lower" (p. 16). As do Birkeland and Prescott (2007), Jeske similarly acknowledges that social security reform has beneficial long-term effects but that, in the short run, "a large portion of the population will be worse off...[T]he problem of privatization is the unfunded liability to pay for current retirees" (p. 22).

Conesa and Garriga (2008) study the optimal financing of the transition from a PAYG to an FF system. By maximizing over the entire policy space and following an optimal fiscal policy approach, the authors find it is possible to finance such a transition in a Pareto-improving manner for all generations. In their model, the fiscal authority changes the labor income tax over time: first by lowering the labor income tax during the transition generation, issuing government debt to fund existing obligations, and then raising taxes over time to repay debt. Measured as equivalent variation in consumption, the authors find that in the transition from a PAYF to an FF system future newborn generations experience a welfare increase between 3 percent and 8 percent. Such a scheme allows for welfare gains for both actual and future generations, and the key aspect is the reduction of the distortions introduced by the tax system.

None of the above-mentioned papers considers the political ramifications of social security reform, nor do they address the social justice issues underlying the need for and extent of a social security contract between a government and its constituents. Our goal is to show the basic mechanisms by which a shift to an FF system can, in some cases, be welfare improving, even for the transition generation. As Orszag and Stiglitz (1999) point out, initial conditions matter and as such, it is an "issue of whether a *shift* to individual accounts would be socially beneficial" and not an issue of whether or not "in a *tabula rasa* sense, an individual account system would have been preferable to a public defined benefit system in the first place" (p. 5).

A MODEL ECONOMY WITH A PAY-AS-YOU-GO SYSTEM

We follow the framework of Samuelson (1958, 1975) and Diamond (1965) and consider the simplest scenario of a two-period overlapping generations model. Individuals work during the first period of their life and are retired during the second period. Consider the problem faced by an individual born in period t (who will be retired in period t+1). During the working period, individuals provide labor, denoted as ℓ_t, and are paid in return a wage rate per unit of labor, w_t. They consume some goods in the first period, $c_{1,t}$; they pay social security taxes on their wage income, denoted by the tax rate, τ_t; and they can save for the next period, a_{t+1}, where savings today are assets that pay interest in the next period. Hence the budget constraint for these individuals is

(1) $$c_{1,t} + a_{t+1} \leq w_t \ell_t - \tau_t w_t \ell_t .$$

During the retirement period—which is the next period, t+1—individuals do not produce any goods; instead, they merely consume the principal and interest on their private savings and their pension payments. Therefore, their budget constraint is given by

(2) $$c_{2,t+1} \leq \left(1 + r_{t+1}\right) a_{t+1} + p_{t+1},$$

where $c_{2,t+1}$ denotes the consumption during retirement (in period t+1) of the individuals born in period t, r_{t+1} denotes the interest payments collected on savings, and p_{t+1} denotes the social security payments collected by the retired in period t+1.

If there is no exogenous restriction on the sign and magnitude of savings,[12] both budget constraints can be combined as

$$(3)\quad c_{1,t} + \frac{c_{2,t+1}}{1+r_{t+1}} \le (1-\tau_t) w_t \ell_t + \frac{p_{t+1}}{1+r_{t+1}}.$$

This expression simply states that the net present value of consumption over the life cycle of individuals cannot be larger than the net present value of after-tax payments (labor income when individuals are young and pensions when old). Given this constraint, individuals would choose ($c_{1,t}$, $c_{2,t+1}$,ℓ_t) given (τ_t,w_t,p_{t+1},r_{t+1}) and their preferences (i.e., how much they value consumption in the present with respect to consumption in the future and how much they value consumption today with respect to how much they dislike working).

Finally, in this world the government operates the social security system in a standard PAYG fashion: Social security contributions of the current working-age population finance the pension payments of the currently retired population. In particular, if there are (1+n) workers per retired person (think of n as a constant growth rate of population), the social security system would be balanced when

$$(4)\quad (1+n)\tau_t w_t \ell_t = p_t.$$

Notice it is an unfunded system in the following sense. The individuals born in period t will contribute to the system $\tau_t w_t \ell_t$. However, when they retire they will collect pensions p_{t+1} = (1+n)$\tau_{t+1} w_{t+1} \ell_{t+1}$ that are not related to their own past contributions.

A FULLY FUNDED SYSTEM

An alternative to the implicit guarantee of the PAYG structure is an FF system of social security. Under our previous notation, an FF system is a defined-contribution plan similar to a 401(k) program.[13] In the model, workers save throughout their lifetimes. Whether this savings program is mandatory is irrelevant, since workers could also save (or borrow) privately if they chose to do so. Hence, given an optimal consumption allocation if the government increases compulsory savings, individuals would respond by decreasing their private savings by the same amount.

This can be easily seen by consolidating the budget constraints as follows:

$$(5)\quad c_{1,t} + a_{t+1} + a_{t+1}^m \le w_t \ell_t,$$

where $a_{t+1}^m = \tau_t w_t \ell_t$ denotes mandatory contributions to an IRA, computed as a fraction of current labor income. These contributions are then capitalized at the market rate of return and constitute the funding of the future retirement pension.

Hence, next period's budget constraint will include a pension, denoted

$$(6)\quad c_{2,t+1} \le (1+r_{t+1}) a_{t+1} + p_{t+1},$$

where now p_{t+1} = (1+r_{t+1})a_{t+1}^m and therefore the pensions are funded by defined contributions.

Notice, though, that the net present value budget constraint is

$$(7)\quad c_{1,t} + \frac{c_{2,t+1}}{1+r_{t+1}} \le (1-\tau_t) w_t \ell_t + \frac{p_{t+1}}{1+r_{t+1}} = w_t \ell_t.$$

In other words, whether savings are compulsory does not matter; in the end, the net present value of consumption is independent of the level of compulsory contributions to social security. Effectively, then, in this simple model environment an FF system is equivalent to private savings.

COMPARISON OF THE TWO SOCIAL SECURITY SYSTEMS

We now compare the two social security systems. For simplicity, we consider a stationary

[12] The presence, for example, of borrowing constraints might complicate the analysis.

[13] It is important to make the distinction that it is not a necessary condition for FF system funds to be invested in equities. As Orszag and Stiglitz (1999) state, "...prefunding and privatization are distinct concepts, and conflating them confuses rather than informs the debate" (p. 9). Indeed, an FF system could be fully invested in government securities, with low risk and lower (but sometimes guaranteed) returns.

world where social security contributions are constant over time, $\tau_{t+1} = \tau_t$; wages grow at some constant rate g—that is, $w_{t+1} = (1+g)w_t$; and hours worked by each generation of workers are also constant, $\ell_{t+1} = \ell_t$.

Under this scenario, the PAYG social security pensions are given by

$$p_{t+1} = (1+n)\tau_{t+1}w_{t+1}\ell_{t+1} = (1+n)(1+g)\tau_t w_t \ell_t \tag{8}$$

and therefore the return of a PAYG social security system is equal to $(1+n)(1+g)$. Clearly, if $(1+r_{t+1}) > (1+n)(1+g)$, then each individual born in this stationary world will benefit more from a funded system, since the right-hand side of equation (7) is larger than the right-hand side of equation (3).

Usually both empirical data and economic theory tend to confirm that, on average, the return on private investment is larger than the growth rate of a stationary economy, $(1+n)(1+g)$. Notice that a systematic violation of this condition would imply that an economy is inefficiently overaccumulating capital (see Samuelson, 1975).

Clearly, if the return on private investment is larger than the growth rate of the economy, one would fare better born in a world with an FF system than a PAYG system. Nevertheless, as previously discussed, this is the correct answer to the wrong question. The relevant issue relates to the following question: Given that the current world has a PAYG social security system, is anything gained by switching to an FF system? Answering this question requires consideration of the events that occur during the transition, which in our theoretical model exists for just one period.

TRANSITION BETWEEN SOCIAL SECURITY SYSTEMS

Transition I: Default on the Currently Retired

Let us consider the scenario in which our economy is operating under a PAYG structure. Workers in this period contribute to the social security fund by paying a social security tax, under the assumed social contract that they will in turn receive a pension benefit when they retire during the next period. Their pension benefit will then be paid by the young generation of the next period, and so on.

Consider we are now in period T. Imagine that the government decides to switch immediately to a funded system during this period without honoring the implicit debt of the current retirees (the workers of the previous period), that is, $p_T = 0$. The current old generation would incur a welfare loss equal to the sum of the pension obligations to the current retirees. If private savings earn a higher return than the implicit return of the social security system, then the generation working in period T is better off (as are all subsequent generations). However, the contributions of the current workers are invested to finance their own future pensions, $a^m_{T+1} = \tau_t w_t \ell_t$. Effectively, the government has defaulted on its obligations to the currently retired. With this transition scheme the initial cohorts of retired individuals (or those close to retirement) bear the cost at the expense of current young and future generations.

Transition II: Default on Future Generations

Consider now that we decide to keep the promise to the currently retired, and as such, we still need to pay their pensions. We finance these pensions by issuing debt that must be repaid during the next period. The government budget constraint is now $B_{T+1} = p_T$ and the currently young are the only ones who absorb the new debt (in a closed economy) and their budget constraint becomes

$$c_{1,T} + a_{T+1} + \frac{B_{T+1}}{1+n} \leq w_T \ell_T. \tag{9}$$

Notice than per capita debt of the young must be

$$\frac{B_{T+1}}{1+n} = \frac{p_T}{1+n} = \frac{(1+n)\tau_T w_T \ell_T}{1+n} = \tau_T w_T \ell_T. \tag{10}$$

Clearly, the per capita debt of the young now is equal to the social security contributions under the original PAYG system. In the next period, the budget constraint of the retired will be

$$(11)\quad \begin{aligned} c_{2,T+1} &\leq (1+r_{T+1})s_{T+1} + (1+r_{T+1})\frac{B_{T+1}}{1+n} \\ &= (1+r_{T+1})s_{T+1} + (1+r_{T+1})\tau_T w_T \ell_T . \end{aligned}$$

Since the rate of return on private savings (or in government debt) is larger, the transition cohort will benefit more than it would with the PAYG social security system.

However, future generations will experience a welfare loss, which can be seen by examining the budget constraint of the workers born in period T+1:

$$(12)\quad c_{1,T+1} + a_{T+2} + \frac{B_{T+2}}{1+n} \leq w_{T+1}\ell_{T+1} - t_{T+1}.$$

These households will absorb the outstanding debt and will have to pay taxes, denoted by t_{T+1}. The reason can be seen from the government budget constraint:

$$(13)\quad B_{T+2} + T_{T+1} = (1+r_{T+1})B_{T+1},$$

where total taxes collected equal the taxes per worker multiplied by the number of workers; that is, $T_{T+1} = (1+n)t_{T+1}$. Notice that in the absence of this new tax, the outstanding debt would explode to infinity, creating a Ponzi scheme that cannot be in equilibrium.

To maintain a constant level of debt, $B_{T+2} = B_{T+1}$, the new tax should be enough to cover the interest payments on the initial debt issued:

$$(14)\quad \begin{aligned} T_{T+1} &= r_{T+1}B_{T+1} \\ (1+n)t_{T+1} &= r_{T+1}p_T . \end{aligned}$$

Hence households now must pay more in net present value to the tax authority than under the PAYG framework:

(15)

$$\text{NOW:}\quad c_{1,T+1} + \frac{c_{2,T+2}}{1+r_{T+2}} \leq w_{T+1}\ell_{T+1} - t_{T+1}$$

$$\text{PAYG:}\quad c_{1,T+1} + \frac{c_{2,T+2}}{1+r_{T+2}} \leq w_{T+1}\ell_{T+1} - \tau_{T+1}w_{T+1}\ell_{T+1} + \frac{p_{T+2}}{1+r_{T+2}}.$$

Notice that taxes paid now equal the interest payments on the government debt initially issued:

$$(16)\quad t = \frac{rp}{1+n}.$$

In the PAYG regime, total taxes net of discounted pensions paid by households are

$$(17)\quad \tau w\ell - \frac{p}{1+r} = \frac{p}{(1+n)(1+g)} - \frac{p}{1+r},$$

where the equality comes from using expression (8).

Simple algebra shows that expression (16) is larger than expression (17) and, as such, all future generations now fare worse than under the original PAYG system.

Welfare-Neutral Transition: No-Default Case

We now consider an economy that honors the implicit debt of the PAYG system but does not benefit one generation at the expense of others, as was the case in the previous two examples. We follow the approach of Conesa and Garriga (2008) and allow the government to issue recognition bonds equal to the value of the government's implicit pension obligations to current workers.[14]

The no-default plan could proceed as follows: Since current workers are still paying their social security contributions (to honor the benefits of the current retired generation), the government will issue these workers a direct monetary transfer financed by government debt (equivalently, recognition bonds could be issued) equal to the net present value of their (lost) pension in the next period. The transfer received by the current workers is then equal to $p_{T+1}/(1 + r_{T+1})$. By construction, current workers are indifferent between this arrangement and the previous PAYG system.

The budget constraints of the transition generation are now defined as

$$(18)\quad c_{1,T} + a_{T+1} + b_{T+1} \leq w_T\ell_T - \tau_T w_T \ell_T + tr_T$$

$$(19)\quad c_{2,T+1} \leq (1+r_{T+1})a_{T+1} + (1+r_{T+1})b_{T+1}.$$

Notice, however, that these are the same budget constraints as in the PAYG model, once we under-

[14] This idea is not restricted to the particular strategy explored in Conesa and Garriga (2008); the term "recognition bond" has been used elsewhere, including by Feldstein and Liebman (2002). Diamond (1996) also highlights this example from the experience of Chile in 1981.

stand that (i) the transfer collected by workers is equal to debt issued $tr_T = b_{T+1}$ and (ii) principal and interest on the debt (or the recognition bonds) are equal to the pensions $(1 + r_{T+1})b_{T+1} = p_{T+1}$.

In the next period, the total amount of the recognition bonds (or government bonds to pay for the transfer) must be paid with interest. By construction, though, this quantity is exactly equal to the amount of the pensions, so the new generation of workers born in period T+1 must pay taxes and again be compensated by a transfer in the exact manner as the previous generation, and so on until infinity.

The introduction of recognition bonds does not increase the level of debt for our model's government; rather it makes the debt explicit. Pakko (2009) provides a brief overview of the U.S. federal deficit and explains the distinction between "on-budget" and "off-budget" items. Social Security is an off-budget item and is reported only as part of the unified budget. Currently, the Social Security Trust Fund in the United States is counted as a surplus since tax receipts are larger than benefit payments. This surplus appears in reported figures of the combined budget. For example, in 2008 the official measure reported by the government—the unified budget deficit—was $455 billion. The on-budget deficit was $638 billion, with an off-budget surplus of $183 billion, funded primarily from Social Security. Pakko (2009) questions which deficit measure—on-budget or unified budget—the government will report starting in 2017 when Social Security outlays exceed revenues. The implicit debt guarantee to future generations is currently not reported, even though it is politically unlikely that the federal government will default on these future obligations.

Chile approached a similar transition in 1981 by incurring no debt; instead, they began building a fiscal surplus three years before the reform started. Chilean GDP grew at an average of 8 percent per year during this period, and the high growth fueled increased tax revenue. Diamond (1996) states that "it may be that a surplus is a contributing condition for a successful privatization" (p. 80). Valdés-Prieto (1998) acknowledges that a surplus is sufficient, but not a necessary condition, to reform. During a review of the reforms in several Latin American countries in the early and mid-1990s, he finds that even countries emerging from hyperinflation have successfully managed a transition by issuing debt after the inflationary period.

Welfare-Improving Transition: No-Default and Lower Labor Distortions

We now use the intuition and fiscal policy approach from Conesa and Garriga (2008, 2009) to present the case in which a switch to an FF defined-contribution model can be welfare improving. The existing literature has shown that in a dynamically efficient economy, it is not possible to raise aggregate welfare by redistributing resources across generations (a result that goes back to Diamond, 1965), which is basically our approach in the previous reform scenarios. The first two scenarios were situations in which one cohort might benefit at the expense of others; the last one was a Pareto-neutral privatization (nobody won or lost; we simply made the implicit debt explicit). However, if the economy is inefficient because of distortions, then we can increase aggregate welfare by removing the distortions.

Furthermore, we have seen that in the case of a decline in the labor force (or a corresponding increase in the dependency ratio), the two policy options are increasing the payroll tax or cutting benefits. Increasing the payroll tax worsens the distortion on labor.

In the baseline scenario the distortion in the economy comes from the tax on labor, which is used to finance the PAYG system. Pensions are viewed as a pure transfer, while contributions are viewed as a pure tax. Notice that actual pension systems do have some connection between labor income and pension entitlements so that, in reality, individuals may realize the link between their individual contributions and their pension entitlements, thus reducing the distortion. However, insofar as the connection between contributions and pensions is not one to one (because of redistributive considerations usually present in most systems), there will still be a distortion.

Another way to view the distortion is by looking at the equivalent economy in which the implicit liabilities of the PAYG system have been made explicit. From equation (18), we see that workers are paying taxes as a function of their labor income, $\tau_T w_T \ell_T$, and at the same time they are receiving a compensatory transfer, tr_T, independent of their labor supply decisions. Basic economic principles imply a deadweight loss because of this scheme. Moreover, this distortion would increase in the labor supply elasticity. Reducing both the tax and the transfer would result in an efficiency gain.

Most countries have additional distortions that are important in studying social security reforms. For example, mandatory retirement rules could be eliminated during a reform, as suggested by Conesa and Garriga (2003). Without this restriction, the transition to an FF system requires a lower level of compensating transfers and ensures a faster convergence to the new steady state. In addition, the government can change the tax treatment of capital income of retirees as an alternative compensation mechanism. In considering these different distortions, the relevant set of budget constraints becomes

$$(20)\quad c_{1,T} + a_{T+1} + b_{T+1} \le w_T \ell_T - \tau_{1,T} w_T \ell_T + tr_T$$

$$(21)\quad \begin{aligned} c_{2,T+1} \le &\left(1-\tau_{2,T+1}\right) w_{T+1} \ell_{T+1} \left(1 + r_{T+1}\left(1-\theta_{T+1}\right)\right) a_{T+1} \\ &+ \left(1 + r_{T+1}\right) b_{T+1}. \end{aligned}$$

In addition to the implicit debt of the social security system, a larger set of distortions has been made explicit. Distortions on labor income of the young are now denoted by $\tau_{1,T}$, while the distortion on labor income of the old is $\tau_{2,T+1}$ (with compulsory retirement this is equal to 100 percent), and θ_{T+1} denotes distortions on investment decisions.

Given this scenario, it is possible for a government to reduce the distortions and generate welfare improvements, implying a lower level of compensatory transfers or a lower level of recognition bonds. Such a strategy is illustrated in Conesa and Garriga (2008, 2009). They show that the optimal social security reform consists of providing compensatory transfers to the initial old generation (transfers almost as large as their social security entitlements) financed with debt and lowering labor income taxes on impact to increase them later.

The introduction of capital income taxes in the analysis allows for the generation of additional welfare gains since it drastically reduces the need for compensatory transfers for the initial generations alive. On average, capital income taxes translate into very large subsidies, especially for the oldest cohorts. Effectively, changing the fiscal treatment of capital income can become a close substitute for compensatory lump-sum transfers to the initial old generation.

CONCLUSION AND POLICY IMPLICATIONS

We have examined some of the "myths" surrounding social security as presented by Orsagz and Stiglitz (1999), particularly by focusing on comparing the transition between systems as opposed to *tabula rasa* comparisons between a PAYG and an FF system. We have presented the findings of other researchers who have documented the welfare gain under an FF defined-contribution system without considering the welfare cost for the transition generation. We build on the notion that a PAYG social security system is just an implicit liability for the tax authority, and as such it could be converted into an explicit liability (i.e., government debt) without cost. After such conversion the government can focus on designing reforms without inevitably generating welfare losses for some generations. The key insight is that this scenario is possible only if the distortions (introduced either by the financing of social security or other types) are reduced.

Hence, the focus should be shifted from the nature of the social security system itself and the debate centered instead on the distortions introduced by all tax-transfer schemes currently present in the economy. The debate on social security reform then becomes a debate on how to allocate to different cohorts the efficiency gains generated by the reduction of these distortions.

REFERENCES

Auerbach, Alan J. and Kotlikoff, Laurence J. “Evaluating Fiscal Policy with a Dynamic Simulation Model.” *American Economic Review*, May 1987, *77*(2), pp. 49-55.

Bateman, Hazel and Piggott, John. “Mandatory Retirement Saving in Australia.” *Annals of Public and Cooperative Economics*, December 1998, *69*(4), pp. 547-69.

Birkeland, Kathryn and Prescott, Edward C. “On the Needed Quantity of Government Debt.” Federal Reserve Bank of Minneapolis *Quarterly Review*, November 2007, *31*(1), pp. 2-15; www.minneapolisfed.org/research/QR/QR3111.pdf.

Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. *The 2009 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*. Washington, DC: U.S. Government Printing Office, 2009; www.socialsecurity.gov/OACT/TR/2009/.

Butelmann, Andrea and Gallego, Francisco. “Household Savings in Chile: Microeconomic Evidence.” Working Paper No. 63, Central Bank of Chile, February 2000.

Conesa, Juan C. and Garriga, Carlos. “The Status Quo Problem in Social Security Reforms.” *Macroeconomic Dynamics*, November 2003, *7*(6), pp. 691-710.

Conesa, Juan C. and Garriga, Carlos. “Optimal Fiscal Policy in the Design of Social Security Reforms.” *International Economic Review*, February 2008, *49*(1), pp. 291-318.

Conesa, Juan C. and Garriga, Carlos. “Optimal Response to a Transitory Demographic Shock in Social Security Financing,” in Robert Fenge, Georges de Ménil, and Pierre Pestieau, eds., *Pension Strategies in Europe and the United States*. Cambridge, MA: MIT Press, pp. 98-113. Reprinted in Federal Reserve Bank of St. Louis *Review*, January/February 2009, *91*(1), pp. 33-48; http://research.stlouisfed.org/publications/review/09/01/Conesa.pdf.

Conesa, Juan C. and Krueger, Dirk. “Social Security Reform with Heterogeneous Agents.” *Review of Economic Dynamics*, October 1999, *2*(4), pp. 757-95.

Coronado, Julia Lynn. “The Effects of Social Security Privatization on Household Saving: Evidence from the Chilean Experience.” *Contributions to Economic Analysis and Policy*, 2002, *1*(1), Article 7, pp. 152-75.

Diamond, Peter A. “National Debt in a Neoclassical Growth Model.” *American Economic Review*, December 1965, *55*(5 Part 1), pp. 1126-50.

Diamond, Peter A. “Proposals to Restructure Social Security.” *Journal of Economic Perspectives*, Summer 1996, *10*(3), pp. 67-88.

Diamond, Peter A. “Social Security.” *American Economic Review*, March 2004, *94*(1), pp. 1-24.

Disney, Richard; Emmerson, Carl and Smith, Sarah. “Pension Reform and Economic Performance in Britain in the 1980s and 1990s.” NBER Working Paper, No. 9556, National Bureau of Economic Research, March 2003; www.nber.org/papers/w9556.pdf?new_window=1.

Drucker, Peter F. and Munnell, Alicia H. "The Financial Crisis and Restoring Retirement Security." Testimony before the Committee on Education and Labor. U.S. House of Representatives, February 24, 2009; http://edlabor.house.gov/documents/111/pdf/testimony/20090224AliciaMunnellTestimony.pdf.

Federal Reserve Board. *2007 Survey of Consumer Finances*. www.federalreserve.gov/pubs/oss/oss2/2007/scf2007home.html.

Feldstein, Martin. "Would Privatizing Social Security Raise Economic Welfare?" NBER Working Paper No. 5281, National Bureau of Economic Research, September 1995; www.nber.org/papers/w5281.

Feldstein, Martin. *Privatizing Social Security*. Chicago: University of Chicago Press, 1998.

Feldstein, Martin and Liebman, Jeffrey B. "Social Security," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*. Volume 4, Part 6. Amsterdam: Elsevier Science, 2002, pp. 2245-324.

Gramlich, Edward M. "Mending But Not Ending Social Security: The Individual Accounts Plan." Federal Reserve Bank of St. Louis *Review*, March/April 1998, *80*(2), pp. 7-10; http://research.stlouisfed.org/publications/review/98/03/9803eg.pdf.

Granville, Brigitte and Mallick, Sushanta. "Does Capital Market Reform Boost Savings? Evidence from the UK Pension Reforms." Working paper, Royal Institute of International Affairs, December 2002; www.chathamhouse.org.uk/files/3065_pension.pdf.

Hepp, Stefan. "Mandatory Occupational Pension Schemes in Switzerland: The First Ten Years." *Annals of Public and Cooperative Economics*, December 1998, *69*(4), pp. 533-45.

Huang, He; Selahattin, Imrohoroglu and Sargent, Thomas J. "Two Computations to Fund Social Security." *Macroeconomic Dynamics*, January 1997, *1*(1), pp. 7-44.

Jeske, Karsten. "Pension Systems and Aggregate Shocks." Federal Reserve Bank of Atlanta *Economic Review*, First Quarter 2003, *88*(1), pp. 15-31; www.frbatlanta.org/filelegacydocs/erq103_jeske.pdf.

Johnson, Paul. "The Reform of Pensions in the UK." *Annals of Public and Cooperative Economics*, December 1998, *69*(4), pp. 517-32.

Johnson, Paul and Stears, Gary. "Should the Basic State Pension Be a Contributory Benefit?" *Fiscal Studies*, February 1996, *17*(1), pp. 105-12.

Kotlikoff, Laurence J. "Privatizing U.S. Social Security: Some Possible Effects on Inter-generational Equity and the Economy." Federal Reserve Bank of St. Louis *Review*, March/April 1998, *80*(2), pp. 31-37; http://research.stlouisfed.org/publications/review/98/03/9803lk2.pdf.

Orszag, Peter R. and Stiglitz, Joseph E. "Rethinking Pension Reform: Ten Myths About Social Security Systems." Presented at the World Bank Conference, "New Ideas About Old Age Security," September 14-15, 1999; www.iza.org/de/calls_conferences/pensionref_pdf/panel_stiglitz.pdf.

Pakko, Michael. "Deficits, Debt and Looming Disaster: Reform of Entitlement Programs May Be the Only Hope." Federal Reserve Bank of St. Louis *Regional Economist*, January 2009, pp. 4-9; http://stlouisfed.org/publications/re/2009/a/pdf/debts.pdf.

Pecchenino, Rowena A. and Pollard, Patricia S. "Reforming Social Security: A Welfare Analysis." Federal Reserve Bank of St. Louis *Review*, March/April 1998, *80*(2), pp. 19-30.

Samuelson, Paul A. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy*, December 1958, *66*(6), pp. 467-82.

Samuelson, Paul A. "Optimal Social Security in a Life-Cycle Growth Model." *International Economic Review*, October 1975, *16*, pp. 539-44.

Social Security Administration. "The Future of Social Security." SSA Publication No. 05-10055, 2008.

Sundén, Annika. "The Swedish NDC Pension Reform." *Annals of Public and Cooperative Economics*, December 1998, *69*(4), pp. 571-83.

Rangel, Antonio. "Forward and Backward Intergenerational Goods: Why Is Social Security Good for the Environment?" *American Economic Review*, June 2003, *93*(3), pp. 813-34.

Social Security and Medicare Boards of Trustees. "Status of the Social Security and Medicare Programs." Social Security Administration Actuarial Publications, 2010; www.ssa.gov/OACT/TRSUM/index.html.

Valdés-Prieto, Salvador. "The Latin American Experience with Pension Reform." *Annals of Public and Cooperative Economics*, December 1998, *69*(4), pp. 483-516.

World Bank. *Averting the Old Age Crisis: Policies to Protect the Old and Promote Growth*. New York: Oxford University Press, 1994; www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1994/09/01/000009265_3970311123336/Rendered/PDF/multi_page.pdf.

What Explains the Growth in Commodity Derivatives?

Parantap Basu and William T. Gavin

This article documents the massive increase in trading in commodity derivatives over the past decade—growth that far outstrips the growth in commodity production and the need for derivatives to hedge risk by commercial producers and users of commodities. During the past decade, many institutional portfolio managers added commodity derivatives as an asset class to their portfolios. This addition was part of a larger shift in portfolio strategy away from traditional equity investment and toward derivatives based on assets such as real estate and commodities. Institutional investors' use of commodity futures to hedge against stock market risk is a relatively recent phenomenon. Trading in commodity derivatives also increased along with the rapid expansion of trading in all derivative markets. This trading was directly related to the search for higher yields in a low interest rate environment. The growth was both in organized exchanges and over-the-counter (OTC) trading, but the gross market value of OTC trading was an order of magnitude greater. This growth is important to note because a critical factor in the recent crisis was counterparty failure in OTC trading of mortgage derivatives. (JEL G120, G130, G180)

Federal Reserve Bank of St. Louis *Review*, January/February 2011, *93*(1), pp. 37-48.

The recent financial crisis was caused by large financial firms taking on too much risk (leverage) using complicated instruments in opaque trading environments.[1] Commodity derivatives trading was one such area. Commodity derivatives include futures and options traded on organized exchanges as well as the forwards and options traded over the counter. Organized exchanges monitor trading of standardized contracts and require margin accounts that protect investors against counterparty risk. The exchange is the counterparty in all trades. Over-the-counter (OTC) trades are bilateral exchanges of customized contracts. Margins are not required and such trading has not been monitored. On July 21, 2010, President Obama signed the Dodd-Frank Wall Street Reform and Consumer Protection Act into law. As of this writing, the regulatory rules have yet to be finalized, but the proposed regulations are intended to limit the use of derivatives by banks and make OTC trading more transparent.

The market failure that led to the recent financial crisis was centered in the opaque, bilateral OTC trading by firms that policymakers at the Federal Reserve and the Treasury considered too big to fail. Because of the potential risks involved, it is important to understand mechanisms that large financial firms can use to exploit the government's safety net. In this article, we document the massive increase in trading in commodity derivatives over the past decade. This growth far

[1] See remarks by Gensler (2010).

Parantap Basu is a professor of macroeconomics at Durham University, Durham, U.K. William T. Gavin is a vice president and economist at the Federal Reserve Bank of St. Louis. The authors thank Kevin Kliesen and Kenji Wada for helpful comments.

outstrips the growth in commodity production and the need for derivatives to hedge risk by commercial producers and users of commodities.

During the past decade, many institutional portfolio managers added commodity derivatives as an asset class to their portfolios. This addition resulted in substantial growth in the use of commodity derivatives—growth out of proportion with the historical levels associated with commercial hedging. This shift was part of a larger change in portfolio strategy away from traditional equity investment and toward derivatives based on assets such as real estate and commodities.

Trading in derivatives does not affect the fundamentals of supply and demand in any obvious way. The derivative trades sum to zero—for every winner there is a loser, for every gain there is an equal loss. Financial firms can write an arbitrarily large number of contracts betting on a future price without necessarily affecting the level of that price. However, an arbitrarily large number of contracts means that there can be an arbitrarily large number of losers. The important policy question is whether the taxpayer is at risk for counterparty failure in OTC trading when some financial firms incur large losses. If a large portion of these investments is made by financial firms that would likely fall under the protection of the government's safety net, then the firms that win will retain their profits while those that lose may shift the burden of their losses to the taxpayer. There is a public interest in preventing large-scale betting by institutions protected by the government's safety net. It is not a zero-sum game for the taxpayer.

In this article, we explore the reasons for the explosive growth in trading in commodity derivatives and advance two main reasons for that growth. First, investors used commodity futures to hedge against equity risk. Both academic and industry economists argued that a negative correlation between returns on equity and commodity futures offered an unexploited hedging opportunity in using commodity derivatives as an asset class.

Second, trading in commodity derivatives increased along with the rapid expansion of trading in all derivative markets. This trading was directly related to the search for higher yields in a low interest rate environment. The search for higher yields refers to the tendency of both individual and institutional investors to choose riskier assets when the return on safe assets is low.[2] Jiménez et al. (2008) used a large dataset from the credit register in Spain to show that bank borrowers are more likely to default if the loans are made when central bank interest rates are relatively low. They also showed that (i) the price of risk tends to be low when short-term interest rates are low and (ii) if the interest rate is low for a long time, the economy's "portfolio" of loans tends to be riskier.

Many derivative instruments that grew rapidly after 2000, such as commodity futures index funds and derivatives on mortgage-backed securities (MBS) such as collateralized debt obligations, were developed in the 1980s and 1990s. Dybvig and Marshall (1997) described the newly developed risk-management processes that included ever more-complex derivatives. Their description noted the possibility of the good, the bad, and the ugly outcomes of using such financial instruments. The good is the new opportunity for more-precise hedging and risk reduction.[3] The bad is the possibility that CEOs and portfolio managers may not fully understand the ramifications of using these complex new instruments. The ugly is the possibility that firms could use OTC derivatives to intentionally take risks that could not be observed by regulators or other market participants. All three outcomes have been evident over the past decade, but it is the ugly outcome that is most responsible for the worldwide financial crisis.

The paper is organized as follows. The second section documents some facts about growth in commodity futures and provides indirect evidence that the rise in derivatives trading was associated with institutional investors using commodity derivatives as an asset class. The third section advances arguments why a negative correlation

---

[2] See, for example, Rajan (2005), Ferguson et al. (2007), and Gerlach et al. (2009).

[3] See Banerji and Basu (2009) for an example showing how banks could use new and creative contracts to offer new risk-bearing services that would be expected to reduce the risk premium in equity markets.

between stock and futures returns may not necessarily offer a hedging opportunity to investors. The concluding section discusses the reform legislation and prospects for continued trading in commodity derivatives.

TRADING IN COMMODITY DERIVATIVES: THE FACTS

The large increase in trading in commodity derivatives was not due to a large increase in hedging by commercial users. It is important to distinguish between the commercial hedgers who produce and use commodities and the institutional investors who use commodity futures to hedge equity and bond risk. For example, commodity futures index funds were marketed to institutional investors as an asset class. Figure 1A depicts the growth of these funds using year-end data for 1994 to 2008. Contracts for these funds are an investment in a long position in a value-weighted portfolio of commodity futures. In 2002, there were fewer than $20 billion in these index-fund contracts. At year-end 2008 these funds had grown to more than $250 billion, about one-fourth to one-third of the notional amounts of commodity futures traded on organized exchanges. In 2007 the Commodity Futures Trading Commission (CFTC) began collecting information on the amount of funds invested in these index funds. Figure 1B reports the CFTC data through September 2010. Note that the exchange trading of commodity futures has rebounded and has nearly recovered to the peak achieved in June of 2008.

Trading in OTC commodity derivatives markets also grew rapidly during the period, as shown by the gross market value of commodity derivatives (Figure 2A). Gross market value is a measure of the funds that investors have at risk on both sides of the bet; for example, it includes funds at risk on both the long and short sides of a forward contract. Figure 2A also depicts the gross market value of equity derivatives contracts. The gross market value of commodity derivatives rose by a factor of 25 between June 2003 and June 2008—reaching $2.13 trillion in June 2008. Figure 2B shows the gross market values of commodity derivatives (excluding precious metals) and gold derivatives.[4] Traditionally, institutional investors have used gold as a hedge against inflation and other risks. There was no surge in the volume of gold derivatives as there was for other commodities.

Figure 3 shows prices for the Standard and Poor's (S&P) Goldman Sachs Commodity Index (GSCI), gold, and two ABX indexes that are for derivatives on insurance contracts for MBS.[5] From the day the S&P GSCI peaked, July 3, 2008, to the day Lehman Brothers filed bankruptcy, September 15, 2008, the S&P GSCI price index fell 37 percent (Figure 3).[6] Investors with a short position made large profits, but investors with a long position lost hundreds of billions of dollars. These were investments traded over the counter, so it is difficult to know what part, if any, these losses played in the financial panic that accompanied Lehman's default.

Oil was about 40 percent of the weight in the S&P GSCI and drove the broad pattern in the S&P GSCI. The commodity price index (see Figure 3) rose very sharply with the trading volume of the commodity derivatives market (see Figures 2A and 2B) and peaked in July 2008 when oil prices peaked. It then fell sharply through the second half of 2008. The gold price was much less volatile (see Figure 3), with no unusual rise in the trading volume of gold derivatives (see Figure 2B). Note that the gold price and the commodity price index rose together until mid-March 2008 (see Figure 3), when the Federal Reserve rescued the counterparties to Bear Stearns. The commodity price index (see Figure 3) and trading volume of commodity derivatives then grew very rapidly while the trading volume of gold derivatives was flat to down a bit (see Figure 2B). The commodity price index started falling 10 weeks before financial markets panicked with the Lehman bank-

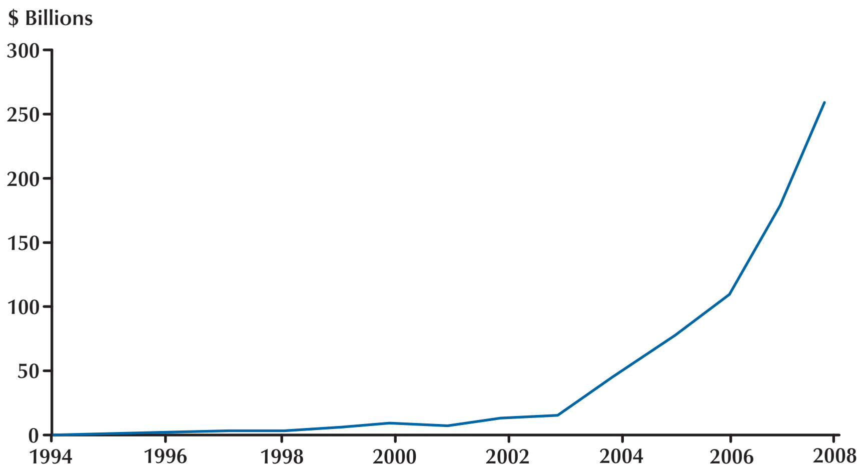
[4] Non-gold precious metals were a small percentage relative to gold and are ignored here.

[5] The gold price is a monthly average of the London PM fix; the source for all prices is Haver Analytics.

[6] We assume that the S&P GSCI represents the market price for the underlying asset in the OTC commodity contracts.

Figure 1A

Commodity Index-Fund Investment (year-end)

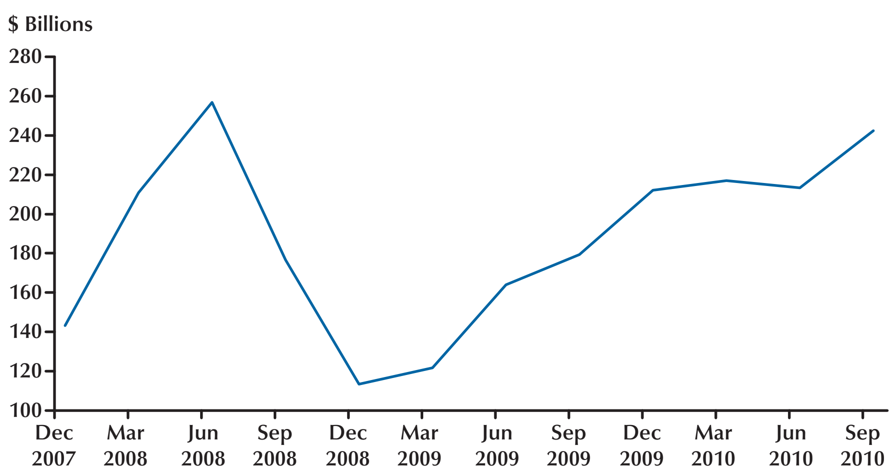


NOTE: 2008 data are through March only.

SOURCE: Masters (2008, Chart 1).

Figure 1B

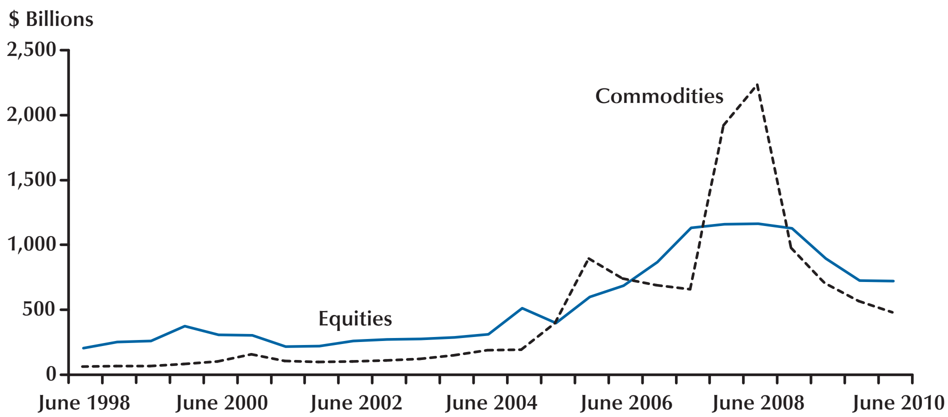
Notional Long Positions Invested in Commodity Futures Index Funds



SOURCE: CFTC.

Figure 2A

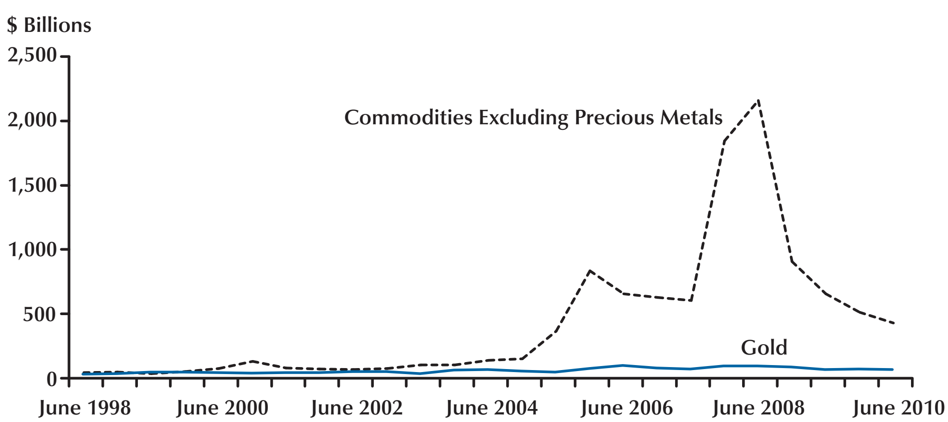
OTC Trading in Commodity and Equity Derivatives (gross market value)



SOURCE: Bank for International Settlements derivatives statistics.

Figure 2B

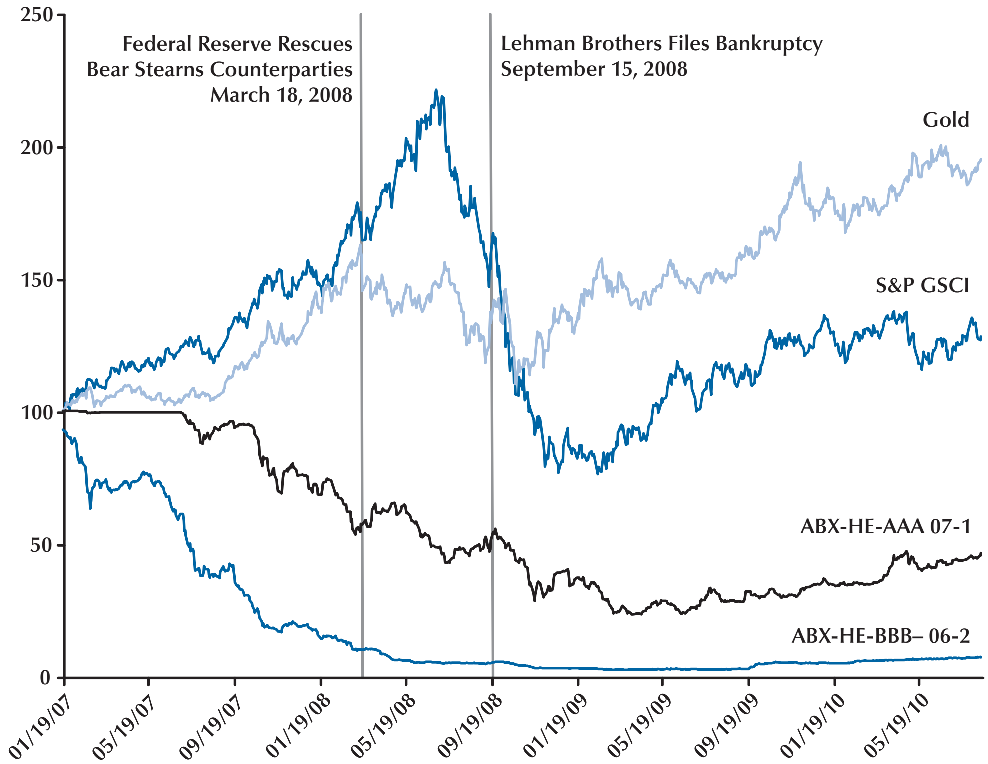
OTC Trading in Commodity Derivatives (gross market value)



SOURCE: Bank for International Settlements derivatives statistics.

Figure 3

Prices on Commodities and ABX Contracts



NOTE: The gold index and the S&P GSCI are each equal to 100 on January 19, 2007. The indexes for the price of ABX contracts are equal to 100 when they are launched. AAA and BBB– are the highest- and lowest-rated mortgage derivatives, respectively. The ABX-HE-BBB– 06-2 is an index for the value of mortgage loan insurance derivatives on the riskiest of the subprime mortgages originated before the second half of 2006. The launch date for an index depends on when the underlying mortgages were originated; the 06-2 index was launched on July 19, 2006, and the 07-1 index on January 19, 2007.

SOURCE: Haver Analytics.

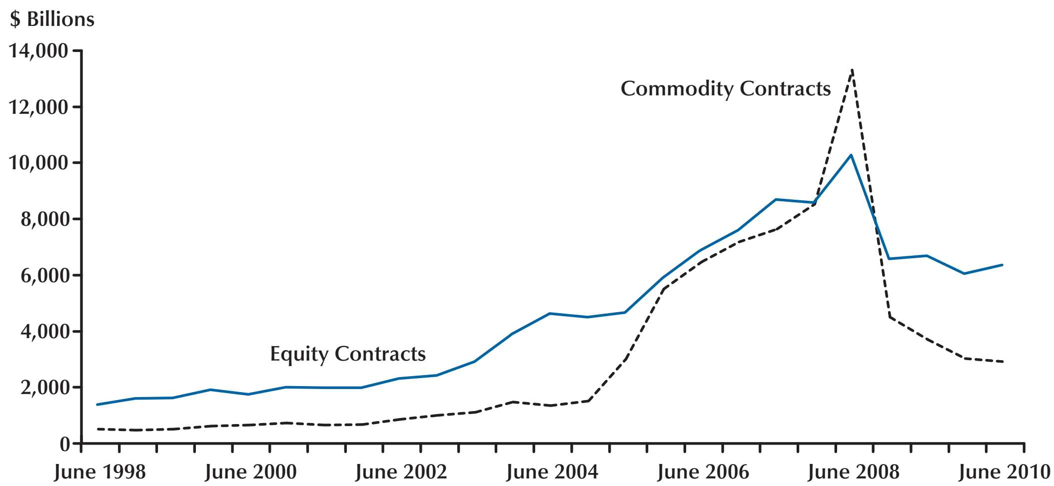
ruptcy filing. The fourth quarter of 2008 was very bad for the economy and financial markets. After year-end, the prices of gold and other commodities as measured by the S&P GSCI began an upward trend that continued through to December 2010.

It is possible that the unusual spike in prices and trading volume for commodity futures was influenced by the loss of confidence in MBS and associated derivatives. Figure 3 shows the loss of confidence in both the highest-rated (AAA) and lowest-rated (BBB–) mortgage derivatives. The ABX BBB– index—for derivatives on mortgage insurance for subprime MBS—began to decline in December 2006 and had fallen 60 percent by August 2007 when the possibility of a wider financial crisis became apparent. By that time, confidence in the highest-rated mortgage paper was also falling. The prices and trading volume of commodity derivatives rose sharply as confidence in the market for subprime mortgages collapsed.

The sharp spike in the price of commodity futures in July 2008 and subsequent collapse by the end of that year is hard to explain. The S&P GSCI was driven mainly by oil prices. Although the longer-term rise in oil prices is often attributed to rising demand associated with growth

Figure 4

OTC Trading in Derivatives (notional amounts)



SOURCE: Bank for International Settlements derivatives statistics.

in emerging market economies, a secular rise in demand cannot explain the 2008 boom and bust.[7]

Figure 4 shows the outstanding notional amounts of commodity derivatives contracts (their face value): The amount tripled between June 1998 and June 2003 and then rose 19-fold in the next 5 years, peaking at $13 trillion in June 2008. During this period, trading in commodity derivatives grew to exceed trading in equity derivatives. Note that, in contrast to trading on organized exchanges, OTC trading in commodity derivatives has continued to decline since the summer of 2008.

To provide some perspective on the size of derivative positions, consider that world GDP rose from $30 trillion in 1998 to $61.1 trillion in 2008.[8] Commodity prices almost quadrupled over the decade before their peak in July 2008. Even at 2008 prices, the total output of commodities was less than half the notional value of outstanding commodity derivatives contracts (nearly $13 trillion).[9] The ratio of the notional amount of commodity derivatives contracts in June 1998 to world GDP rose from 1.5 percent in 1998 to 21.6 percent in 2008. Over the same period, the ratio of equity derivatives to world GDP rose from 4.2 percent to 16.7. At first glance, this shift appears to be consistent with the rising use of commodity derivatives as an asset class in institutional portfolios.

TWO EXPLANATIONS FOR THE RISE IN COMMODITY DERIVATIVES TRADING

One explanation for the rise in commodity derivatives trading is that it was simply part of a widespread increase in risky investing during the past decade that was attributed to a "search for yield." A second explanation for the rise is

[7] See, for example, Kilian (2009).

[8] We are using World Bank estimates of world gross domestic product (GDP) in U.S. dollars.

[9] Even at its peak price in July 2008, total world production of oil in 2008 was less than $4.5 trillion. Oil constitutes the largest share of total commodity production. For example, the estimated worldwide production of corn, wheat, and soybeans in 2009 was less than $100 billion. See, for example, www.nue.okstate.edu/Crop_Information/World_Wheat_Production.htm.

that it was driven by a mistaken notion that an investment in commodity futures can be used to hedge equity risk. An early paper by Greer (2000) and later papers by Erb and Harvey (2006) and Gorton and Rouwenhorst (2006) found a negative correlation between returns to a passive long investment in commodity futures and returns to equity.

The Search for Yield Hypothesis

The term "search for yield" is somewhat vague. In an efficient market model, all investors are assumed to optimize over combinations of risk and return. One should not choose more risk unless the expected returns also rise. One way to interpret the search for yield is to argue that, at low interest rates, investors are willing to take on relatively more risk for only small increases in return. In such a case, investors will bid up the price of risky assets and, all else equal (including default probabilities), the price of risk will decline. This search for yield may explain why risk premiums were so low in 2003 and 2004 and offers one reason (among many) for the high leverage in household mortgages and financial institutions.

During the period of rapid growth in commodity derivatives, managers of pension funds, university endowment funds, and other institutional funds began to include commodity derivatives as an asset class in their portfolios. There was a shift out of domestic equities into commodities.[10] One argument was that investing in such real assets could increase returns without adding much risk. This leads us to the second hypothesis: Brokers and dealers selling commodity derivatives also argued that commodity futures could be used to hedge equity risk.

Hedging Hypothesis

> Fully collateralized commodity futures historically have offered the same return and Sharpe ratio as U.S. equities. Although the risk premium on commodity futures is essentially the same as that on equities for the study period, commodity futures returns are negatively correlated with equity returns and bond returns. The negative correlation is the result, primarily, of commodity futures' different behavior over a business cycle (Gorton and Rouwenhorst, 2006, p. 47).

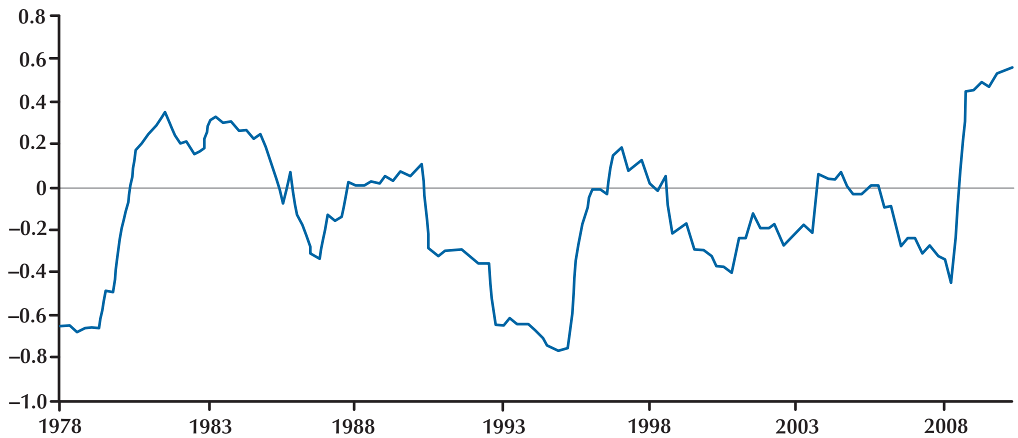
While the use of commodities to hedge inflation risk was widely appreciated, their use to hedge equity or business cycle risk is more controversial. Using data from July 1959 to December 2004, Gorton and Rouwenhorst (2006) calculated the return to holding a rolling long investment in a value-weighted portfolio of commodity futures. They reported that the correlation was nearly zero for short horizons and negative, but not statistically significant, for horizons up to one year. This is consistent with research at the CFTC by Büyükşahin, Haigh, and Robe (2008), who found that the unconditional correlation between equity and commodity futures returns is near zero. But their results changed as the investment horizon lengthened. Gorton and Rouwenhorst (2006) also reported that if this investment was rolled-over for a longer period, the return was negatively correlated with the returns from comparable bond and equity portfolios. They found that the average correlation between returns on equities and commodity futures was a statistically significant –0.42 if the investments were held for 5 years.

Figure 5 reports a rolling 5-year correlation between returns on an index of S&P 500 equities and the index of commodities included in the S&P GSCI. When commodity prices peak in June 2008, the correlation is negative on average. However, following the collapse of commodity prices in the summer of 2008 and the subsequent financial panic in September 2008, the correlation becomes highly positive, reaching a record 0.56 in February 2010. Thus, portfolios that included commodity derivatives to hedge equity risk did very badly over the last 2 two years studied. In the years building up to the crisis and since, portfolios that included commodity derivatives were more volatile than equities-only portfolios. The high returns in 2004 through 2006 reflected very risky investments—not only those in mortgage derivatives. Note that this is the first business cycle following the widespread adoption of this new investment strategy.

[10] See Cohn and Symonds (2004), Symonds (2004), and Palmeri (2006).

Figure 5

Rolling Correlation Coefficient Between Daily Equity and Commodity Returns (5-year window with 5-year returns)

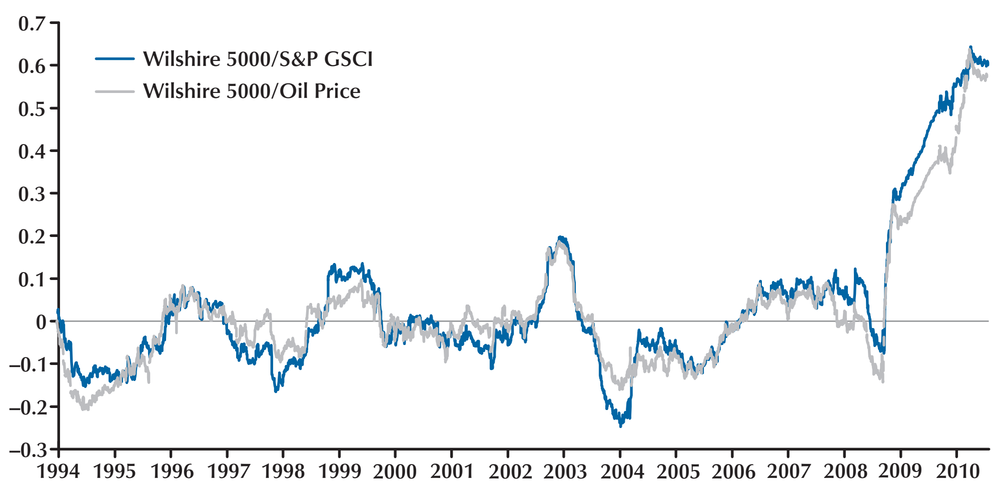


NOTE: The equity index is the S&P 500. The commodity index is the S&P GCSI. The returns are calculated as the percent change in the total return.

SOURCE: Haver Analytics and authors' calculations.

Figure 6

Rolling Correlation Coefficient Between Daily Equity and Commodity Returns (1-year window)



NOTE: The equity index is the Wilshire 5000. The commodity index is the S&P GCSI. The oil price is the domestic spot price on West Texas Intermediate crude oil.

SOURCE: Haver Analytics and authors' calculations.

Similar changes are seen in the correlation of daily returns. Figure 6 reports a rolling correlation coefficient between total returns to investments in the Wilshire 5000 and the S&P GSCI using a 1-year window. The correlation is relatively small and generally not significantly different from zero until the onset of the financial crisis. During and following the crisis, the correlation is very large and positive. Because the S&P GSCI is heavily weighted in oil, we also show the daily correlation between the Wilshire 5000 and the daily spot price of West Texas Intermediate crude oil. This correlation makes it clear that the S&P GSCI is heavily influenced by the oil market.[11]

Erb and Harvey (2006) argued that the most important source of expected return from a portfolio of commodity futures comes from diversification across individual commodities that have uncorrelated returns. They described the different schemes used to construct weights to aggregate the component commodities and explained why the excess returns depend on there being little correlation among returns for the individual component commodities. They also warned against assuming that historical return correlations will persist. Tang and Xiong (2010) showed that the introduction of index trading led to a rise in the correlation among the individual commodities included in an index, thus reducing or even eliminating the gains to diversification within individual index funds. They further showed that the rise in the correlations among the individual components began in 2004, well before the onset of the crisis, and became higher over the next few years as open interest in commodity index futures rose.

Figures 5 and 6 show that the correlation between returns to equity and commodity futures can change sign over time. In a general equilibrium model in which there are no unexploited hedging opportunities, it is straightforward to show that the equilibrium correlation can be either negative or positive, depending on the nature of shocks to the world economy.[12] In particular, the correlations shown in Figures 5 and 6 depend on investors' perceptions about how the domestic economy and commodity production will respond to various shocks.

CONCLUSION

We offer two possible explanations for the surge in trading commodity derivatives. The first also explains the massive increase in trading of risky mortgage debt and all financial derivatives: Investors were searching for more substantial yields in an environment with very low returns paid on safe assets. This also explains why investors moved from real estate derivatives to commodity derivatives when the problems in the subprime market became apparent.

The second reason is a prevailing notion among institutional investors that commodity derivatives are an asset class that can be used to hedge equity risk, a notion we argue is mistaken. Even if the observed correlation between equity and commodity futures returns were reliably negative, it is likely that this negative correlation would be an equilibrium arbitrage phenomenon that should be expected in a world where no unexploited hedging profit opportunity exists. The rise in commodity derivative trading thus poses a challenge to asset-pricing theorists to explain in a well-articulated rational asset pricing model.

The lesson from this financial crisis is not that the government should prevent firms and investment funds from investing in commodity futures. As we noted, it was the unregulated, opaque OTC trading that was a critical factor in the financial crisis. The Dodd-Frank Act is intended to limit this type of trading and to make it more transparent. This outcome is already suggested by the incoming data. On organized exchanges (where traders are monitored and protected against counterparty failure), trading of commodity derivatives has nearly recovered

[11] Table 3 in Erb and Harvey (2006) reports the portfolio weights for three commodity futures indexes as of May 2004. Crude oil is about 40 percent of the S&P GSCI and all energy commodities make up two-thirds of the weight in the index. This does not include grains used for ethanol. They also report that 86 percent of the open interest in commodity futures indexes was in the S&P GSCI.

[12] See, for example, Basu and Gavin (2010).

to the peak achieved in June of 2008, while OTC trading in commodity derivatives has continued to decline.

A lesson from the crisis is that regulators and policymakers should monitor financial innovations closely to learn whether they are being used to take excessive risks—that is, risks firms would not take if they were operating outside the government's safety net. Under new regulations, the CFTC will collect information that should make trading in commodity derivatives more transparent. Banks argue that they need to use commodity derivatives to help customers manage risks. This may be true, but the recent experience in commodity futures did not reduce risks but exacerbated them just at the wrong time. The challenge to the government is to prevent too-big-to-fail firms from using current and yet invented derivatives to increase overall risk in the financial system.

REFERENCES

Banerji, Sanjay and Basu, Parantap. "Universal Banking and the Equity Risk Premium." Unpublished manuscript, Durham University, November 2010.

Basu, Parantap and Gavin, William T. "Negative Correlation between Stock and Futures Returns: An Unexploited Hedging Opportunity?" Unpublished manuscript, Federal Reserve Bank of St. Louis, December 2010.

Büyüks¸ahin, Bahattin; Haigh, Michael S. and Robe, Michel A. "Commodities and Equities: A 'Market of One'?" Working paper, Commodity Futures Trading Commission, June 9, 2008; ww.cftc.gov/ucm/groups/public/@aboutcftc/documents/file/amarketofone_update0608.pdf.

Cohn, Laura and Symonds, William C. "Striking Gold in Commodities." *Business Week*, October 25, 2004.

Dybvig, Philip H. and Marshall, William J. "The New Risk Management: The Good, the Bad, and the Ugly." Federal Reserve Bank of St. Louis *Review*, November/December 1997, *79*(6) pp. 9-22; http://research.stlouisfed.org/publications/review/97/11/9711pd.pdf.

Erb, Claude B. and Harvey, Campbell R. "The Strategic and Tactical Value of Commodity Futures." *Financial Analysts Journal*, March/April 2006, *62*(2), pp. 69-97.

Ferguson, Roger W. Jr.; Hartmann, Philip; Panetta, Fabio and Portes, Richard. *Geneva Report on the World Economy 9: International Financial Stability*. Geneva, Switzerland: Centre for Economic Policy Research and the International Center for Monetary and Banking Studies, 2007.

Gensler, Gary. "Remarks." Presented at the Sandler O'Neill Global Exchange and Brokerage Conference, New York, June 3, 2010; www.cftc.gov/pressroom/speechestestimony/opagensler-46.html.

Gerlach, Stefan; Giovannini, Alberto; Tille, Cédric and Viñals, José. *Geneva Reports on the World Economy 10: Are the Golden Years of Central Banking Over? The Crisis and the Challenges*. Geneva, Switzerland: Centre for Economic Policy Research and the International Centre and the International Center for Monetary and Banking Studies, 2009.

Gorton, Gary and Rouwenhorst, K. Geert. "Facts and Fantasies about Commodity Futures." *Financial Analysts Journal*, March/April 2006, *62*(2), pp. 47-68.

Greer, Robert J. "The Nature of Commodity Index Returns." *Journal of Alternative Investments*, Summer 2000, *3*(1), pp. 45-53.

Jiménez, Gabriel; Ongena, Steven; Peydró, José-Luis and Saurina, Jesús. "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk-Taking?" Working Paper No. 0833, Bank of Spain, 2008; www.bde.es/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/08/Fic/dt0833e.pdf.

Kilian, Lutz. "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market." *American Economic Review*, June 2009, *99*(3) pp. 1053-69.

Masters, Michael W. Testimony before the Committee on Homeland Security and Governmental Affairs. United States Senate, May 20, 2008; http://hsgac.senate.gov/public/_files/052008Masters.pdf.

Palmeri, Christopher. "CalPERS' New Crusade: Commodities Are Where the Hefty Returns Are, Says Investment Chief Russell Read." *BusinessWeek, June 5, 2006;* www.businessweek.com/print/magazine/content/06_23/b3987080.htm?chan=gl.

Rajan, Raghuram G. "Has Financial Development Made the World Riskier?" in *The Greenspan Era: Lessons for the Future*. Proceedings of the symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, August 25-27, 2005. Kansas City, MO: Federal Reserve Bank of Kansas City, 2005, pp. 313-67; www.kansascityfed.org/publicat/sympos/2005/pdf/Rajan2005.pdf.

Symonds, William C. "How to Invest Like Harvard." *BusinessWeek*, December 27, 2004.

Tang, Ke and Xiong, Wei. "Index Investment and the Financialization of Commodities." Working paper, Princeton University, August 2010; www.princeton.edu/~wxiong/papers/commodity.pdf.

Real-Time Forecast Averaging with ALFRED

Chanont Banternghansa and Michael W. McCracken

This paper presents empirical evidence on the efficacy of forecast averaging using the ALFRED (ArchivaL Federal Reserve Economic Data) real-time database. The authors consider averages over a variety of bivariate vector autoregressive models. These models are distinguished from one another based on at least one of the following factors: (i) the choice of variables used as predictors, (ii) the number of lags, (iii) use of all available data or only data after the Great Moderation, (iv) the observation window used to estimate the model parameters and construct averaging weights, and (v) the use of either iterated multistep or direct multistep methods for forecast horizons greater than one. A variety of averaging methods are considered. The results indicate that the benefits of model averaging relative to Bayesian information criterion-based model selection are highly dependent on the class of models averaged The authors provide a novel decomposition of the forecast improvements that allows determination of the most (and least) helpful types of averaging methods and models averaged across. (JEL E52, E58, C53)

Federal Reserve Bank of St. Louis *Review*, January/February 2011, *93*(1), pp. 49-66.

This paper provides evidence on the ability of various forms of forecast averaging to improve the real-time forecast accuracy of monthly bivariate vector autoregressive (VAR) forecasts of headline and core consumer price index (CPI)-based inflation, growth in industrial production (IP), and the unemployment rate. We consider a range of approaches to averaging forecasts obtained by a variety of primitive methods for managing the estimation of each bivariate VAR model. The averaging methods include equally weighted averages, medians, mean square error (MSE)-weighted averages, Bayesian model averages based on a Bayesian information criterion (BIC) approximation, and averages based on the top 10 percent of models that have performed best historically. For each averaging approach, we construct forecasts of each variable using real-time data from the ALFRED (ArchivaL Federal Reserve Economic Data) database. We compare our model-averaging results with those obtained with BIC-based model selection.

Model averaging for forecasting is nothing new. An abundance of evidence suggests that model averaging can improve forecast accuracy relative to model selection. Empirical examples of this evidence include, but are certainly not limited to, Stock and Watson (2004), Kapetanios, Labhard, and Price (2008), and Kascha and Ravazzolo (2010). Theoretical results include Hansen (2008), Elliott and Timmermann (2004), Clark and McCracken (2008), and many others.

In some instances (e.g., Clark and McCracken, 2010, and Faust and Wright, 2009), forecasting with model averages accounts for the real-time nature of the data. Even so, such examples are the exception and not the norm. Here we use the

Chanont Banternghansa was a research assistant and Michael W. McCracken is a research officer and economist at the Federal Reserve Bank of St. Louis.

ALFRED database to mimic the type of data that would have been accessible to forecasters at each point in time as they construct their monthly forecasts. Using real-time data is important because it accounts for the fact that economic data are often subject to revision and hence the actual value of a variable may change across forecast origins. In addition, using real-time data accounts for the fact that most macroeconomic data become available only after a substantial lag and, moreover, these time lags can vary widely across variables from as short as a week (for employment figures) to as long as two months (for trade data). Finally, by using the ALFRED database as the universe of potential predictors, we allow for the availability of new series across time and existing series that are sometimes discontinued.

In accordance with the literature, our results indicate that model averaging can—but does not always—improve forecast accuracy relative to the more-standard BIC-based approach to model selection. Put differently, model averaging per se is not a panacea for improving forecast accuracy. Improvements from model averaging depend critically on the type of models averaged across. Preselecting which primitive models should be used in the averaging process appears to offer some advantage. For example, when forecasting core CPI-based inflation there appear to be substantial gains in forecast accuracy at all horizons when averaging over only those models estimated with a rolling observation window of fixed size rather than a recursive, expanding observation window. In contrast, we find improved IP forecasting accuracy when averaging over only those models estimated with a recursive window rather than a rolling window of observations.

With these two examples in mind, we provide a novel decomposition of the relative root mean square error (RMSE) improvements for each dependent variable at each forecast horizon, which allows us to determine which primitive model types and model-averaging techniques are, on average, most (and least) beneficial. In some, though not all, instances our decomposition meshes well with the permutations of types of models and types of averaging procedures that produce the most accurate forecasts.

The remainder of the paper proceeds as follows. The next section describes the real-time data used in our analysis. We then provide a synopsis of the primitive models we average over, followed by a section describing the types of model averaging we consider. Finally, we present our results on forecast accuracy, our decomposition, and our conclusions.

DATA

We obtained our data from the ALFRED database maintained by the Federal Reserve Bank of St. Louis. This database consists of collections of vintages of data for each variable—that is, vintages that vary across time as either new data are released or existing data are revised by the relevant statistical agency. Using this database ensures that at each monthly forecast origin we are using only data that were available as of the date of the forecast origin. We therefore define "real-time" forecasting as using any data available by the end of the month from which we are forecasting.[1]

Choosing the end of a month as the forecast origin is nontrivial. Nearly all monthly macroeconomic data are released after the end of the month the data reference. A model needing data for January 1996 must therefore be constructed after that month has ended. If we choose the first day of February 1996 as our forecast origin, the forecast would be very timely but there would be almost no data for January to use, thus reducing the accuracy of the forecast. On the other hand, if we choose the first day of May as our forecast origin, all the data for January would be available but the forecast would be very outdated. As a middle ground we choose the end of the month following the most recent data vintage as the relevant forecast origin. For example, this implies that one-step-ahead forecasts, constructed using January 1996 vintage data, made at the end of

[1] The ALFRED database (http://alfred.stlouisfed.org/) allows retrieval of vintage versions of economic data available on specific dates in history. In general, economic data for past observation periods are revised as more accurate estimates become available. Currently, vintage data are available for 24,293 series in 14 categories.

February, will be forecasts of data associated with February 1996.[2]

Our analysis uses a total of 238 unique monthly macroeconomic series from the ALFRED database. Of these 238 series, 67 are available for the January 1996 vintage data. As we progress across time, we allow the number of variables to increase or decrease with data availability. For example, the number of series available more than doubles in November 1996. By the end of our forecasting exercise in December 2008 a total of 193 series are used either as dependent variables or as predictors. This is less than the total number of variables because 45 series were discontinued or did not have enough observations at some point in time to adequately estimate either the model parameters or model-averaging weights.[3] There are 29 output and production series; 8 income, outlays, and savings series; 40 labor market series; 52 monetary aggregate and reserve series; 35 exchange rate series; 38 financial market and interest rate series; 34 price series; and 2 survey series. The detailed list is available from the authors on request.

For brevity, in our forecasting exercise we focus exclusively on forecasting four of the most publicly visible nominal and real monthly frequency variables: headline and core CPI-based inflation, IP growth, and the unemployment rate. Specifically, at each forecast origin starting in February of 1996, we construct forecasts of three variables: headline CPI-based inflation, IP growth, and the unemployment rate. We begin forecasting core CPI-based inflation using December 1996 vintage data—the first available vintage for this series. For each of the four variables we construct h = 1-, 3-, 6-, 12- and 24-month-ahead forecasts. For unemployment, the target variable being forecast is y_{t+h}, the unemployment rate at the forecast horizon h. For CPI and IP, the target variable being forecast is the average annualized monthly rate of growth over the forecast horizon and hence interpretation of the target variable varies with the forecast horizon. More precisely, if we let y_t denote the time t log difference in, say, headline CPI, the target variable being forecast at horizon h is

$$y_{t+h}^{(h)} = \left(\tfrac{1200}{h}\right)\sum_{i=1}^{h} y_{t+i}.$$

In constructing our forecast errors, we use the third release (or, equivalently, the second revision) of the variable as the realized value of our target variable. In total, because December 2008 is the final vintage used to evaluate our forecasts, for each model we have roughly 155 1-month-ahead forecast errors that we use to measure accuracy. This number shrinks to 151, 145, 133, and 109 for the 3-, 6-, 12-, and 24-month-ahead forecasts, respectively. Following Marcellino, Stock, and Watson (2006), each variable is transformed to ensure stationarity using differences or log differences. For the dependent variables, we treat the unemployment rate as stationary in levels but treat headline CPI, core CPI, and IP as stationary in log-first differences. These transformations are made across all vintages uniformly. We do not allow for differences in the type of transformation across vintages. After transforming the variables we then check for outliers, defined as observations greater than six times the interquartile range. The outliers are replaced with the mean of the series (without the outlier) from the relevant vintage. This replacement is done vintage by vintage and hence the outlier detection is not influenced by observations not available at each forecast origin. Note that across the forecasting period the CPI and IP indices have been periodically renormalized so that the units of measurement are not the same across all vintages. To avoid mixing and matching, we renormalized each vintage relative to the December 2008 vintage.

METHODS

In this section we describe the primitive models over which we average. All models have one thing in common: They all take the form of

[2] Giannone, Reichlin, and Small (2008) refer to this type of forecast as a “nowcast.”

[3] In our analysis, we set a few basic rules for inclusion of variables: (i) We do not use seasonally unadjusted data when the seasonally adjusted version is available, (ii) we do not use regional data for our analysis, (iii) we omit a variable if fewer than 10 years of data are available for estimating the model parameters, and (iv) we omit a variable if we do not have at least 24 pseudo out-of-sample forecast errors to calculate the MSE-weighted forecasts.

an OLS-estimated bivariate VAR in the variable to be predicted and one additional predictor (see the section "Iterated Multistep and Direct Multistep Forecasts" for a caveat). Otherwise, all the primitive models differ by at least one of six features: (i) the series from the ALFRED database used as an additional predictor, (ii) the number of lags of the dependent variable used as a predictor, (iii) the number of lags of the additional predictor used, (iv) whether the model is estimated using all available data (i.e., the recursive scheme) or a moving window of observations (i.e., the rolling scheme), (v) whether the model is estimated using only post-Great Moderation data or data as far back as available for that vintage, or (vi) for forecast horizons greater than one step ahead, whether iterated multistep (IMS) or direct multistep (DMS) methods are used to create our primitive forecasts.

Predictors

As noted previously, we use the ALFRED database for our real-time forecasting exercise. In particular, we treat it as the universe of potential variables that could be used as a predictor for any one of our four dependent variables. Since the number of variables in ALFRED changes across forecast origins, the number of primitive models over which we average changes across forecast origins. At the beginning of our sample, January 1996, we have a total of only 66 potential predictors for each dependent variable. At the last potential forecast origin, November 2008, we have a total of 192 potential predictors for 1-step-ahead forecasts. While the number of predictors typically grows—sometimes dramatically, as for November 1996—in a few instances the number of predictors falls as various variables are discontinued or dropped because of insufficient data.[4]

Full Sample and Great Moderation Sample

For each model, we estimate the regression parameters using one of two subsets of data. In the first, the full sample, we use all available data in that vintage. While the date of the first observation varies across individual variables, many date back to as early as January 1959. In the second, the post sample, we restrict attention to only those data available starting in January 1983, roughly the time frame for the start of the Great Moderation. Note that for the post sample, this implies that for each vintage used for estimation, any pre-1983 observations are discarded.

We consider both subsets of data because there is considerable evidence, including that in D'Agostino, Giannone, and Surico (2007), that the predictability of many macroeconomic variables has changed since the onset of the Great Moderation. Even so, there is a trade-off. Using less information to estimate model parameters may generate estimates that are more likely to be unbiased because older data come from a different macroeconomic regime, but less information also can decrease the precision of the estimates. In practice, this trade-off may favor using more (or less) data to estimate parameters due to a bias-variance trade-off.

Recursive and Rolling Windows

For each model, and conditional on whether we use the full or post sample, we estimate the bivariate VAR using one of two observation windows. In the recursive scheme, we estimate the model by OLS using all available data. Hence as we move forward from one month to the next, we use one more observation to estimate the model parameters. In the rolling scheme, we estimate the model by OLS using only the past 10 years of available data. Hence when using the rolling scheme, as we move forward from one month to the next we use the same number of observations to estimate the model parameters.

In some ways, our decision to consider two subsets of data (full vs. post) and two types of observation windows (recursive vs. rolling) may seem redundant. We view the two choices, however, as distinct but related. In the former, we essentially assume a discrete break in 1983 and see how doing so helps forecast accuracy. For the latter, we assume a somewhat smoother sequence of breaks. Since we are unsure which is the proper way to manage forecasting in the presence of

[4] See footnote 3 for more detail.

uncertain forms of potential structural change, we consider both. See Clark and McCracken (2010) for further discussion on this issue.

Iterated Multistep and Direct Multistep Forecasts

For each permutation of predictor, sample, and observation window, we estimate our bivariate VAR forecasting model using two different methods: the textbook method that induces an IMS forecast and the somewhat easier-to-implement method of DMS forecasting. The following text provides a brief description of each approach.

Let y_t denote either the time t level of the unemployment rate or the time t log-first difference of headline or core CPI or IP. In addition, recall that the target variable to be forecast at forecast horizon h is

$$y_{t+h}^{(h)} = \left(\tfrac{1200}{h}\right)\sum_{i=1}^{h} y_{t+i}$$

for the CPI and IP indices but is simply y_{t+h} for unemployment. For the IMS forecasting approach, at each forecast origin t we first use OLS to estimate the bivariate VAR model,

$$(1)\quad \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \alpha_{y,0} \\ \alpha_{x,0} \end{pmatrix} + A(L)\begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix},$$

where $A(L)$ denotes a lag operator of appropriate dimension for the given number of lags used in both the y and x equations. With the regression parameter estimates in hand, the recursive nature of the VAR is used to generate a sequence of 1- through h-step-ahead forecasts $\hat{y}_{t+i}$ 1 = 1,...,h. For the unemployment rate, $\hat{y}_{t+h}$ is the resulting forecast of our target variable. For the other dependent variables, we follow Marcellino, Stock, and Watson (2006) and define our h-step-ahead IMS forecast as

$$\left(\tfrac{1200}{h}\right)\sum_{i=1}^{h} \hat{y}_{t+i}.$$

Note that for each forecast horizon, the same parameter estimates are used to construct the forecasts.

For the DMS forecasting approach, a distinct model is estimated separately for each forecast horizon h. For the unemployment rate and a fixed value of h, this model takes the form

$$(2)\quad y_{t+h} = \alpha_{y,0} + A_y(L)y_t + A_x(L)x_t + \varepsilon_{y,t+h},$$

where $A_y(L)$ and $A_x(L)$ denote lag operators of appropriate dimension for the given number of lags used for y and x, respectively. For each separate forecast horizon the forecast is defined as

$$\hat{y}_{t+h} = \hat{\alpha}_{y,0} + \hat{A}_y(L)y_t + \hat{A}_x(L)x_t.$$

For the CPI and IP indices, the model takes the slightly different form of

$$(3)\quad y_{t+h}^{(h)} = \alpha_{y,0} + A_y(L)y_t + A_x(L)x_t + \varepsilon_{y,t+h}.$$

For each separate forecast horizon the forecast is similarly defined as

$$\hat{y}_{t+h}^{(h)} = \hat{\alpha}_{y,0} + \hat{A}_y(L)y_t + \hat{A}_x(L)x_t.$$

Note that in each of the above examples, the parameter estimates from these models vary with the forecast horizon.

Lags

Each of the IMS and DMS specifications requires choosing the number of lags of y and x to use as predictors. The textbook approach would be to use a model-selection procedure such as BIC. Such a choice, however, contrasts with our goal of providing evidence on the benefits of model averaging relative to model-selection techniques. In addition, because of the considerable evidence suggesting a change in the degree of persistence in inflation (e.g., Levin and Piger, 2006), one might consider the possibility that the lag order structure of the model, for inflation in particular, has changed over time. We therefore consider all 144 permutations of up to 12 lags of either the y or x variable.

AVERAGING METHODS

After considering all the permutations of model elements discussed above, for each variable we have 76,128 1-month-ahead forecasting models estimated in January 1996 and 221,280

1-month-ahead forecasting models estimated in November 2008.[5] With this rich collection of individual forecasting models as building blocks, we consider a range of approaches to model averaging with an eye toward determining which types of model averaging are most useful and moreover, which types of primitive models are the most useful for averaging over.

Simple Model Averages

Our first set of model averages is the simplest. We consider the equally weighted average and the median forecast from among these models. While these methods are not statistically exciting, substantial evidence suggests that simple forms of model averaging can perform quite well (e.g., Smith and Wallis, 2009). Note that this form of model averaging implies model weights invariant to the forecast horizon.

Weighted Model Averages: Inverse Mean Square Error Weights

We then consider two distinct forms of weighted model averaging. In the first, we follow Stock and Watson (2004) (among others) and consider relative inverse mean square forecast error (MSE)-based weights to combine our models. The intuition is that if historical evidence suggests some models are more accurate than others, it may be beneficial to give those particular models more weight. Computationally, if $MSE_{i,t,h}$ denotes the known MSE associated with individual model i at forecast origin t associated with a sequence of past h-step-ahead forecast errors, the weight given to model i is

$$\frac{MSE_{i,t,h}^{-1}}{\sum_{j=1}^{N_t} MSE_{j,t,h}^{-1}},$$

where $j = 1,\ldots,N_t$ denotes an index of all the available primitive models at forecast origin t.

In our application, for the relevant vintages of data needed to estimate a particular model at forecast origin t, we conduct a pseudo out-of-sample forecasting exercise to generate these MSEs. The particulars of the exercise depend on whether (i) the full or post sample and (ii) the recursive or rolling scheme are used to construct our forecasts. If the recursive (rolling) scheme is used for the model forecast, then the recursive (rolling) scheme is used for the pseudo out-of-sample forecasts used to construct the model weights. If the full sample is used, the first pseudo out-of-sample forecast is based on parameters estimated using data from January 1960 to December 1969 and iterates forward until the availability of real-time data, at time t, is insufficient to calculate a forecast error using the third release of the relevant dependent variable. If the post sample is used, the first pseudo out-of-sample forecast is based on parameters estimated using data from January 1984 to December 1993 and iterates forward as discussed. Since our forecasting exercise starts in January 1996, this implies that the model weights constructed with the full sample are estimated based on an average MSE that uses many more squared forecast errors than those constructed with the post sample.

Weighted Model Averages: Bayesian Weights

We also consider an approximate Bayesian model-averaging strategy in which we calculate a posterior probability from prior probabilities and marginal likelihoods for each model, with each model assigned the same prior probability. Following Garratt, Koop, and Vahey (2006), the marginal likelihood of a given model is approximated using its BIC. In our analysis, for each vintage we estimate each model using the relevant subset of the available data (i.e., the full or post sample) and, based on the subsequent residuals, calculate the value of the BIC. Computationally, if we let $BIC_{i,t,h}$ denote the value of the BIC associated with the residuals from individual model i at forecast origin t, the weight given to model i is

$$\frac{\exp\left(-0.5 * BIC_{i,t,h}\right)}{\sum_{j=1}^{N_t} \exp\left(-0.5 * BIC_{j,t,h}\right)}.$$

For the IMS models the BIC is constructed in the typical fashion using equation (1), which implic-

[5] The number of models not only changes across forecast origins but also varies slightly across forecast horizons due to data availability. See footnote 3.

itly assumes that the residuals are serially uncorrelated. For the DMS models, however, we know that when $h > 1$ the residuals from equation (2) are not serially uncorrelated and hence the typical formulation is invalid.[6] For simplicity, we use the standard BIC formula regardless.

Weighted Model Averages with Trimming

In addition to the previously described weighted forecasts that average across all models, we also considered a variant that filters out the models considered "less accurate" by some metric and averages over only those remaining. Specifically, at each forecast origin *t* we follow Aiolfi and Timmermann (2006) and Clark and McCracken (2010) by calculating a top 10 percent MSE-weighted and a top 10 percent BIC-weighted average constructed using only the top 10 percent of the available models. For the top 10 percent MSE models this is done by averaging over only the models with the lowest 10 percent of pseudo out-of-sample MSEs based on the data available as of the forecast origin. Similarly, for the top 10 percent BIC models this is done by averaging over only the models with the lowest 10 percent of values of BIC based on the data available as of the forecast origin.

Benchmark Forecast

In reporting our results it is useful to gain some perspective on the magnitude of the benefits of model averaging. Doing so requires choosing a baseline for comparison. Since our goal is to observe the benefits of model averaging relative to model selection, using a fixed autoregressive model with known lags is insufficient. Not only does that baseline fail to capture the time-varying nature of model selection in a real-time forecast setting, in many cases it does not even serve as a particularly difficult benchmark to "beat." For example, we could have used the standard random walk benchmark but, as seen below, while this is a strong benchmark for the unemployment rate, it is a horrible benchmark for IP and both CPI indices.

Instead, we use the recursively estimated, IMS, BIC-selected forecast estimated over the full sample as our benchmark. At each forecast origin *t* this entails calculating the value of the BIC for each IMS model from equation (1), estimated by (i) using the full sample, separately across all possible lag permutations and choices of additional predictor, and (ii) then choosing the model with the lowest BIC as the model that is used to construct the forecast. The reason for our selection is that this particular BIC-selected forecast is the conventional methodology that a textbook in time-series econometrics would suggest. For completeness, we also report the relative RMSEs associated with the random walk model.

Before we proceed, it is important to clarify two things about our "benchmark model." First, it is chosen in real time in the sense that at each forecast origin we use only the vintage of data available at that forecast origin.[7] In particular, we use only the vintage of data available at the time the forecast is constructed to compute the value of the BIC for each possible model. Second, across time there is no single benchmark model. That is, as we proceed across forecast origins, it is possible for the model with the smallest value of BIC to change. This can occur for any number of reasons: the presence of unmodeled structural change, revisions in the data across vintages, or even changes in the collection of models considered as the universe of variables in ALFRED expands or contracts across time. Because of this possibility, the benchmark model is not so much a "model" as it is a forecasting method.

Summary of Methods

For each variable and each horizon, we consider six different forms of model averaging: average, median, (inverse) MSE-weighted, BIC-weighted, top 10 percent (inverse) MSE-weighted, and top 10 percent BIC-weighted. Each form of

[6] See Hansen (2010) for a discussion of how this affects the definition of BIC.

[7] Recall that the phrase "full sample" is intended to denote that for a given forecast origin the entirety of the corresponding vintage of data is used for estimation. This is in contrast to the phrase "post sample," which uses only the portion of the corresponding vintage that coincides with the Great Moderation.

Table 1

RMSEs of Out-of-Sample Forecasts of Nominal Variables

Variables	Forecast horizon: 1 month	3 month	6 month	12 month	24 month
Headline CPI					
BIC, recursive, IMS, full*	3.560	**2.741**	1.622	1.146	0.800
Random walk	1.151	1.519	2.298	2.851	4.234
Median	0.995	1.018	**0.990**	0.934	0.760
Average, all forecasts	0.995	1.021	1.008	0.952	0.816
MSE weight, all forecasts	0.995	1.018	0.993	0.934	0.778
MSE weight, top 10%	1.000	1.030	1.006	0.943	0.821
BIC weight, all forecasts	**0.994**	1.024	1.007	0.946	0.781
BIC weight, top 10%	**0.994**	1.023	0.996	**0.913**	**0.667**
Core CPI					
BIC, recursive, IMS, full*	1.233	0.805	0.606	0.586	0.591
Random walk	1.198	1.580	1.858	1.855	1.954
Median	0.938	0.942	0.899	0.867	0.827
Average, all forecasts	**0.931**	0.962	0.944	0.944	0.967
MSE weight, all forecasts	0.934	0.938	**0.884**	**0.840**	0.827
MSE weight, top 10%	0.958	0.949	0.893	0.848	0.834
BIC weight, all forecasts	0.936	0.946	0.918	0.918	0.922
BIC weight, top 10%	0.946	**0.932**	0.888	0.842	**0.810**

NOTE: *Values associated with the first row in each panel are RMSEs. The remaining values are ratios of RMSEs relative to that of the first row. For each forecast horizon, the best relative RMSE is shown in bold type. BIC, Bayesian information criterion; CPI, consumer price index; full, full sample (all available data in that vintage); IMS, iterated multistep; MSE, mean square error. See text for details.

averaging is then applied separately to several distinct classes of models, which are indexed by their type of construction using (i) the full and/or post samples, (ii) the recursive and/or rolling schemes, and (iii) the IMS and/or DMS approaches to forecasting. Note that since we allow for averaging over, for example, models estimated using either the recursive or rolling schemes, there are $3^3 = 27$ model classes that we consider. In all, this gives us $6 \times 3^3 = 162$ distinct permutations of forms of model averaging and the types of models that are averaged over.

RESULTS

In this section, we discuss our results on the benefits of using forecast averaging as a tool for improving forecast accuracy. For brevity, however, we do not present the tables associated with all 162 model-averaging and model class variants. Instead, Tables 1 and 2 present results for each type of model averaging when we average over *all* models. Table 1 presents results for headline and core CPI-based inflation and Table 2 presents results for growth in IP and the unemployment rate. The values in the first row of each panel of these tables are the RMSEs associated with the benchmark model chosen using BIC at each forecast origin. The remaining values in each panel are relative RMSEs. Values greater than 1 favor the benchmark model, while values less than 1 favor the form of model averaging denoted in the first column. For each forecast horizon, the best relative RMSE is shown in bold type.

Table 2

RMSEs of Out-of-Sample Forecasts of Real Variables

	Forecast horizon				
Variables	1 month	3 month	6 month	12 month	24 month
Industrial production					
BIC, recursive, IMS, full*	9.951	6.229	5.050	4.136	**2.837**
Random walk	1.125	1.263	1.313	1.561	2.281
Median	**0.985**	0.972	0.986	**0.994**	1.070
Average, all forecasts	**0.985**	**0.970**	**0.981**	**0.994**	1.055
MSE weight, all forecasts	0.986	**0.970**	0.982	0.996	1.056
MSE weight, top 10%	0.988	0.974	0.982	1.011	1.067
BIC weight, all forecasts	0.987	0.972	0.983	1.001	1.072
BIC weight, top 10%	0.989	0.990	1.005	1.023	1.094
Unemployment rate					
BIC, recursive, IMS, full*	0.167	0.301	0.463	0.696	1.023
Random walk	0.995	0.958	0.947	0.982	1.031
Median	0.936	0.882	0.864	0.922	0.936
Average, all forecasts	0.935	0.877	0.857	0.917	0.922
MSE weight, all forecasts	0.935	0.877	0.856	0.916	0.926
MSE weight, top 10%	**0.922**	**0.865**	**0.836**	**0.910**	0.969
BIC weight, all forecasts	0.935	0.879	0.862	0.918	**0.919**
BIC weight, top 10%	0.938	0.895	0.889	0.953	0.985

NOTE: *Values associated with the first row in each panel are RMSEs. The remaining values are ratios of RMSEs relative to that of the first row. For each forecast horizon, the best relative RMSE is shown in bold type. BIC, Bayesian information criterion; CPI, consumer price index; full, full sample (all available data in that vintage); IMS, iterated multistep; MSE, mean square error. See text for details.

Root Mean Square Errors of Nominal Variables

The first panel of Table 1 provides the results of forecasts for headline CPI-based inflation averaged across all models. At the three shortest horizons there are few, if any, advantages to forecast averaging across all models in terms of RMSEs. When averaging over all the primitive models, the benchmark is either better than model averaging or only marginally worse. However, as the forecast horizon increases to 12 months, model averaging improves accuracy by roughly 5 percent and at the longest horizon, forecast averaging improves accuracy by as much as 30 percent. In each of these latter horizons, the top 10 percent BIC-weighted forecasts yielded the lowest RMSEs.[8]

The second panel of Table 1 provides the results for core CPI-based inflation. In contrast to the results for headline inflation, consistent improvements are noted at all horizons for model averaging across all models. At the shortest horizons, the gains were on the order of a modest 5 percent, but as the horizon increases the improvements rise to about 15 percent. Across all horizons, no single averaging approach consistently gives the greatest improvements: The average, MSE-weighted, and top 10 percent BIC-weighted forecasts each perform best in at least one horizon.

[8] We do not test for statistical significance in our results because there is no known method for doing so when the baseline model is allowed to change across time and the competing model forecast is not based on a model per se but is instead an average across many models.

Root Mean Square Errors of Real Variables

The first and second panels of Table 2 parallel those in Table 1 in terms of the benefits of model averaging. As for headline CPI, model averaging across all models provides little to no improvement relative to model selection when forecasting IP growth at the shortest horizons. In fact, model averaging typically is worse than model selection at the longest horizons with losses of roughly 5 percent.

But again, in contrast to the results in the first panel, model averaging across all models consistently improves forecast accuracy relative to model selection when forecasting the unemployment rate. Each model-averaging procedure improves forecast accuracy at every horizon. Somewhat surprisingly, the improvements are (inverse) U shaped: The improvements in RMSE are roughly 7 percent at the shortest and longest horizons but are closer to 12 percent at the intermediate horizons. Across all but the longest horizon, the top 10 percent MSE-weighted forecast has the largest improvements relative to the benchmark. At the longest horizon the BIC-weighted average performs best.

Decomposition Regression Analysis

Tables 1 and 2 show that while model averaging can improve forecast accuracy, it does not always do so relative to our model selection-based benchmark. Moreover, when model averaging does provide improvements, the best form of model averaging varies across both dependent variables and forecast horizons. Finally, though obviously not apparent in Tables 1 and 2 (which present results averaged over *all* the primitive models), comparable conclusions can be reached if we report all the remaining permutations of types of model averaging and model classes for each dependent variable and each horizon.

Even so, it may be that on average across all these permutations, some simple patterns emerge that could help in identifying the best types of model averaging and the classes of models that should be averaged over. To parse out such effects we estimate a regression in which we use dummy variables for the types of model averaging and model classes as predictors for the corresponding relative RMSEs. Specifically, for each dependent variable and each forecast horizon, we use OLS to estimate the following regression:

$$
\begin{aligned}
&RMSE_i^h - 1 = \alpha_1 \text{DMS} + \alpha_2 \text{Post} + \alpha_3 \text{Roll} \\
(4)\quad &+\beta_1 \text{IMS/DMS} + \beta_2 \text{Full/Post} + \beta_3 \text{Rec/Roll} \\
&+\gamma_1 \text{Equal} + \gamma_2 \text{Weight} + \gamma_3 \text{Top } 10\% + \gamma_4 \text{MSE} + \varepsilon_i^h,
\end{aligned}
$$

where $RMSE_i^h$ is the relative RMSE of permutation $i = 1,\dots,162$ and *Rec* and *Roll* denote the recursive and rolling window schemes, respectively. By subtracting 1 the coefficients are more easily interpreted as indicating percent improvement (a negative coefficient) or percent deterioration (a positive coefficient) relative to our benchmark.

The α coefficients in equation (4) are associated with variables that indicate how an individual forecast is made: *DMS* takes the value 1 if only DMS models are included and 0 otherwise, *Post* takes the value 1 if only Great Moderation data are used and 0 otherwise, and *Roll* takes the value 1 if only a rolling window of observations is used to estimate the model parameters and 0 otherwise. The β coefficients are associated with the different combinations of the α coefficients: *IMS/DMS* takes the value 1 if the weighted forecast combines both IMS and DMS forecasts and 0 otherwise, *Full/Post* takes the value 1 if the weighted forecast combines both the full and post samples and 0 otherwise, and *Rec/Roll* takes the value 1 if the weighted forecast combines both recursive and rolling estimation schemes and 0 otherwise. The γ coefficients are associated with how the weighted forecasts are constructed: *Equal* takes the value 1 if either the average or median averaging methods are used and 0 otherwise, *Weight* takes the value 1 if the models are weighted unequally and 0 otherwise, *Top 10%* takes the value 1 if the averaging uses only the top 10 percent of forecasts and 0 otherwise, and *MSE* takes the value 1 if MSE-based weights are used and 0 otherwise.

Results for Nominal Variables

Table 3 shows decomposition results for both headline and core CPI-based inflation. In each

Table 3

Decomposition Regression of Nominal Variables

Variables	Forecast horizon 1 month	3 month	6 month	12 month	24 month
Headline CPI					
DMS	0.000	–0.004	–0.001	–0.016*	0.014
DMS/IMS	0.000	–0.001	0.001	–0.005	–0.000
Post	–0.011***	–0.027***	–0.049***	–0.066***	–0.147***
Post/Full	–0.009***	–0.019***	–0.035***	–0.050***	–0.113***
Roll	–0.004***	–0.045***	–0.075***	–0.086***	–0.141***
Rec/Roll	–0.012***	–0.031***	–0.050***	–0.078***	–0.177***
Equal	0.018***	0.074***	0.088***	0.082***	0.087**
Weighted	–0.001	–0.002	–0.000	0.001	–0.017
Top 10%	0.000	–0.002	–0.011*	–0.020**	–0.047***
MSE	–0.001	0.002	–0.006	–0.022***	–0.011
N	162	162	162	162	162
Core CPI					
DMS	–0.000	–0.010*	–0.034***	–0.085***	–0.152***
DMS/IMS	–0.000	–0.004	–0.011	–0.025	–0.044
Post	0.001	–0.041***	–0.075***	–0.130***	–0.230***
Post/Full	–0.002	–0.034***	–0.060***	–0.102***	–0.176***
Roll	–0.002*	–0.069***	–0.134***	–0.236***	–0.414***
Rec/Roll	–0.013***	–0.056***	–0.105***	–0.182***	–0.317***
Equal	–0.048***	0.046***	0.096***	0.215***	0.439***
Weighted	–0.004***	–0.013**	–0.011	–0.015	–0.027
Top 10%	0.009***	–0.011**	–0.025**	–0.049***	–0.073***
MSE	0.004**	0.000	–0.020**	–0.040**	–0.041
N	162	162	162	162	162

NOTE: Each column in each panel provides the coefficients associated with a distinct OLS-estimated version of equation (4). *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. BIC, Bayesian information criterion; CPI, consumer price index; DMS, direct multistep; Full, full sample (all available data in that vintage); IMS, iterated multistep; MSE, mean square error; Post, only data that occur starting in January 1983; Rec, recursive window scheme; Roll, rolling window scheme. See text for details.

panel, the first six rows relate to the selection of models to average over and the next four rows relate to the type of averaging method. We begin by studying panel 1 (that associated with headline CPI-based inflation).

In the first two rows of panel 1 (those associated with averaging over DMS models, IMS models, or both), there appears to be little statistically significant advantage to any of these particular forecasting methods. The sole exception is at the 12-month horizon, where DMS models appear to be favored. The results are stronger for the choice of data used to estimate the models. Across all horizons, the use of only Great Moderation data to estimate the models appears to be a significant advantage: Not only are the coefficients on post samples significantly different from 0 and negative, they also are more negative than the coefficients associated with averaging over both the full and post samples. The results for the choice of sampling scheme are a bit more muddled but still instructive. At the shortest and longest horizons, combining the recursive and rolling schemes—as suggested by Clark and McCracken (2008)—appears to offer the most advantage in terms of reducing RMSEs. At the other horizons, using the rolling scheme tends to be the best choice.

In the next four rows of panel 1, results for the type of averaging method clearly indicate that the simple equally weighted averaging methods perform significantly worse than the benchmark. At all horizons the coefficient associated with *Equal* is positive and different from 0. Unfortunately, the remaining three rows are not as easy to interpret. While the *MSE*, *Weight*, and *Top 10%* coefficients are typically negative—suggesting that a top 10 percent MSE-weighted average might be best—the coefficients are statistically significant only in a few instances at the longer horizons.

The results in panel 2 (those associated with core inflation) are similar to those for headline inflation, with a few specific differences. The evidence in favor of using the DMS approach to forecasting is stronger at all horizons and significantly so. Again, for all but the shortest horizon, the evidence favors using only Great Moderation data to estimate the model parameters. Similarly, using the rolling scheme or a combination of the rolling and recursive schemes is the preferred approach.

In the seventh through ninth rows of panel 2, the results for the type of averaging method are much sharper than those for headline inflation. In all but the shortest horizons, the simple equally weighted averaging methods perform significantly worse than the benchmark. But at the 1-month horizon, it appears that a simple averaging method does provide significant gains in forecast accuracy and, moreover, those gains are larger than when some form of weighting is used. For horizons longer than 1 month, the coefficients on *Top 10%* are all significantly negative, which along with the negative *MSE* and *Weight* coefficients suggests that a top 10 percent MSE-weighted average might be best.

Results for Real Variables

The results for the real variables (Table 4), particularly those for IP, are quite different from those for the nominal variables. A quick glance at the first six rows of panel 1 indicates quite clearly that the preferred model types for averaging are now IMS forecasting models estimated recursively using the full sample—a sharp contrast to the type of models chosen for both headline and core CPI-based inflation. Moreover, in the next four rows of panel 1, it appears that while some evidence favors MSE weighting relative to BIC weighting, the majority of the evidence suggests even better results would be obtained using the simple equally weighted averages rather than a weighted or top 10 percent weighted average.

The results in panel 2 (those associated with the unemployment rate) are less clear cut than those for IP and even those for headline and core CPI-based inflation. At the 3- and 6-month horizons, the DMS approach to forecasting appears to perform best but at the longest horizon the IMS appears to perform best. Similarly, at the intermediate horizons, using the post (Great Moderation) sample appears to perform best but at the longest horizon the full sample appears to perform best. And while the rolling scheme or a combination of the recursive and rolling schemes tends to perform best at the shortest horizons, the recursive

Table 4

Decomposition Regression of Real Variables

	Forecast horizon				
Variables	1 month	3 month	6 month	12 month	24 month
Industrial production					
DMS	–0.000	0.003***	0.013***	0.024***	0.025***
DMS/IMS	–0.000	0.000	0.004**	0.001	–0.003
Post	0.000	0.000	0.006***	0.006**	–0.001
Post/Full	–0.000	–0.000	0.004**	0.004	–0.001
Roll	0.012***	0.013***	0.049***	0.089***	0.147***
Rec/Roll	0.004***	0.006***	0.025***	0.050***	0.085***
Equal	–0.018***	–0.034***	–0.047***	–0.053***	–0.009**
Weighted	0.001***	0.002***	0.000	–0.000	0.003
Top 10%	0.003***	0.007***	0.007***	0.002	0.003
MSE	–0.002***	–0.006***	–0.003*	0.002	–0.006
N	162	162	162	162	162
Unemployment rate					
DMS	0.000	–0.006***	–0.016***	0.004	0.077***
DMS/IMS	0.000	–0.002	–0.004*	–0.002	0.001
Post	0.001	–0.009***	–0.010***	–0.009***	0.026***
Post/Full	–0.001	–0.007***	–0.008***	–0.009***	0.014
Roll	–0.014***	–0.002	0.013***	0.054***	0.159***
Rec/Roll	–0.009***	–0.004**	0.004	0.023***	0.071***
Equal	–0.054***	–0.107***	–0.127***	–0.088***	–0.151***
Weighted	0.003***	0.003*	0.003	0.002	0.001
Top 10%	–0.007***	–0.005**	–0.006*	0.001	0.034***
MSE	–0.008***	–0.009***	–0.014***	–0.010***	–0.006
N	162	162	162	162	162

NOTE: Each column in each panel provides the coefficients associated with a distinct OLS-estimated version of equation (4). *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. BIC, Bayesian information criterion; CPI, consumer price index; DMS, direct multistep; Full, full sample (all available data in that vintage); IMS, iterated multistep; MSE, mean square error; Post, only data that occur starting in January 1983; Rec, recursive window scheme; Roll, rolling window scheme. See text for details.

scheme clearly tends to dominate at the 6-month and longer horizons. Finally, as for IP, it appears that some evidence favors MSE weighting relative to BIC weighting, but the majority of the evidence suggests even better results would come from using one of the equally weighted averages rather than a weighted or top 10 percent weighted average.

Rankings

Tables 3 and 4 give some indication of which model-averaging types should be used and which model classes should be averaged over. However, we emphasize that these results are indicators of average treatment effects across all 162 permutations of averages and model classes. They do not necessarily indicate which permutations actually do perform best. Tables 5 and 6 provide a brief description of the permutations that perform best. In particular, we list the 10 best-performing permutations of averaging methods and model classes and their respective relative RMSEs for each variable and each of the 1-, 3-, and 12-month horizons.[9] In addition, we provide the five worst-performing permutations for the sake of comparison.

The first panel of Table 5 provides the rankings for headline CPI-based inflation. There are several striking features. In line with the results from Table 1, at the 1- and 3-month horizons there are few, if any, gains to model averaging irrelevant of model class. But as the horizon increases to 12 months, gains of roughly 10 percent are available when top 10 percent weighted averages are used; these gains are consistent with the decomposition results from Table 3. In addition, across all horizons, the 10 best-performing permutations use either the rolling scheme or a combination of the rolling and recursive schemes. In contrast, the five worst-performing permutations exclusively use the recursive scheme. Finally, as suggested in Table 3, all but one of the five worst-performing permutations use the simple equally weighted averaging schemes.

The second panel of Table 5 (that associated with core inflation) offers a slightly different picture of the benefits of model averaging relative to model selection. In particular, as in Table 1, model averaging is consistently beneficial at all horizons provided the right permutations of model averages and model classes are used. The 10 best-performing permutations outperform the benchmark by roughly 7 percent at the shortest horizon and by as much as 25 percent at the longest horizon. On the other hand, the 5 worst-performing permutations outperform the benchmark at the 1-month horizon but not at the 3- and 12-month horizons.

Interestingly, the types of model averages that perform best and worst for core inflation coincide nicely with the results in Table 3. At the shortest horizon, equally weighted averages tend to perform best but as the horizon increases, the top 10 percent weighted averages begin to dominate. In general, the class of models to average over also coincides with the results in Table 3: The 10 best-performing permutations are dominated by DMS forecasting models estimated over the post (Great Moderation) sample or an average of the post and full samples, using the rolling scheme or a combination of the rolling and recursive schemes. One result that does not coincide is at the 12-month horizon, where the BIC-weighted averages appear to perform best while the results in Table 3 suggest the MSE-weighted average would perform better.

The first panel of Table 6 provides the rankings for IP growth. As in Table 2, the advantages to model averaging relative to model selection, while feasible, are not particularly large, with a maximum of only 5 percent at the 12-month horizon. As indicated in the decomposition (see Table 4), the equally weighted averages seem to perform best at the 1-month horizon but as the horizon increases to 3 months, top 10 percent weighted averages appear to gain some traction among the best-performing averaging methods—a sharp contrast to the decomposition. Apparently part of the problem is that many of the worst-performing models are also top 10 percent weighted averages; hence, in averaging across all permutations, the decomposition indicates the equally weighted averages should perform better. One point that clearly matches our decomposition is the choice of sampling scheme: Nearly

[9] We present these three horizons for brevity. A complete set of results is available from the authors on request.

Table 5

Ranking of Model Averagings for Nominal Variables

	Forecast horizon					
Ranking	1 month	Relative RMSE	3 month	Relative RMSE	12 month	Relative RMSE
CPI						
1	BIC-DMS/IMS-Post-Rec/Roll	0.993	Top 10% BIC-DMS/IMS-Post-Roll	1.005	Top 10% MSE-DMS-Full-Roll	0.894
2	BIC-IMS-Post-Rec/Roll	0.993	Top 10% BIC-DMS/IMS-Full-Roll	1.005	Top 10% BIC-IMS-Post-Rec/Roll	0.899
3	BIC-DMS-Post-Rec/Roll	0.993	Top 10% BIC-DMS/IMS-Full/Post-Roll	1.005	Top 10% BIC-DMS/IMS-Post-Rec/Roll	0.900
4	Simple Avg.-IMS-Post-Rec/Roll	0.994	Top 10% BIC-IMS-Full-Roll	1.005	Top 10% MSE-DMS-Full-Rec/Roll	0.902
5	Simple Avg.-DMS-Post-Rec/Roll	0.994	Top 10% BIC-IMS-Post-Roll	1.005	Top 10% BIC-IMS-Full/Post-Rec/Roll	0.911
6	Simple Avg.-DMS/IMS-Post-Rec/Roll	0.994	Top 10% BIC-IMS-Full/Post-Roll	1.005	Top 10% BIC-DMS/IMS-Full/Post-Rec/Roll	0.913
7	BIC-DMS/IMS-Full/Post-Rec/Roll	0.994	Top 10% BIC-DMS-Post-Roll	1.006	Top 10% MSE-DMS/IMS-Full-Roll	0.915
8	BIC-IMS-Full/Post-Rec/Roll	0.994	Top 10% BIC-DMS-Full-Roll	1.006	Top 10% MSE-DMS/IMS-Full-Rec/Roll	0.919
9	BIC-DMS-Full/Post-Rec/Roll	0.994	Top 10% BIC-DMS-Full/Post-Roll	1.006	MSE-DMS-Full-Roll	0.919
10	Top 10% BIC-DMS-Post-Rec/Roll	0.994	Median-DMS-Full/Post-Roll	1.006	BIC-DMS-Post-Rec/Roll	0.921
⋮						
158	Simple Avg.-IMS-Full-Rec	1.032	BIC-IMS-Full-Rec	1.113	Median-DMS/IMS-Full-Rec	1.195
159	Simple Avg.-DMS-Full-Rec	1.032	Median-DMS-Full-Rec	1.113	Median-DMS-Full-Rec	1.196
160	Median-DMS-Full-Rec	1.040	Simple Avg.-IMS-Full-Rec	1.114	Simple Avg.-DMS/IMS-Full-Rec	1.197
161	Median-DMS/IMS-Full-Rec	1.040	Median-DMS/IMS-Full-Rec	1.117	Median-IMS-Full-Rec	1.201
162	Median-IMS-Full-Rec	1.040	Median-IMS-Full-Rec	1.120	Simple Avg.-DMS-Full-Rec	1.212
Core CPI						
1	Simple Avg.-DMS-Full-Rec/Roll	0.929	Top 10% BIC-DMS-Full-Roll	0.928	Top 10% BIC-DMS-Post-Roll	0.761
2	Simple Avg.-DMS/IMS-Full-Rec/Roll	0.929	Top 10% BIC-DMS-Post-Roll	0.928	Top 10% BIC-DMS-Full-Roll	0.761
3	Simple Avg.-IMS-Full-Rec/Roll	0.929	Top 10% BIC-DMS-Full/Post-Roll	0.928	Top 10% BIC-DMS-Full/Post-Roll	0.761
4	MSE-DMS-Full-Rec/Roll	0.930	MSE-IMS-Post-Roll	0.929	Top 10% BIC-DMS-Full-Rec/Roll	0.769
5	MSE-DMS/IMS-Full-Rec/Roll	0.930	MSE-IMS-Full/Post-Roll	0.929	Top 10% BIC-DMS-Full/Post-Rec/Roll	0.780
6	MSE-IMS-Full-Rec/Roll	0.930	Top 10% BIC-DMS/IMS-Full-Rec/Roll	0.930	Top 10% BIC-DMS-Post-Rec/Roll	0.783
7	Simple Avg.-DMS-Full/Post-Rec/Roll	0.931	Top 10% BIC-DMS-Full-Rec/Roll	0.930	MSE-DMS-Post-Roll	0.797
8	Simple Avg.-DMS/IMS-Full/Post-Rec/Roll	0.931	Top 10% BIC-DMS-Full/Post-Rec/Roll	0.931	MSE-DMS-Full/Post-Roll	0.798
9	Simple Avg.-IMS-Full/Post-Rec/Roll	0.931	MSE-DMS/IMS-Post-Roll	0.931	Top 10% MSE-DMS-Full-Roll	0.799
10	Top 10% MSE-DMS-Full-Rec/Roll	0.932	MSE-DMS/IMS-Full/Post-Roll	0.931	BIC-DMS-Full/Post-Roll	0.801
⋮						
158	Top 10% MSE-DMS/IMS-Post-Roll	0.966	BIC-DMS/IMS-Full-Rec	1.105	Median-IMS-Full-Rec	1.397
159	Simple Avg.-IMS-Full-Rec	0.967	Median-IMS-Full-Rec	1.108	Top 10% BIC-DMS/IMS-Full-Rec	1.436
160	Simple Avg.-DMS/IMS-Full-Rec	0.967	BIC-IMS-Full-Rec	1.115	BIC-DMS/IMS-Full-Rec	1.478
161	Simple Avg.-DMS-Full-Rec	0.967	Simple Avg.-DMS/IMS-Full-Rec	1.116	BIC-IMS-Full-Rec	1.484
162	Top 10% MSE-IMS-Post-Roll	0.967	Simple Avg.-IMS-Full-Rec	1.133	Simple Avg.-IMS-Full-Rec	1.523

NOTE: The values indicate the relative RMSEs of the model-averaging permutation in each row relative to the baseline. Avg., averaging; BIC, Bayesian information criterion; CPI, consumer price index; DMS, direct multistep; Full, full sample (all available data in that vintage); IMS, iterated multistep; MSE, mean square error; Post, only data that occur starting in January 1983; Rec, recursive window scheme; Roll, rolling window scheme. See text for details.

Table 6

Ranking of Model Averagings for Real Variables

Ranking	1 month	Relative RMSE	3 month	Relative RMSE	12 month	Relative RMSE
Industrial production						
1	Median-IMS-Full/Post-Rec	0.978	Top 10% BIC-DMS-Full-Rec	0.962	BIC-IMS-Full-Rec	0.949
2	Median-DMS-Full/Post-Rec	0.978	Top 10% MSE-IMS-Full/Post-Rec	0.964	Simple Avg.-IMS-Full-Rec	0.949
3	Median-DMS-Full/Post-Rec	0.978	Top 10% MSE-DMS-Full-Rec	0.964	BIC-DMS/IMS-Full-Rec	0.949
4	Simple Avg.-DMS-Full/Post-Rec	0.981	Top 10% BIC-IMS-Full-Rec	0.964	Top 10% BIC-DMS/IMS-Full-Rec	0.949
5	Simple Avg.-DMS/IMS-Full/Post-Rec	0.981	Top 10% MSE-DMS/IMS-Full-Rec	0.965	Median-IMS-Full-Rec	0.950
6	Simple Avg.-IMS-Full/Post-Rec	0.981	Simple Avg.-IMS-Post-Rec	0.965	MSE-IMS-Full-Rec	0.950
7	Top 10% BIC-DMS-Full-Rec	0.981	Median-IMS-Post-Rec	0.965	MSE-DMS/IMS-Full-Rec	0.951
8	Top 10% BIC-IMS-Full-Rec	0.981	Top 10% MSE-IMS-Full-Rec	0.965	Simple Avg.-DMS/IMS-Full-Rec	0.952
9	Top 10% BIC-DMS/IMS-Full-Rec	0.981	Top 10% MSE-DMS/IMS-Full/Post-Rec	0.965	Median-DMS/IMS-Full-Rec	0.952
10	BIC-DMS/IMS-Full/Post-Rec	0.981	MSE-IMS-Post-Rec	0.965	Simple Avg.-IMS-Full/Post-Rec	0.954
⋮						
158	Top 10% BIC-IMS-Full-Roll	1.001	Top 10% BIC-IMS-Full-Roll	0.996	MSE-DMS-Full-Roll	1.088
159	Top 10% BIC-DMS-Full-Roll	1.001	Top 10% BIC-IMS-Post-Roll	0.996	Median-DMS-Full/Post-Roll	1.090
160	Top 10% BIC-DMS-Full/Post-Roll	1.001	Top 10% MSE-DMS-Post-Roll	0.997	Median-DMS-Post-Roll	1.090
161	Top 10% BIC-IMS-Post-Roll	1.001	Top 10% BIC-DMS-Full/Post-Rec/Roll	0.997	Median-DMS-Full-Roll	1.090
162	Top 10% BIC-IMS-Full/Post-Roll	1.001	Top 10% BIC-DMS-Post-Rec/Roll	0.998	Top 10% MSE-DMS-Full-Roll	1.093
Unemployment rate						
1	Top 10% MSE-IMS-Post-Roll	0.905	Top 10% MSE-DMS-Post-Rec/Roll	0.848	Top 10% MSE-DMS-Post-Rec	0.855
2	Top 10% MSE-DMS/IMS-Post-Roll	0.905	Top 10% MSE-DMS-Post-Rec	0.849	Top 10% MSE-DMS-Post-Rec/Roll	0.860
3	Top 10% MSE-DMS-Post-Roll	0.905	Top 10% MSE-DMS-Post-Roll	0.851	Top 10% BIC-DMS-Full-Rec	0.863
4	Top 10% MSE-IMS-Full/Post-Roll	0.913	Top 10% MSE-DMS/IMS-Post-Rec/Roll	0.851	Top 10% BIC-DMS-Post-Rec	0.867
5	Top 10% MSE-IMS-Post-Rec/Roll	0.913	Top 10% MSE-DMS/IMS-Post-Rec	0.851	Top 10% MSE-DMS-Full/Post-Rec	0.867
6	Top 10% MSE-DMS/IMS-Full/Post-Roll	0.913	Top 10% MSE-DMS/IMS-Post-Roll	0.853	Top 10% BIC-DMS-Full/Post-Rec	0.874
7	Top 10% MSE-DMS/IMS-Post-Rec/Roll	0.914	Top 10% MSE-IMS-Post-Rec	0.854	Top 10% MSE-DMS-Full/Post-Rec/Roll	0.875
8	Top 10% MSE-DMS-Full/Post-Roll	0.914	Top 10% MSE-IMS-Post-Rec/Roll	0.854	Top 10% MSE-DMS/IMS-Post-Rec	0.877
9	Top 10% MSE-DMS-Post-Rec/Roll	0.915	Top 10% MSE-IMS-Post-Roll	0.856	Top 10% MSE-DMS/IMS-Full/Post-Rec	0.892
10	Top 10% MSE-IMS-Full/Post-Rec/Roll	0.921	Top 10% MSE-DMS-Full/Post-Roll	0.863	Top 10% BIC-DMS/IMS-Full-Rec	0.892
⋮						
158	Simple Avg.-DMS-Post-Rec	0.951	Top 10% BIC-IMS-Post-Roll	0.899	MSE-DMS-Full-Roll	0.987
159	Simple Avg.-IMS-Post-Rec	0.951	Top 10% BIC-IMS-Full-Roll	0.899	Median-DMS-Full-Roll	0.989
160	Median-DMS-Post-Rec	0.952	Top 10% BIC-IMS-Full/Post-Roll	0.899	Median-DMS-Full/Post-Roll	0.989
161	Median-DMS/IMS-Post-Rec	0.952	Median-DMS-Full-Rec	0.899	Median-DMS-Post-Roll	0.989
162	Median-IMS-Post-Rec	0.952	Top 10% BIC-IMS-Full-Rec/Roll	0.900	Top 10% MSE-DMS-Full-Roll	0.995

NOTE: The values indicate the relative RMSEs of the model-averaging permutation in each row relative to the baseline. Avg., averaging; BIC, Bayesian information criterion; CPI, consumer price index; DMS, direct multistep; Full, full sample (all available data in that vintage); IMS, iterated multistep; MSE, mean square error; Post, only data that occur starting in January 1983; Rec, recursive window scheme; Roll, rolling window scheme. See text for details.

all the best-performing permutations average across models estimated with the recursive scheme while all the worst-performing permutations average across models estimated with the rolling scheme.

In the second panel of Table 6 (that associated with forecasts of the unemployment rate), a few things are immediately apparent. First, model averaging uniformly improves forecast accuracy across all horizons and all permutations. In fact, at the 3-month horizon, the worst-performing model average provides an improvement of 10 percent relative to the benchmark. Also, across all horizons the 10 best-performing types of model averaging are of the top 10 percent form. This is in sharp contrast with the decomposition results, which predicted that the equally weighted averages tended to perform best. Even so, as for the decomposition shown in Table 4, it appears that at the shortest horizon the rolling scheme appears to perform best but as the horizon increases the recursive scheme becomes preferred. At the 12-month horizon, the 5 worst-performing permutations use the rolling scheme.

CONCLUSION

We use the ALFRED real-time database to provide empirical evidence on the real-time benefits of model averaging monthly-frequency forecasts of headline and core CPI-based inflation, growth in IP, and the unemployment rate. Our results support those discussed in much of the literature on forecasting: Model averaging typically improves forecast accuracy relative to a benchmark chosen using model selection. Even so, we emphasize a different point that is typically glossed over in the literature on forecast averaging: The choice of models averaged across can greatly influence the efficacy of the averaging methods.

This of course raises the question of how to choose the correct class of models to average across. Based upon a novel decomposition of the benefits of forecast averaging relative to using model-selection methods, a few rules of thumb seem evident. First, DMS forecasting models estimated over the post (Great Moderation) sample (or an average of the post and full samples) using the rolling scheme (or a combination of the rolling and recursive schemes) seem to perform best for forecasting either headline or core CPI-based inflation. Second, averaging over models estimated using the recursive scheme (or an average of the rolling and recursive) seems to perform best for forecasting either IP growth or the unemployment rate. Third, the top 10 percent averaging approach frequently provides the best improvements in forecast accuracy, but it is not immune to poor performance relative to equally weighted averages because past model performance does not always ensure future model performance.

REFERENCES

Aiolfi, Marco and Timmermann, Allan. “Persistence in Forecasting Performance and Conditional Combination Strategies.” *Journal of Econometrics*, 2006, *135*(1-2), pp. 31-53.

Clark, Todd E. and McCracken, Michael W. “Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts.” *International Economic Review*, 2008, *50*(2), pp. 363-95.

Clark, Todd E. and McCracken, Michael W. “Averaging Forecasts from VARs with Uncertain Instabilities.” *Journal of Applied Econometrics*, January/February 2010, *25*(1), pp. 5-29.

D’Agostino, Antonello; Giannone, Domenico and Surico, Paolo. “(Un)Predictability and Macroeconomic Stability.” Working Paper Series No. 605, European Central Bank, April 2006; www.ecb.int/pub/pdf/scpwps/ecbwp605.pdf.

Elliott, Graham and Timmermann, Allan. "Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions." *Journal of Econometrics*, September 2004, *122*(1), pp. 47-79.

Faust, Jon and Wright, Jonathan H. "Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset." *Journal of Business and Economic Statistics*, October 2009, *27*(4), pp. 468-79.

Garratt, Anthony; Koop, Gary and Vahey, Shaun P. "Forecasting Substantial Data Revisions in the Presence of Model Uncertainty." *Economic Journal*, 2008, *118*(53), pp. 1128-44.

Giannone, Domenico; Reichlin, Lucrezia and Small, David. "Nowcasting: The Real Time Informational Content of Macroeconomic Data Releases." *Journal of Monetary Economics*, May 2008, *55*(4), pp. 665-76.

Hansen, Bruce E. "Least-Squares Forecast Averaging." *Journal of Econometrics*, 2008, *146*(2), pp. 342-50.

Hansen, Bruce E. "Multi-Step Forecast Model Selection." Presented at the 20th Annual Meetings of the Midwest Econometrics Group, October 1-2, 2010, Olin Business School, Washington University in St. Louis; http://apps.olin.wustl.edu/MEGConference/Files/pdf/2010/61.pdf.

Kapetanios, George; Labhard, Vincent and Price, Simon. "Forecasting Using Bayesian and Information Theoretic Model Averaging: An Application to U.K. inflation." *Journal of Business and Economic Statistics*, January 2008, *26*, pp. 33-41.

Kascha, Christian and Ravazzolo, Francesco. "Combining Inflation Density Forecasts." *Journal of Forecasting*, 2010, *29*(1-2), pp. 231-50.

Levin, Andrew T. and Piger, Jeremy M. "Is Inflation Persistence Intrinsic in Industrial Economies?" Working Paper No. 2002-0231; Federal Reserve Bank of St. Louis; October 2002, revised November 2003; http://research.stlouisfed.org/wp/2002/2002-023.pdf.

Marcellino, Massimiliano; Stock, James H. and Watson, Mark W. "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Time Series." *Journal of Econometrics*, November-December 2006, *135*(1-2), pp. 499-526.

Smith, Jeremy and Wallis, Kenneth F. "A Simple Explanation of the Forecast Combination Puzzle." *Oxford Bulletin of Economics and Statistics*, June 2009, *71*(3), pp. 331-55.

Stock, James H. and Watson, Mark. "Combination Forecasts of Output Growth in a Seven-Country Data Set." *Journal of Forecasting*, September 2004, *23*(6), pp. 405-30.

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