# Essays in Macroeconomics:

Author: David Schenck



Boston College Electronic Thesis or Dissertation, 2020

Boston College

The Graduate School of Arts and Sciences

Department of Economics

# ESSAYS IN MACROECONOMICS

a dissertation

by

DAVID SCHENCK

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May 2020

# Essays in Macroeconomics

**David Schenck**

Advised by Professor Susanto Basu (Chair), Professor Peter Ireland,

and Associate Professor Ryan Chahrour

## Abstract

My dissertation consists of three independent chapters analyzing parameter estimation and structural change in applied macroeconomics. A first theme linking these papers is structural change, especially as it relates to the monetary policy transmission mechanism through the Phillips curve. A second theme is an assessment of small-sample statistical inference for impulse response functions after estimating macroeconomic models. Two of my chapters provide simulation studies of statistical coverage of standard test statistics after estimating impulse response functions in both atheoretical (local projection) and highly structural (dynamic stochastic general equilibrium) models.

The first chapter of my dissertation, "Using Survey Expectations to Estimate the New Keynesian Phillips Curve," provides new estimates of the parameters in the New Keynesian Phillips Curve, exploiting survey based expectations data provided by the Survey of Professional Forecasters and the Michigan Survey of Consumers. I find that the use of survey expectations in US data improves the fit of the textbook Phillips Curve model to the data and provides economically sensible estimates of its coefficients. The estimated model pro-

vides stable parameter estimates until the Great Recession, after which inflation becomes less dependent on marginal cost. Household and professional forecasts each contribute to the forward-looking component of inflation expectations, with household forecasts given more weight.

The second chapter of my dissertation, "Estimating Structural Breaks in Impulse Response Functions via the Local Projection Estimator," proposes an estimator for parameter instability in impulse response functions that are estimated by local projections. I use the estimator to investigate the presence of parameter instability in the Romer–Romer monetary policy shocks. I find evidence of a structural break in the impulse response coefficients in the late 1970s. In the early period, there is strong evidence that monetary policy shocks have real effects. There is little evidence that monetary policy shocks have real effects in the later period. Tax and oil price shocks exhibit little change in their effects on output throughout the postwar period.

The third chapter of my dissertation, "Standard Errors for Impulse Response Functions of Estimated DSGE Models," provides a method for constructing appropriate asymptotic standard errors for impulse responses of estimated dynamic stochastic general equilibrium models. The method requires only the matrices characterizing the model solution, the derivatives of those matrices with respect to the underlying structural parameters, and the covariance estimate of the structural parameters themselves. I provide simulation evidence on the small-sample properties of these standard errors.

*To my parents*

# Acknowledgments

I am extremely grateful to my advisors Susanto Basu, Peter Ireland, and Ryan Chahrour for guidance and comments. I also thank Robert Murphy, Sanjay Chugh, and numerous colleagues in the Boston College Economics Department for their suggestions and for fostering a vibrant intellectual environment.

I thank David Drukker, Vince Wiggins, and the econometrics group at Stata for valuable feedback on the second and third chapters of this dissertation, and Stata for its hospitality while writing those two chapters.

I owe a special debt to Dr. Leo Kim, Dr. Petros Carvounis, the ophthalmology team of Massachusetts Eye and Ear Infirmary, and the ophthalmology team of Baylor College of Medicine for critical medical support at key points of my graduate studies.

Finally, this work would not have been possible without the loving support of my family. I thank my aunt Peggy Cooper for endless encouragement and financial support. I thank my father, who encouraged me to pursue excellence. And I thank my mother, who taught me patience, persistence, and faith.

# Contents

# List of Tables

*List of Tables*

# List of Figures

# Chapter 1

# Using Survey Expectations to Estimate the New Keynesian Phillips Curve

## 1.1 Introduction

Macroeconomic models of price-setting center around a dynamic problem in which firms choose a path of prices subject to adjustment costs or constraints on their ability to re-adjust prices in the future. Once aggregated, firm-level choices about the dynamic path of prices lead to an aggregate equation, the New Keynesian Phillips Curve, which links inflation to expectations of future inflation and marginal cost. The Phillips Curve can then be taken to data and its parameters can be estimated. This paper investigates the degree to which observable survey measures of inflation expectations can improve the fit of Phillips Curve estimates, the degree to which differences in expectations across economic agents matters for inflation determination, and the degree to which the Phillips Curve's parameters exhibit structural change.

Inflation dynamics are interesting both in their own right and as a necessary step towards understanding the real effects of monetary policy. Researchers have sought to capture inflation dynamics with small models containing few variables and linked tightly to economic theory; one key discussion that has emerged is the role of forecasts of future inflation relative to lagged inflation in determining current inflation. A related discussion has centered around the role of slack in the economy in driving inflation. In older models, economic slack as measured by detrended real GDP or estimates of the output gap entered directly in the Phillips Curve; in modern models, slack covaries with marginal cost, which in turn serves as a direct determinant of inflation. It is of interest whether these theoretical models can be successfully fit to data and whether there has been structural change in the relative importance of inflation determinants during the recent recession.

Inflation expectations are unobservable. In the absence of a direct measure of firms' expectations, researchers have turned to two approaches to accounting for expectations in empirical work. One line of research has leaned heavily on rational expectations, replacing inflation expectations with actual future inflation and instrumenting accordingly. A second line of work has used surveys of economic agents as a proxy for firms' expectations. Galı and Gertler (1999) is exemplary of the first approach and provides the benchmark study of inflation dynamics. They find evidence for forward-looking behavior in the Phillips Curve alongside a substantial backward-looking component; this leads them to propose a hybrid

Phillips Curve specification that has been in wide use since. Roberts (1995) is an early example of the survey approach.

This paper investigates the extent to which survey measures of expected future inflation can improve the fit of Phillips Curve estimates. I find that surveys measures of expectations enter significantly into Phillips Curve estimates and that using survey expectations yields stable evidence for the Phillips Curve's main theoretical prediction of a link between inflation and marginal cost. However, the survey expectations Phillips Curve does not reject a lagged inflation term. Further, consumer forecasts of inflation also enter significantly into the estimates, indicating that firms' expectations may be proxied by a combination of professional and household forecasts.

The first result is that using survey expectations yields stable estimates of the marginal cost coefficient in the New Keynesian Phillips Curve even when the Great Recession is included in the dataset. The marginal cost coefficient is stable across multiple specifications of the real variable (labor share, HP -filtered labor share, and output gap). These results contrast with the rational expectations approach, which tends to find only weak support for a marginal cost term and indeed sometimes finds the marginal cost term entering with the "wrong" sign. These results complement the work of Nunes (2010), which finds that expectations from the Survey of Professional Forecasters add value to rational expectations estimates. However, while the Phillips Curve coefficients are stable when estimated through the Great Moderation, they show structural change during the Great Recession: since 2007, inflation has become less dependent on marginal cost and has become more persistent.

The second result is that measurement matters for answering two substantive economic questions: which real variable best fits the data, and the role of inflation persistence in Phillips Curve estimates. Using the GDP deflator and the log-labor share (the closest available analogues to the model's objects) yields stable Phillips Curve estimates that nevertheless also imply a strong role for lagged inflation in determining current inflation. However, a Phillips Curve estimated on consumer price inflation yields an only marginally significant link between the output gap and inflation, and estimates no significant link between marginal cost and inflation. In addition, estimates of a CPI Phillips Curve lead one to reject a lagged inflation term and provide support for a fully forward-looking specification of the Phillips Curve (c.f. Fuhrer, 2013), while using the GDP deflator lends support towards a hybrid model with forward- and backward- looking components receiving approximately equal weight in determining current inflation. When taken to data that closely map to the model objects, survey expectations support the inflation-marginal cost link but cannot resolve the "inflation persistence" puzzle.

Several different surveys of inflation expectations are available, capturing both professional and household forecasts. The third result is that consumer forecasts of inflation matter just as much as, or even somewhat more than, professional forecasts in the estimation – a result which lends further support to the findings in Coibion and Gorodnichenko (2015). Across several specifications, consumer forecasts of "the expected rise in general prices" enter strongly into the GDP deflator Phillips Curve; indeed, consumer expectations are given more weight than professional forecasts in most specifications. If we take these surveys as proxying for underlying unobserved firm expectations, then firms have expectations that are a convex combination of household and professional forecasts with a weight of one-half to two-thirds on the household forecast.

The rest of this paper is organized as follows. Section 1.2 reviews the theory of the New Keynesian Phillips Curve and derives the estimating equation. Section 1.3 carefully discusses data issues to be considered in the empirical exercise. Section 1.4 describes the results. Section 1.5 discusses applications to recent inflation dynamics. Section 1.7 presents some implications of these results for macro modelling generally, and offers conclusions.

## 1.2 Phillips Curve Estimation

### 1.2.1 The Estimating Equation

Two models of price adjustment have emerged as workhorses for the study of inflation dynamics: the staggered price setting model of Calvo (1983) and Yun (1996), and the quadratic cost of price adjustment model of Rotemberg (1982). To a first-order approximation, the two models yield identical estimating equations. I will go through the derivation of the quadratic cost model, in part because it yields convenient closed-form expressions both in the level and the log- deviation of its variables.

Consider a representative goods-producing firm which maximizes the expected discounted flow of dividends, subject to a demand constraint and a cost of adjusting prices. In order to focus on the dynamic aspects of the problem, suppose the firm minimizes cost within each period and has marginal cost $\varphi_{it}$. Then the firm's dynamic problem will be:

$$\begin{aligned} \max_{P_{it}, Y_{it}} \quad & E_0 \sum_{t=0}^{\infty} \beta^t m_t \left[ \left( \frac{P_{it}}{P_t} \right) Y_{it} - \varphi_{it} Y_{it} - Y_t \frac{\phi_p}{2} \left( \frac{P_{it}}{\pi P_{i,t-1}} - 1 \right)^2 \right] \\ \text{s.t.} \quad & Y_{it} = \left( \frac{P_{it}}{P_t} \right)^{-\theta} Y_t. \end{aligned}$$

The firm takes aggregate output $Y_t$, the aggregate price level $P_t$, and the stochastic discount factor $m_t$ as given. The firm's demand constraint relates demand for the firm's product to its relative price and aggregate income; it is the demand function that would be generated by a consumer minimizing the cost of a Dixit-Stiglitz aggregate of goods. $\theta$ indexes the degree of substitutability across varieties. The parameter $\phi_p$ captures the degree of price rigidity. By taking the demand constraint as given, the firm pledges to sell output on demand at its chosen price $P_{it}$ . $\phi_p$ captures the cost of price adjustment, $\pi$ the trend rate of inflation in the economy, and $\beta$ is the discount factor.

Substituting the demand constraint into the objective function, the firm faces an unconstrained problem in optimally choosing a sequence of nominal prices, $P_{it}$ for $t = 1, 2, 3 \ldots$. The first-order condition for an individual firm is:

$$\begin{aligned} (\theta - 1) \left( \frac{P_{it}}{P_t} \right)^{-\theta} \frac{Y_t}{P_t} + \phi_p \left( \frac{P_{it}}{\pi P_{i,t-1}} - 1 \right) \frac{Y_t}{\pi P_{i,t-1}} & \\ = \theta \left( \frac{P_{it}}{P_t} \right)^{-\theta-1} \varphi_{it} + \beta \phi_p E_t & \left[ \frac{m_{t+1}}{m_t} \left( \frac{P_{i,t+1}}{\pi P_{i,t}} - 1 \right) \frac{Y_{t+1}}{P_{it}} \frac{P_{i,t+1}}{\pi P_{it}} \right] \end{aligned}$$

This first-order condition relates the firm's pricing choice to current marginal cost and the expected future path of prices.

Imposing symmetry across firms, we can write the nonlinear New Keynesian Phillips Curve as:

$$\frac{(\theta-1)}{\phi_p} + \left(\frac{\pi_t}{\pi} - 1\right)\frac{\pi_t}{\pi} = \frac{\theta}{\phi_p}\varphi_t + \beta E_t\left[\frac{m_{t+1}}{m_t}\frac{Y_{t+1}}{Y_t}\left(\frac{\pi_{t+1}}{\pi} - 1\right)\frac{\pi_{t+1}}{\pi}\right] \tag{1.1}$$

which links price inflation to expected future inflation and marginal cost.

Up to a first-order approximation, the previous equation can be written as:

$$\hat{\pi}_t = \beta E_t \hat{\pi}_{t+1} + \lambda \hat{\varphi}_t \tag{1.2}$$

where $\lambda = (\theta - 1)/\phi_p$ and hats denote deviations of each variable from its steady-state value. It is this log-linearized Phillips Curve that is most often taken to the data.

The log-linearized equation is often augmented with a lagged inflation term, and the main empirical specification is:

$$\hat{\pi}_t = \beta\left[\alpha\hat{\pi}_{t-1} + (1-\alpha)E_t\hat{\pi}_{t+1}\right] + \lambda\hat{\varphi}_t \tag{1.3}$$

see, among others, Ireland (2004). From here, the main questions of interest are the size of $\lambda$ and the size and significance of the backward-looking coefficient $\alpha$. The hybrid specification nests two interesting special cases: fully forward-looking behavior ($\alpha = 0$) and fully backward-looking behavior ($\alpha = 1$). Fully backward-looking behavior is consistent with older accelerationist theories of the Phillips Curve; intermediate ranges of $\alpha$ indicate a hybrid Philips Curve that weighs both past and expected future inflation.

### 1.2.2 Standard Estimation

We can re-write the log-linearized Phillips Curve as:

$$E_t\left[\hat{\pi}_t - \beta\left(\alpha\hat{\pi}_{t-1} + (1-\alpha)\hat{\pi}_{t+1}\right) - \lambda\hat{\varphi}_t\right] = 0$$

If agents form expectations rationally, then the conditional expectation operator is the true statistical expectation, and we can write:

$$E\left[\hat{\pi}_t - \beta\left(\alpha\hat{\pi}_{t-1} + (1-\alpha)\hat{\pi}_{t+1}\right) - \lambda\hat{\varphi}_t | \Omega_t\right] = 0$$

where $\Omega_t$ contains all information known at time $t$. Hence the Euler equation along with the rational expectations assumption naturally generates a set of moment conditions. This is exactly an instrumental variables/method of moments problem, with any variable known at time $t$ or before serving as a valid instrument. Hence,

$$E\left[\left(\hat{\pi}_t - \beta\left(\alpha\pi_{t-1} + (1-\alpha)\hat{\pi}_{t+1}\right) - \lambda\hat{\varphi}_t\right) z_t\right] = 0$$

with $z_t$ being lagged data serving as moment conditions. Since the null hypothesis of rational expectations implies that *any* information known at time $t$ or before is a valid instrument, the above equation provides researchers a rich array of instrument sets in practice.

### 1.2.3 Estimation in This Paper

I replace $E_t\pi_{t+1}$ with the mean forecast of inflation in the GDP deflator as delivered by the Survey of Professional Forecasters. Hence my estimating equation will be of the form:

$$\hat{\pi}_t = \beta_1\hat{\pi}^e_{t+1|t} + \beta_2\hat{\pi}_{t-1} + \lambda\hat{\varphi}_t + \varepsilon_t \tag{1.4}$$

directly substituting the rational expectation with the SPF survey expectation.

By making this substitution, I assume that firm's inflation forecasts are at minimum colinear with professional forecasts. The parameter estimates should be interpreted as semi-structural. I do not provide estimates of the discount factor or the relative weight on past expectations; I restrict my results to the coefficients on survey expectations, lagged inflation, and real activity themselves. Backing out estimates of the deeper parameters requires placing implausible assumptions on the degree to which measured surveys of householdss and professional forecasters map onto the expectations of price-setters.

The time-$t$ dated error term contains two pieces: all differences between observed SPF and unobserved firm expectations and the potential measurement error in the marginal cost proxy. I instrument all regressions using lags two through four of inflation, the labor share, commodity price inflation, the output gap, and nominal wage inflation. These variables are in principle available to agents' information sets at time $t$.

## 1.3 Data Issues

The theoretical objects examined above must be mapped to data prior to estimation. The approach of Gali and Gertler is to replace inflation expectations with actual future inflation, instrumenting appropriately and exploiting the moment properties of rational expectations. I replace expected inflation with the Survey of Professional Forecasters' inflation expectation. There are a variety of options for proxying the real variable, as marginal cost is unobservable; theory provides some guidance on the measurement of inflation and marginal cost. Given data limitations, the choice of data and their links to the model's objects require some comment.

All data decisions in this exercise involve some degree of compromise. There is no perfect link among measures of inflation, inflation expectations, and real activity. For inflation, the primary choice is between the nonfarm business deflator, which is the closest analogue to the theory, and the GDP deflator, which is the inflation measure for which we have survey expectations available. As for the slack variable, the labor share has traditionally been used to proxy for real activity in the marginal cost NKPC. HOwever, the labor share has shown considerable downweard secular trend since 2000, even after measurement and other conceptual issues have been cleaned out; see, for example, Elsby, Hobijn, and Şahin (2013). The nonstationarity of the labor share necessitates that some pre-filtering be applied before the data can be brought to the model. Theory provides little guidance on how this filtering is to be done. Hence some stand must be taken on a filtering method, before the model can even be tested.

My main specification uses the nonfarm business deflator as the measure of inflation, to conform to the theory; survey expectations of the GDP deflator, as it is the closest object to the NFB deflator for which we have expectations data; and a filter for the labor share suggested by Hamilton (2018), which essentially uses a $h-$step ahead regression forecast error to model the cyclical component of a series. Robustness exercises are included throughout to assess the sensitivity of results to these data choices.

### 1.3.1 Inflation data

Throughout, I report results using three measures of inflation. The GDP deflator is the broadest measure I consider; it is often employed in estimated DSGE models, as in Ireland (2004), Christiano, Eichenbaum, and Evans (2005), and Smets and Wouters (2007). Second, I provide results for the inflation rate of the implicit price deflator for the nonfarm business sector. This measure is commonly used in single–equation studies of the Phillips Curve such as Sbordone (2002) and King and Watson (2012). Galı and Gertler (1999) provides results using both the GDP deflator and the nonfarm business deflator. Finally, I include results using the CPI inflation rate.

The GDP deflator is a standard measure of inflation. However, it includes components that are not priced competitively, such as defense expenditures and some education expenditures. Thus, not all components of the GDP deflator should be considered as generated from the staggered price setting process emphasized in the NKPC. Nonfarm private business prices, on the other hand, more closely match that description. A drawback is that survey expectations are reported with reference to the GDP deflator, not the NFB deflator.

### 1.3.2 The Real Variable

The New Keynesian Phillips Curve links realized inflation to expectations of future inflation and real activity. As has been noted in Coibion, Gorodnichenko, and Kamdar (2018), the empirical strength of the NKPC is sensitive to the measure of real activity used. The theoretically appropriate real variable is the real marginal cost of production. In some situations, real marginal cost can itself be linked to the output gap, providing justification for output gap Phillips curve estimation. Measurement of real marginal cost or the output gap presents challenges that may introduce measurement error into the estimating equation.

Prior studies have focused on labor's share of income as a proxy for marginal cost. This decision follows a line of work, beginning with Galı and Gertler (1999); which has attempted to align Phillips curve estimation with the theory. The proper relationship that emerges from a dynamic optimization problem links inflation to marginal cost; only under restrictive assumptions does marginal cost in turn linearly relate to the output gap. Further, under cost-minimization, real marginal cost can be linked to the firm's wage bill.

I use labor's share of nonfarm business business sector. The labor share was approximately stable from 1968 to 2000, but has shown considerable secular decline since 2000. The decline of the labor share has been a global phenomenon, as documented in Karabarbounis and Neiman (2013). As the secular decline in the labor share is not the primary focus of this paper, a filtered version of the labor share will be used in the main empirical specification. I use the forecast error filter proposed in Hamilton (2018). The filtered value of a variable $\hat{y}_{t+k}$ is the forecast error of a regression of $y_{t+k}$ on the contemporaenous and most recent three lags, $y_t$, $y_{t-1}$, $y_{t-2}$, $y_{t-3}$. I use four quarters as the forecast horizon $k$; the results are not substantively affected by alternatives such as a six-quarter-ahead horizon or an eight-quarter-ahead horizon.

An older literature uses the output gap as the real variable, instead of a proxy for marginal cost; I operationalize this idea with the CBO-reported output gap. This variable more broadly captures the idea that some measure of economic slack ought to enter into the Phillips Curve; however it does not connect closely with the model's prediction that marginal cost specifically ought to enter into the Phillips Curve. The estimated output gap is further confounded by possible measurement error. I will provide some evidence on the output gap as a contrast to the labor share specification.

To link the two ideas, consider a model without capital or net exports. Further suppose that households have isoelastic preferences with regard to consumption and work effort; let $\sigma$ be the coefficient of risk aversion and $\eta$ be the labor supply elasticity. Then marginal cost $\varphi_t$ is linked to the output gap $x_t$ via

$$\hat{\varphi}_t = (\eta + \sigma)\hat{x}_t$$

implying two possible specifications of the Phillips Curve:

$$\hat{\pi}_t = \beta \hat{\pi}^e_{t+1} + \lambda \hat{\varphi}_t + \varepsilon_t \tag{1.5}$$

$$\hat{\pi}_t = \beta \hat{\pi}^e_{t+1} + \kappa \hat{x}_t + \varepsilon_t \tag{1.6}$$

Given data on both marginal cost and the output gap, both versions model will be tested. The link between the two objects is not exactly linear in the presence of variable capital utilization, but nevertheless this exercise allows for separate estimation of the coefficients on the output gap and marginal cost. As inflation is not detrended in the following regressions, I provide two measures of marginal cost: the labor share detrended with the Hodrick- Prescott filter, and the logarithm of the level of the labor share. While statistical significance will not differ much across these two specifications, the size of the coefficient will. For the output gap, I will use the logarithm of GDP less the logarithm of potential GDP as estimated by the Congressional Budget Office.

### 1.3.3 Expectations

There are two primary sources of inflation expectations data. The Survey of Professional Forecasters provides expectations of GDP deflator and CPI inflation from forecasters. This data is available quarterly since 1968 for the deflator and since 1981 for the CPI; both series include one-quarter- ahead projections of the inflation rate. The SPF survey is typically taken in the middle of the quarter and hence includes all information in prior quarters and partial information in the current quarter. In all regressions, the forecast variable is mapped to the dependent variable: deflator expectats in regressions involving the GDP deflator or the NFB price index, and CPI expectations to regressions involving CPI inflation.

A second source of expectations data is the Michigan survey of households. This data is available since 1978 and tracks households' expectations of the percentage change of "prices in general" over the next year. While the theory is clear that the expectations of firms is the relevant variable, I will explore the role of household expectations in the Phillips Curve as well. I will provide results with these data and contrast them with the results of the SPF data, noting that the Michigan Survey question is a slightly different frame than the SPF question.

Throughout, samples begin when the SPF data become available. Estimation with the SPF forecasts of the GDP deflator begins in 1968; estimation with the SPF forecast of the CPI begins in 1981; and estimation with the Michigan Survey begins in 1978. All regressions include data through 2018 unless otherwise noted.

## 1.4 Results

### 1.4.1 Estimates of the Phillips Curve

I begin by examining two benchmark variants of the Phillips Curve specifications:

$$\pi_t = \beta\pi^e_{t+1|t} + \lambda\varphi_t + \varepsilon_t \tag{1.7}$$

$$\pi_t = \beta\pi^e_{t+1|t} + \kappa x_t + \varepsilon_t \tag{1.8}$$

Here, $\pi^e_{t+1|t}$ is the forecast of inflation one period ahead. The marginal cost term $\varphi_t$ is Hamilton's filter of labor's share of income, while output gap $x_t$ is taken from the CBO estimate. Estimation is by the generalized method of moments. Instruments include four lags of inflation, the labor share, commodity price inflation, the output gap, and nominal wage inflation.

Throughout, I provide one specification test of the forward-looking Phillips Curve model, the "hybrid" Phillips Curve:

$$\pi_t = \beta_1\pi^e_{t+1|t} + \beta_2\pi_{t-1} + \lambda\varphi_t + \varepsilon_t \tag{1.9}$$

$$\pi_t = \beta_1\pi^e_{t+1|t} + \beta_2\pi_{t-1} + \kappa x_t + \varepsilon_t \tag{1.10}$$

By including a lagged inflation term, the New Keynesian Phillips Curve can be tested against an alternate form which includes both a forward-looking and backward-looking component. Prior studies have found that the backward- looking component enters the estimation with an statistically significant and economically large coefficient, so throughout I will investigate whether survey expectations reduce inflation's dependence on a lagged dependent variable.

The error term $\varepsilon_t$ subsumes several pieces, each with economic content. First, it measures the deviation of professional survey forecasts from those of actual firms. The theoretically appropriate expectation is that of firms; to the extent that firms' and professional forecasters' expectations differ, error is introduced into the regression equation. Second, the error term measures the deviation of professional survey forecasts from rational expectations. If one wishes to maintain the null hypothesis of rational expectations, then any deviation of professional forecasts from rational expectations introduces a time-$t$ error into the regression. Third, it includes measurement error from the deviation of the measured output gap or marginal cost from the unobservable, true output gap or marginal cost.

Begin with the results in Table 1.1, which reports results from the simplest possible specification: inflation from the nonfarm business deflator is regressed on its expected value one period ahead from the Survey of Professional Forecasters and on the filtered labor's share of income, the proxy for marginal cost. I find that expected future inflation enters significantly and with a value around one. The labor share enters significantly and with a value of around 0.19, indicating a shallow but statistically significant Phillips Curve relationship. The inclusion of the lagged dependent variable enters significantly and reduces the coefficient on the expectations term. In the hybrid specification, the lagged inflation term enters with a coefficient of around one-third, and the forward-looking term retains a large coefficient of about 0.75. Although the lagged inflation term is statistically significant, its role is quantitatively small compared to the forward-looking component. The coefficient on the marginal cost term remains steady, with a point estimate of 0.18.

The estimated coefficient on the labor share lies between 0.15 to 0.20. Galı and Gertler (1999) obtains elasticities of inflation with respect to marginal cost of about 0.05 for the period 1960-1997, roughly one-third of the value found here. The discrepancy is not simply due to the differing sample periods; regressions using survey expectations, but restricted to the 1968-1997 sample, also yield output gap elasticities between 0.15 and 0.20. The discrepancy also holds in the 1968-1997 sample when marginal cost is measured as the log-level of the labor share. The elasticity reported here is comparable to that in Nunes (2010). Estimation with survey expectations delivers a Phillips Curve whose marginal cost coefficient is of correct sign, is consistently statistically significant, and with a slope steeper than what is obtained in estimation using rational expectations GMM.

The coefficient on expected future inflation is around unity for the forward- looking specification, indicating that an increase in expected future inflation feeds one-for-one into current inflation. However, the coefficient falls sharply when a lagged dependent variable is included. Furthermore, residuals from the forward-looking specification show some autocorrelation, while residuals from the hybrid specification show no signs of autocorrelation. The sum of forward-looking and backward-looking coefficients in the hybrid Phillips Curve is statistically indistinguishable from unity. The backward-looking component is around one-third, indicating considerable backward-looking behavior despite the use of survey expectations.

Hence the baseline results provide two strong results in favor of the New Keynesian Phillips Curve. First, the coefficient on expected future inflation enters significantly and with a coefficient that is in the range of that suggested by the theory. Second, the coefficient on the labor share is significant, positive, and economically reasonable.

However, the purely forward-looking model is not entirely successful. First, the hybrid model does not reject positive $\alpha$, and the estimated $\alpha$ is economically large. Second, the forward-looking model has highly autocorrelated errors, but the hybrid model does not; this result indicates unobserved persistence in the forward-looking specification but not in the hybrid specification, a point in favor of the hybrid model. Thus using survey expectations as an observable measure of unobserved inflation expectations solves one puzzle, the weak contribution of marginal cost to Phillips Curve estimates, but does not solve a second, the persistent importance of a lagged inflation term. A hybrid, marginal cost Phillips Curve with roughly equal weights on lagged and expected inflation best fit the data.

### 1.4.2 Marginal Cost, the Labor Share, and the Output Gap

The central prediction of the Phillips Curve is of a link between nominal and real variables; as the theory has evolved, the relevant real variable has evolved as well. Beginning as a theory that linked wage inflation to the level of the unemployment rate, modern New Keynesian theories of the Phillips Curve link (producer) price inflation to marginal cost. This section examines the robustness of the above baseline parameter estimates to different specifications of the real variable.

First, the log-level of the labor share may be used instead of its HP- filtered value. As mentioned previously, this strategy closely aligns with the theory. However, the labor share has seen substantial secular drift since 2000, complicating a straightforward application of the theory to the data. Nevertheless, this exercise taken in the first two columns of Table 1.2. Again the dependent variable is inflation as measured by the change in the nonfarm business price index. In both the forward-looking and hybrid specifications, the labor share enters with the wrong sign and is statistically insignificant.

Second, the Congressional Budget Office provides an estimate of the output gap, which provides another variable which might be used to proxy for marginal cost. The specification with the CBO output gap is provided in the latter two columns of Table 1.2. The output gap enters significantly with a coefficient somewhat higher than the log labor share, but lower than the HP-filtered labor share. It is significant and significance remains when a lagged dependent variable is included in the regression. The coefficient on the SPF inflation expectation is around unity in the forward-looking specification and falls to about one-half in the hybrid specification; the implied $\alpha$ remains one-half and the implied $\beta$ does not differ from unity.

Table 1.3 restricts the sample to 1968-2000, but otherwise provides the same information as 1.2. This period is of particular interest in that it conforms to the sample period used in Galı and Gertler (1999) and King and Watson (2012), and because this period is characterized by relative stationarity of the labor share. In this sample, the log labor share enters significantly and with the correct sign; the estimated slope coefficient is relatively steep. The estimated slope of marginal cost Phillips curve is about 0.33, and this estimate does not change much when a lagged inflation term is included. The lagged inflation term enters with a point estimate of less than 0.2 and is statistically insignificant. However, the specification without a lagged dependent variable shows moderate serial correlation in the residuals; the specification with a lagged dependent variable shows no discernable autocorrelation in the residuals. A Phillips curve estimated from 1968-2000 with survey expectations delivers a forward- looking specification with a large, positive, statistically significant coefficient on marginal cost.

In the early sample, the output gap Phillips curve specification provides an output gap coefficient of bout 0.25, somewhat larger than is obtained in the full sample. In the output gap specification, a lagged dependent variable enters significantly in a specification test. As with the marginal cost specification, the output gap specification suffers from autocorrelation in the residuals when a lagged dependent variable is not included.

In using three different measures of marginal cost and the output gap, some general observations emerge. The coefficient on the real variable is significant across most specifications, with a central tendency around 0.15 in the full sample. A lagged dependent variable tends to enter significantly in the full sample specifications. In a sample restricted to 1968-2000, the gap measure enters significantly in all specifications with a relatively steep slope.

### 1.4.3 Alternate Measures of Inflation

The Survey of Professional Forecasters also provides forecasts of CPI inflation, so it is possible to investigate the degree to which differences in the measure of inflation affect the ability of survey expectations to resolve puzzles in Phillips Curve estimation. This section provides key results for the CPI forecast. I provide results with the filtered labor share and with the CBO output gap. The focus of this section is estimation in which the dependent variable and the measure of survey expectations cohere. This involves two specifications: one with inflation measured by the change in the GDP deflator, and one with inflation measured by the change in the CPI. The SPF CPI inflation forecast is available for 1981:I to the present. As before, I include specifications with a lagged inflation term as a test of the forward-looking Phillips Curve against a hybrid alternative. These results are in Table 1.4 and 1.5.

In table 1.4, instrumental variables regressions are run in which the GDP deflator and CPI inflation as the dependent variable, the slack measure is the Hamilton-filtered labor share, and expectations are measured by the relevant one–quarter–ahead SPF forecast. Results for the GDP deflator mimic those for the nonfarm business deflator. The labor share estimate is positive, statistically significant, and economically large. The estimate is about 0.2 for the purely forward-looking specification and is slightly smaller, 0.13, for the hybrid specification. Second, the hybrid Phillps curve provides a better fit, in that lagged inflation enters significantly and the weights on the forward-looking and backward-looking components are evenly split. Third, the hybrid specification is necessary to remove autocorrelation in the residuals.

Results for the CPI specification are weaker. In neither specification does the labor share enter significantly; its point estimates have the wrong sign. The hybrid specification places no weight on lagged CPI inflation. The sum of forward-looking and backward-looking coefficients is below unity regardless of the choice of marginal cost or output gap. Finally, note the explanatory power of the model falls considerably in the CPI specification.

Hence CPI inflation seems to be more forward-looking than GDP deflator inflation, and also less well-characterized by the NKPC. The CPI sample is shorter than the NFB or GDP deflator samples, and runs only from 1981 to present. This sample encompasses the Great Moderation, Great Recession, and subsequent recovery, which may explain the lack of inflation persistence. Simple univariate tests show that the GDP deflator is about twice as persistent as the CPI over the 1981- present period; these reduced-form differences show up significantly in the Phillips curve regressions.

These results show that in the case of Phillips Curve estimation, measurement of inflation matters. Where the theory most closely maps into data – measuring inflation with the GDP deflator, using forecasts of the deflator, and using the labor share as the proxy for marginal cost – the hybrid model fits the data quite well. Even using the output gap leads to strong support for the hybrid model. However, once one turns to different measures of inflation the data begin to reject the model and provide misleading inferences about the strength of inflation persistence. While the broadest notion of a Phillips Curve relationship between inflation and the output gap remains, the more precise link between price inflation and marginal cost breaks down.

### 1.4.4 Household versus Professional Forecasts

The Michigan Survey of Consumers provides quarterly expectations of inflation by households beginning in 1978. The Michigan Survey collects forecasts of " the expected change in prices," not specifying any particular set of prices to be forecasted. Michigan survey forecasts are made over the following year, not the following quarter. Consumers are not necessarily predicting the same object as forecasters. With these caveats, it is possible to investigate whether adding consumer inflation forecasts improves the fit of Phillips Curve regressions.

Figure 1.1 displays consumer expectations of price inflation against professional expectations of GDP deflator inflation, both with a one- year forecast horizon. Consumer inflation expectations stabilized around 3% after the Volcker disinflation and remained consistent for about two decades; professional forecasts declined gradually throughout the 1980s and 1990s. Since 2000 the forecasts of professionals and households have diverged sharply. Professionals became firmly anchored at the Federal Reserve's implicit 2% target, while households' expectations remained elevated. Household expectations of inflation have been stable at 3% since the early 1980s, with some transitory deviations from 3% during 2006–2008. Divergence in inflationary expectations matters to the extent that firms follow one group or the other and to the extent that the monetary authority wishes to keep expectations anchored.

Table 1.6 shows the results of adding household inflation forecasts to the Phillips Curve. The Michigan Survey's forecast is for one year ahead, which introduces an MA(4) error term. I use instruments lagged 5 to 8 periods prior. The dependent variable is inflation in the nonfarm business sector. The professional forecast remains that of the GDP deflator. The slack measure is the filtered labor share. Results for the CBO output gap, not reported here, are similar. Samples begin in 1978:I, the first quarter in which Michigan Survey data are available. The model to be tested is:

$$\pi_t = \alpha\pi_{t-1} + \beta\pi_{t+1}^{SPF} + \gamma\pi_{t+1}^{MICH} + \lambda\varphi_t + \varepsilon_t \tag{1.11}$$

where underlying firm expectations are expressed using professional and household expectations, This specification need not be interpreted as firms literally melding together household and professional forecasts; it only indicates that firms' expectations are a latent variable, about which household and professional expectations provide imperfect signals.

The results are mixed. In column (1) of Table 1.6, the Michigan survey is the only measure of expectations. It enters with a coefficient near unity. The slack variable enters positively, but statistically insignificantly. Further, its point estimate (noisily estimated) is about half of that seen when using the SPF measure of inflation expectations. Column (2) introduces a lagged inflation term. Lagged inflation enters significantly, though the model is still heavily weighted towards the survey measure of expectations. The estimated coefficient on the labor share falls still further and remains imprecisely estimated. Column (3) specifies a purely forward-looking model with both professional and household forecasts. The coefficients on each expectations measurement are similar in magnitude; the coefficient on slack is small, enters with the wrong sign and is statistically insignificant. A similar pattern holds when the model with multiple expectations measures is allowed to incorporate a lagged dependent variable. Weights on the two expectations variables are about even, the slack coefficient remains insignificant, and past inflation enters insignificantly.

Perhaps the most interesting result of this "horse race" between professional and household forecasts is the large weight given to household forecasts. Focusing on the regression with the labor share and lagged dependent variable, household forecasts receive approximately the same weight as professional forecasts receive (both coefficients of around 0.50).

### 1.4.5 Stability Over Time

Overall, the battery of results presented so far indicate that the marginal cost Phillips Curve $\lambda$ is between 0.10 and 0.20. The coefficient on marginal cost is significantly related to inflation, once inflation expectations are operationalized with forecasts from professionals. Inflation is persistent even conditional on expectations and marginal cost, with an autoregressive coefficient of about one-third. Professional forecasts provide a better fit than household forecasts.

Next, I consider the stability of the parameter estimates over time. I run three tests. First, I run rolling estimates of the parameters of the hybrid NKPC with an expanding window, starting with 1981:I. Second, I run formal break tests for known break in 2000 and

2008. The former is the beginning of the secular decline of the labor share in the US, and the latter marks the post-Great Recession period. Third, I use sample splits to estimate the NKPC in selected subsamples of the 1968–2018 period.

Figure 1.2 graphs the labor share coefficient in an expanding window of the sample period. Throughout the estimated coefficient is positive and statistically significant. In the first two decades of the estimation sample, the coefficient value hovers around 0.4, an estimate that implies a relatively steep Phillips curve. After 2000, the estimated coefficient declines before stabilizing around 0.2 by 2010 and remaining there for the remainder of the sample period. Recall that these estimates are for a pre- filtered measure of the labor share, so this decline in the labor share coefficient are above and beyond those that would be seen if the (downward-trending) log-level of the labor share were used.

Figure 1.3 reports the coefficient on the lagged inflation term. It is generally positive and significant throughout the sample period. The coefficient on the lagged inflation term rises suddenly in the mid-2000s, somewhat after the decline in the log labor share. This increase brings the lagged inflation coefficient to about 0.3, so that lagged inflation is an economically meaningful component of inflation but remains about half of the coefficient on expected future inflation. The significance of this increase in lagged inflation coefficient will be explored quantitatively subsequently.

Benati (2008) has provided evidence that inflation persistence is dependent on the monetary regime. I find that the inflation persistence term in the New Keynesian Phillips Curve is particularly stable after 1990, settling at a value of 0.25 during most of the Great Moderation and rising to 0.3 in the Great Recession years. The model does not show any tendency for a reduction in inflation persistence, despite the increasing emphasis of monetary policymakers on inflation expectations management throughout the 1990s and 2000s. This is true despite the fact that survey expectations have been claimed to capture excess inflation persistence.

In addition to rolling estimation, I test for the presence of a single, discrete break in the marginal cost coefficient at given dates. I consider two candidate dates: 2000 and 2008. The first of these corresponds to the beginning of the sharp decline in the US labor share. Although I use the filtered labor share in my main estimates, and the CBO output gap is stationary by construction, testing for a break after 2000 reveals the presence of structural change beyond that which is facored out by filtering. Second, I test for a break in the marginal cost coefficient after 2008, the onset of the Great Recession. A structural break test at this date tests for a quantitatively detectable flattening of the Phillips curve since the Great Recession.

Table 1.7 and 1.8 display results for the NKPC as estimated with the filtered labor share. Table 1.7 tests for a break in the year 2000. The labor share coefficient remains positive, but the standard error widens. Two interaction terms are present: one for the constant, and one for the labor share. The constant term does display a shift after 2000. However, the interaction term on the labor share is not significant in either OLS or GMM estimation. Table 1.8 presents results for a break after the Great Recession. The results are similar to those in the post-2000 test. The coefficent on the labor share is measured less precisely.

Neither the interaction term on the constant nor the interaction term on the labor share enter significantly.

Tables 1.9 and 1.10 provide analogous results for the New Keynesian Phillips Curve as estimated with the CBO output gap as the slack variable. As before, neither the post-2000 nor post-2009 interaction terms are negative, indicating a flattening of the Phillips curve, but these coefficients are not quantitatively or statistically significant.

These results complement those in Mineyama (2018). Mineyama (2018) finds evidence for a decline in the slope of the Phillips curve after 2008 when using a structural break test and an output gap measure (detrended output), but not when using a marginal cost measure. The evidence I find with a formal break test is weaker. The difference between these results stems from a difference in the treatment of the constant term. I allow the intercept to drift after 2008, unlike Mineyama (2018), and find no statistically significant evidence of a flattening of the output gap parameter. My point estimates of the interaction term tend to be negative, but come with large standard errors. Specifying a fully interacted model – allowing for breaks in the intercept and all three coefficients of interest – delivers similar results. The coefficient on the labor share interacted with post-2000 and post-2009 dummies is of roughly equal magnitude and opposite sign as the main labor share coefficient, indicating a post-2009 slope of nearly zero, but is imprecisely estimated.

The third test I undertake consists of NKPC estimation across sample splits. I choose three subperiods. The early period is 1968-1983; the middle period is 1984-2007, corresponding to the Great Moderation; the late period is 2008- 2018, corresponding to the

Great Recession and subsequent recovery. Table 1.11 provides results for the full sample and these three splits, using the filtered labor share as the slack variable. The first two columns repeat the information from 1.1. The second two columns report results from the early subsample. As could be seen in 1.2, the marginal cost coefficient is about 0.5 and highly statistically significant. The expectations term enters positively and significantly with a coefficient above unity in the baseline specification and about 0.88 in the hybrid specification. The lagged inflation term is insignificant in the early sample.

The Great Moderation sample shows a deterioration of the NKPC to account for inflation dynamics. In both the forward-looking and hybrid specifications, the labor share enters with insignificantly and with the wrong sign. Survey expectations remain significant in the forward-looking specification, but enter with a point estimate that is half of its value in the early sample. Lagged inflation enters positively and significantly, with a point estimate of about 0.3. During the Great Moderation, the link between marginal cost and inflation is absent. This result also holds for the GDP deflator.

An even more bleak pattern emerges for the post-2009 period. Inflation since 2009 has essentially been white noise; neither survey expectations, marginal cost, or lagged inflation enter significantly. However, the sample size is small and inflation has shown little variation in the recovery from the Great Recession, and SPF expectations of inflation have also shown little variation since 2009, limiting the amount of signal that can be extracted from this subsample. Phillips curve logic appears to be of little aid in understanding inflation dynamics in the recent recovery. Through three sets of tests, I find that there is qualitative evidence for structural change in the NKPC coefficients. In rolling samples, the marginal cost coefficient drifts downward over time. Although structural break tests are unable to provide evidence for a break post-2000 or post-2009, the point estimates on a break are typically negative, indicating a noisily estimated decline in the marginal cost coefficient. Split-sample tests confirm that the marginal cost term's significance weakens considerably in both the Great Moderation and post-Great Recession subsamples.

## 1.5 Inflation Expectations and Recent Inflation Dynamics

We have a rich array of inflation expectations data: professional forecasts of the GDP deflator, professional forecasts of the Consumer Price Index, household forecasts of "price inflation" generally and, since 2002, financial market forecasts of CPI inflation as inferred from Treasury Inflation-Protected Securities. We have seen above that the behavior of the CPI and GDP deflator differ in some meaningful ways; the deflator tends to be more persistent and tends to be more tightly linked to marginal cost than the CPI. But regardless of the price index used, inflation expectations are a critical determinant of actual inflation. To the extent that the forecasts of households, professionals, and financial market participants agree, the distinction between them is primarily academic and technical, not policy-relevant.

However, since the Great Recession, expectations from these surveys have diverged. Households continue to have elevated inflation expectations and recent data indicates further upward pressure on their expectations. Professional forecasts, even only one quarter out, remain solidly anchored to the Federal Reserve's 2% target. However, inflation expectations as measured from TIPS breakeven rates have trended downward for several years, settling at a level below the Fed's target. This divergence in opinion matters for policymakers, who must sift through conflicting surveys to ascertain the underlying expectational pressures, be they stable (from households) or deflationary (from financial markets). Further, to the extent that forecasts embed individuals' expectations of future policy, persistent inflation expectations below the inflation target imply that the central bank's inflation target is becoming unanchored from below.

The results here show that changes in consumers' inflation forecasts enter significantly into changes in realized inflation, even more so than changes in professional forecasters' expectations. However, inflation expectations gleaned from TIPS spreads are available at much higher frequency. The degree to which weight should be placed on these indicators depends on two factors: the degree to which they feed into current inflation and the degree to which changes in each forecast reflects incoming information about news.

Inflation as inferred from breakeven TIPS rates changes daily and, to the extent that financial markets are informationally efficient, rapidly conveys changes in incoming information about future shocks. Household inflation expectations tend to move slowly, be more persistent, are seemingly require substantial effort to un-anchor from a 3% expectation. However, financial market participants and households may have differing information sets; the latter may be less quick to update their inflation forecast than the former, and the former may be more quick to incorporate news into their forecasts. The importance of consumer forecasts indicates that medium-term inflation management depends on the central bank's ability to effectively anchor consumer expectations; financial market expectations provide information of a different sort, namely about expectations of future developments.

## 1.6 Relationship to the Literature

This paper has investigated the degree to which survey expectations can deliver empirically reasonable Phillips curves and can remedy the shortcomings identified in rational expectations estimates of the Phillips curve. I have focused on the size and significance of the economic slack term, the degree to which the model rquires the inclusion of a lagged inflation term, and stability in the Phillips curve coefficients over time. Full-sample estimates are encouraging, with the survey expectations NKPC delivering significant and quantitatively meaningful estimates of the real variable coefficient across several specifications; however, the survey expectations NKPC is unable to reduce the model's dependence on lagged inflation. Estimation across sub-samples indicates that the weakness of Phillips curve during the Great Moderation and recovery from the Great Recession cannot be wholly mitigated by the incorporation of survey expectations. In this section I address links between the results in this paper and the literature, with a focus on results reported in Coibion, Gorodnichenko, and Kamdar (2018). That paper involves an extended discussion of the survey expectations NKPC, albeit estimated on CPI inflation rather than the GDP or NFB deflators, and using the CBO's unemployment gap rather than any proxy for marginal cost.

**Specification.** Traditional FIRE estimation of the New Keynesian Phillips Curve produced estimates in which the marginal cost coefficient was small and which featured excess inflation persistence. I find that the NKPC as estimated with survey expectations, in the full sample, delivers a larger marginal cost coefficient that is statistically significant throughout the sample range, albeit falling as the most recent decade of data is added. I find that lagged inflation continues to play a role in the model, and the weight on lagged inflation is qualitatively unchanged from Galı and Gertler (1999). I show that the lagged inflation term is more important in specifications with the nonfarm business deflator and the GDP deflator than it is in specifications with CPI inflation. A lagged term remains significant in NKPC estimation, but its weight is diminished relative to the weight given to survey expectations.

**Structural stability.** One proposed advantage of the survey expectations NKPC was its ability to deliver stable coefficients on marginal cost through different regimes. The hope was that survey expectations could absorb variation due to changes in the inflation target, information processing, etc, and yield stable Phillips curve slope coefficients in subsamples of interest. Coibion and Gorodnichenko (2015) and Coibion, Gorodnichenko, and Kamdar (2018), table 8, provides results to this effect when using CPI inflation with the CBO unemployment gap as the slack measure. In their tests, the Phillips curve flattens in later subperiods, though the slope is usually in the theoretically predicted direction. Survey expectations mitigate but do not eliminate the flattening of the Phillips curve since 2000. In addition, both my tests and the tests in Coibion and Gorodnichenko (2015) use filtered variables, so these tests deliver evidence of a flattening of the Phillips curve even above that which would be found when using the raw labor share. The specifications I consider provide additional evidence that the survey expectations NKPC has only limited ability to provide stable coefficients across subperiods. My main specification, a Phillips curve specified with the nonfarm business deflator and using filtered labor cost as the slack variable, flattens considerably over time, especially after the early pre-1985 period. Most of our positive evidence on the slope of the NKPC comes from pre-Great Moderation observations.

Coibion, Gorodnichenko, and Kamdar (2018) test formally for the null hypothesis of no structural break in the NKPC slope coefficient, and find insufficient evidence to reject in most specifications. However, they also find insufficient evidence to reject a zero slope coefficient in many subperiods, so this must be taken as mixed evidence. My table 1.7 through 1.10 provide complementary results, in that the slope coefficient cannot be said to change significantly over the post-2000 and post-2008 samples, but also that a Phillips curve specified with that level of flexibility also indicates a slope that is, at best, imprecisely estimated.

**Forecast accuracy.** Although this paper focuses on parameter estimation rather than on using the NKPC as a forecasting tool, the results of my tests for structural stability lend information on this issue. The ability of the NKPC to forecast inflation has deteriorated through the sample period. In split samples, the labor share, survey expectatiosn, and even lagged inflation have only modest ability to predict inflation uring the Great Moderation. In addition, inflation since 2008 seems to be nearly independent of all three variables in the NKPC.

**The decline of the labor share.** Throughout this paper, I have avoided taking a stand on the low-frequency movements in the labor share by filtering most slack variables prior to estimation. However, this low-frequency behavior can be important for assessing the quarterly-frequency properties of the data. Many factors outside the scope of the NKPC could contribute to drift in the average labor share, but movements in the average markup would affect both the average labor share (i.e., its low-frequency movements) and the slope of the Phillips curve (i.e., the high frequency movements) simultaneously. The short-run behavior of the NKPC along the transition path to a new steady-state with a different level of markup has not been sufficiently explored.

**Which expectations matter?** A plethora of inflation expectations data are available: household expectations, professional forecasts, Greenbook forecasts, breakeven inflation compensation from financial data, and even limited information from new firm surveys. Two related questions arise from this richness in data. First, whose expectations enter more strongly into empirical Phillips curve specifications? Second, which maps most closely to the theory? On the first question, my results complement Coibion, Gorodnichenko, and Kamdar (2018). Both that study and this one find that household expectations enter the Phillips curve with quantitatively large coefficient, even after controlling for professional forecasts. Indeed in a specification with both household and SPF forecasts, the point estimate on household expectations exceeds that of professional forecasters.

Coibion, Gorodnichenko, and Kumar (2018) reports results from a cross-section of New Zealand firms. Given that their sample is a short panel, running from 2013 to 2016, they cannot feasibly estimate Phillips curves with the data. Instead, they explore the dispersion in beliefs across firms and the relationship between firm expectations and the costs of gathering and processing information. They find that firm expectations look qualitatively more like the household expectations in the Michigan survey than the professional forecasts in the SPF. For one, firm managers on average over-estimate inflation, like US households but unlike US forecasters. Second, firm managers show a degree of cross-sectional dispersion that is more like households than it is professional forecasters. Third, they find that inattention and infrequent information updating characterizes many firms. These findings indicate that when assessing the empirical performance of the theoretical NKPC (as opposed to choosing a specficiation for forecasting), specifications with consumer expectations provide a closer analogue to firm expectations than do surveys of professional forecasters.

## 1.7 Extensions and Conclusions

I have found that survey-based expectations perform well in fitting the New Keynesian Phillips Curve to the data through the 1968-2018 period as a whole. In particular, survey expectations enter with correct sign, reasonable magnitude, and high statistical significance. The slack coefficient also enters with the correct sign, reasonable magnitude, and statistically significantly. The post-1968 data do not reject an elasticity of inflation to marginal cost between 0.10 and 0.20. Estimates of the forward-looking component are high, typically around unity in purely forward-looking specifications and over 0.5 in hybrid specifications. The Phillips Curve based around the nonfarm business deflators and GDP deflator admits a lagged inflation term, with estimates of the lagged inflation coefficient between 0.25 and 0.35. The CPI Phillips Curve consistently rejects a lagged inflation term. However, results are weaker for individual subsamples, and most of the evidence for a positive, significant marginal cost coefficient comes from the earliest parts of the sample.

We have seen that survey expectations can play a useful role in fitting economic models at the single-equation level, and fits the data better than usual estimation techniques that rely strongly on rational expectations. This result suggests a research direction that embeds non-rational, or at least non full-information, expectations in a full-blown dynamic macroeconomic model. However, including nonrational expectations in a fully-specified DSGE model requires one to take a stand on exactly how expectations adjust and on the precise way in which rationality fails. There are many ways for full- information rational expectations to fail; these details are left unstated in the single-equation results I presented above but must be specified in a full- blown model environment.

In this paper, I have provided estimates of the Phillips Curve's parameters and the degree to which theoretical relationships among inflation, marginal cost, and inflation expectations are borne out in the data. Rather than assume rational expectations, I assume firms' inflation expectations could be proxied for by surveys of expectations from the Survey of Professional Forecasters and the Michigan Survey of Consumers. I found that these survey data provide coefficient estimates of structural parameters that are statistically significant, fairly stable over time, and economically reasonable. I showed that the dependence of inflation on marginal cost has lessened since the Great Recession and that inflation persistence has increased. The exact measure of inflation used in estimation matters substantively. A mixture of professional and household forecasts provides the best fit, with household forecasts entering with higher weight than professional forecasts.

## 1.8 Tables and Figures

Table 1.1: Baseline specification

| Dependent: NFB inflation | (1) | (2) |
|---|---|---|
| Lagged inflation | | $0.326^{**}$ |
| | | (0.078) |
| Filtered labor share | $0.188^{*}$ | 0.180 |
| | (0.088) | (0.083) |
| SPF expectations | $1.125^{***}$ | $0.751^{***}$ |
| | (0.050) | (0.092) |
| Constant | $-0.764^{**}$ | $-0.486^{*}$ |
| | (0.202) | (0.201) |
| $\pi^e_{t+1} + \pi_{t-1}$ | 1.125 | 1.077 |
| | (0.05) | (0.05) |
| N | 203 | 203 |
| RMSE | 1.424 | 1.347 |
| Resid. Autocorr. | $0.350^{***}$ | 0.080 |
| | (0.066) | (0.070) |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 1.2: Alternate measures of slack, 1968-2018

| | Dependent: NFB inflation | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Lagged inflation | | 0.309*** | | 0.241** |
| | | (0.079) | | (0.076) |
| Log labor share | -0.022 | -0.011 | | |
| | (0.022) | (0.022) | | |
| CBO output gap | | | 0.148** | 0.109* |
| | | | (0.048) | (0.048) |
| SPF expectations | 1.072*** | 0.750*** | 1.054*** | 0.823*** |
| | (0.059) | (0.100) | (0.053) | (0.090) |
| Constant | 9.706 | 4.714 | -0.422* | -0.366* |
| | (10.166) | (10.020) | (0.185) | (0.175) |
| $\pi^e_{t+1} + \pi_{t-1}$ | 1.07 | 1.06 | 1.05 | 1.06 |
| | (0.06) | (0.06) | (0.05) | (0.05) |
| N | 203 | 203 | 203 | 203 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 1.3: Alternate measures of slack, 1968-2000

| | Dependent: NFB inflation | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Lagged inflation | | 0.169 | | 0.230** |
| | | (0.088) | | (0.083) |
| Log labor share | 0.352*** | 0.322*** | | |
| | (0.080) | (0.075) | | |
| CBO output gap | | | 0.235*** | 0.237*** |
| | | | (0.062) | (0.059) |
| SPF expectations | 1.084*** | 0.887*** | 1.253*** | 0.992*** |
| | (0.061) | (0.120) | (0.062) | (0.113) |
| $\pi^e_{t+1} + \pi_{t-1}$ | 1.08 | 1.06 | 1.25 | 1.22 |
| | (0.06) | (0.06) | (0.06) | (0.06) |
| N | 129 | 129 | 129 | 129 |
| Resid. Autocorr. | 0.235** | 0.062 | 0.292** | 0.025 |
| | (0.086) | (0.088) | (0.084) | 0.088 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 1.4: Alternate measures of inflation, labor share

| | GDP deflator | | CPI inflation | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Lagged inflation | | 0.474*** | | -0.075 |
| | | (0.071) | | (0.060) |
| Filtered labor share | 0.192** | 0.130** | -0.095 | -0.096 |
| | (0.060) | (0.049) | (0.094) | (0.097) |
| SPF expectations | 0.998*** | 0.495*** | 0.782*** | 0.837*** |
| | (0.038) | (0.078) | (0.144) | (0.161) |
| Constant | -0.064 | 0.052 | 0.486 | 0.549 |
| | (0.128) | (0.120) | (0.431) | (0.425) |
| $\pi^e_{t+1} + \pi_{t-1}$ | 0.998 | 0.97 | 0.782 | 0.762 |
| | (0.04) | (0.04) | (0.144) | (0.14) |
| N | 203 | 203 | 148 | 148 |
| Resid. Autocorr. | 0.435*** | -0.006 | -0.050 | 0.016 |
| | (0.063) | (0.070) | (0.081) | (0.081) |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 1.5: Alternate measures of inflation, CBO gap

| | GDP deflator | | CPI inflation | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Lagged inflation | | 0.468*** | | -0.083 |
| | | (0.071) | | (0.059) |
| CBO output gap | 0.161*** | 0.124*** | 0.079 | 0.084 |
| | (0.037) | (0.033) | (0.072) | (0.073) |
| SPF expectations | 1.019*** | 0.521*** | 0.783*** | 0.835*** |
| | (0.038) | (0.081) | (0.151) | (0.171) |
| Constant | 0.057 | 0.144 | 0.634 | 0.748 |
| | (0.137) | (0.125) | (0.467) | (0.459) |
| $\pi^e_{t+1} + \pi_{t-1}$ | 1.02 | 0.99 | 0.78 | 0.75 |
| | (0.04) | (0.04) | (0.15) | (0.15) |
| N | 203. | 203 | 148 | 148 |
| Resid. Autocorr. | 0.411*** | -0.034 | -0.068 | 0.006 |
| | (0.064) | 0.070 | 0.081 | 0.081 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 1.6: NKPC stimates using both SPF and Household Forecasts

| | Dependent: NFB inflation | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Lagged inflation | | 0.300*** | | 0.106 |
| | | (0.058) | | (0.065) |
| Filtered labor share | 0.096 | 0.054 | -0.031 | -0.056 |
| | (0.098) | (0.097) | (0.100) | (0.099) |
| SPF expectations | | | 0.568*** | 0.490*** |
| (one quarter ahead) | | | (0.079) | (0.091) |
| Household expectations | 1.100*** | 0.820*** | 0.608*** | 0.573*** |
| (one year ahead) | (0.074) | (0.092) | (0.095) | (0.100) |
| Constant | -1.422*** | -1.166*** | -1.393*** | -1.292*** |
| | (0.249) | (0.244) | (0.212) | (0.223) |
| $\pi^e_{t+1}$ | | | 1.18 | 1.06 |
| | | | (0.06) | (0.10) |
| $\pi^e_{t+1} + \pi_{t-1}$ | 1.10 | 1.12 | 1.18 | 1.17 |
| | (0.7) | (0.07) | (0.06) | (0.06) |
| N | 166 | 166 | 166 | 166 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 1.7: Post-2000 break tests with labor share NKPC

| | (1) OLS | (2) GMM |
|---|---|---|
| Lagged inflation | 0.281*** | 0.178 |
| | (0.068) | (0.093) |
| Filtered labor share | 0.134 | 0.239 |
| | (0.079) | (0.172) |
| SPF expectations | 0.913*** | 1.266*** |
| | (0.103) | (0.208) |
| post2000 | 0.588* | 2.556** |
| | (0.250) | (0.902) |
| Labor share × post2000 | -0.213 | 0.177 |
| | (0.126) | (0.427) |
| Constant | -1.182*** | -2.760*** |
| | (0.301) | (0.827) |
| N | 203 | 203 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

post2000 is an indicator =1 after 2000

Table 1.8: Post-2008 break tests with labor share NKPC

| | (1) | (2) |
|---|---|---|
| | OLS | GMM |
| Lagged inflation | $0.313^{***}$ | $0.338^{***}$ |
| | (0.068) | (0.071) |
| Filtered labor share | 0.060 | 0.198 |
| | (0.070) | (0.141) |
| SPF expectations | $0.813^{***}$ | $0.701^{***}$ |
| | (0.095) | (0.119) |
| post2008 | 0.291 | -0.610 |
| | (0.273) | (0.696) |
| Labor share $\times$ post2008 | -0.105 | -0.278 |
| | (0.152) | (0.422) |
| Constant | $-0.744^{**}$ | -0.262 |
| | (0.238) | (0.415) |
| N | 203 | 203 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

post2008 is an indicator =1 after 2008

Table 1.9: Post-2000 break tests with CBO gap NKPC

| | (1) OLS | (2) GMM |
|---|---|---|
| Lagged inflation | 0.289*** | 0.183* |
| | (0.067) | (0.090) |
| CBO gap | 0.174** | 0.172 |
| | (0.052) | (0.127) |
| SPF expectations | 0.930*** | 1.286*** |
| | (0.102) | (0.223) |
| post2000 | 0.505 | 2.545** |
| | (0.265) | (0.905) |
| CBO gap × post2000 | -0.160 | 0.066 |
| | (0.094) | (0.340) |
| Constant | -1.080*** | -2.640** |
| | (0.296) | (0.857) |
| N | 203 | 203 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

post2000 is an indicator =1 after 2000

Table 1.10: Post-2008 break tests with CBO gap NKPC

| | (1) | (2) |
|---|---|---|
| | OLS | GMM |
| Lagged inflation | $0.316^{***}$ | $0.333^{***}$ |
| | (0.066) | (0.068) |
| CBO gap | $0.162^{**}$ | 0.151 |
| | (0.050) | (0.137) |
| SPF expectations | $0.837^{***}$ | $0.758^{***}$ |
| | (0.093) | (0.127) |
| post2008 | 0.228 | -0.485 |
| | (0.342) | (1.607) |
| CBO gap × post2008 | -0.159 | -0.164 |
| | (0.115) | (0.659) |
| Constant | $-0.690^{**}$ | -0.336 |
| | (0.233) | (0.424) |
| N | 203 | 203 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

2008 is an indicator =1 after 2008

Table 1.11: Sample splits, labor share

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | (1968-2018) | | (1968-1983) | | (1984-2008) | | (2008-2018) | |
| Lagged inflation | | 0.326*** | | 0.223 | | 0.297** | | 0.191 |
| | | (0.068) | | (0.124) | | (0.101) | | (0.154) |
| Filtered labor share | 0.188* | 0.180* | 0.494** | 0.471** | -0.083 | -0.076 | 0.068 | 0.050 |
| | (0.088) | (0.083) | (0.163) | (0.158) | (0.100) | (0.094) | (0.150) | (0.146) |
| SPF expectations | 1.125*** | 0.751*** | 1.127*** | 0.876*** | 0.682*** | 0.479*** | 0.591 | 0.428 |
| | (0.050) | (0.092) | (0.127) | (0.187) | (0.107) | (0.124) | (0.551) | (0.553) |
| Constant | -0.764*** | -0.486* | -0.583 | -0.475 | 0.146 | 0.100 | 0.431 | 0.443 |
| | (0.202) | (0.201) | (0.782) | (0.760) | (0.312) | (0.298) | (1.048) | (1.022) |
| N | 203 | 203 | 61 | 61 | 96 | 96 | 46 | 46 |
| RMSE | 1.424 | 1.347 | 1.948 | 1.889 | 0.952 | 0.909 | 1.046 | 1.026 |

Dependent: nonfarm business sector inflation

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

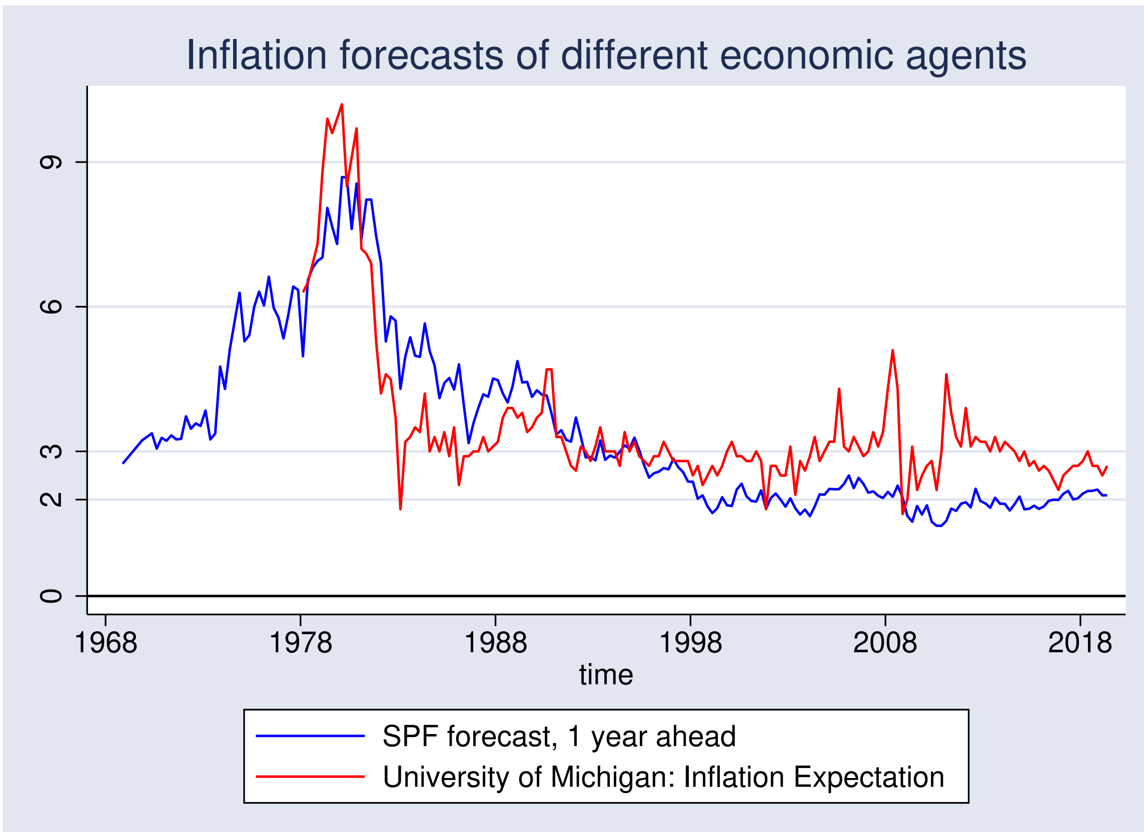


Figure 1.1: SPF expectations and household expectations of inflation

Note: SPF are expectations of one-year-ahead GDP deflator inflation, while household expectations are of "the general rise in prices" one year ahead. The 2% inflation target is indicated for reference.

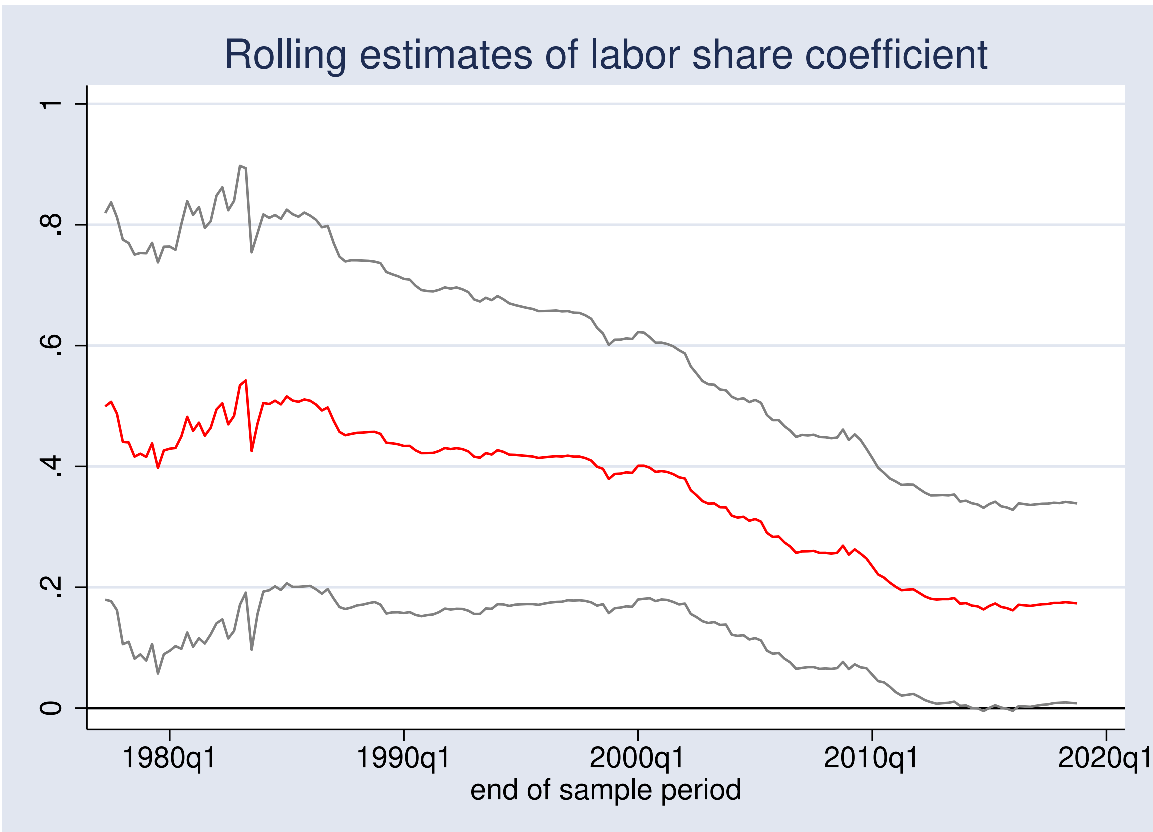


Figure 1.2: Estimates of Phillips Curve coefficients over time

Note: Gap variable is the filtered labor share. Coefficients are estimated from 1968 to quarter $t$ for $t = 1985$ to 2015.

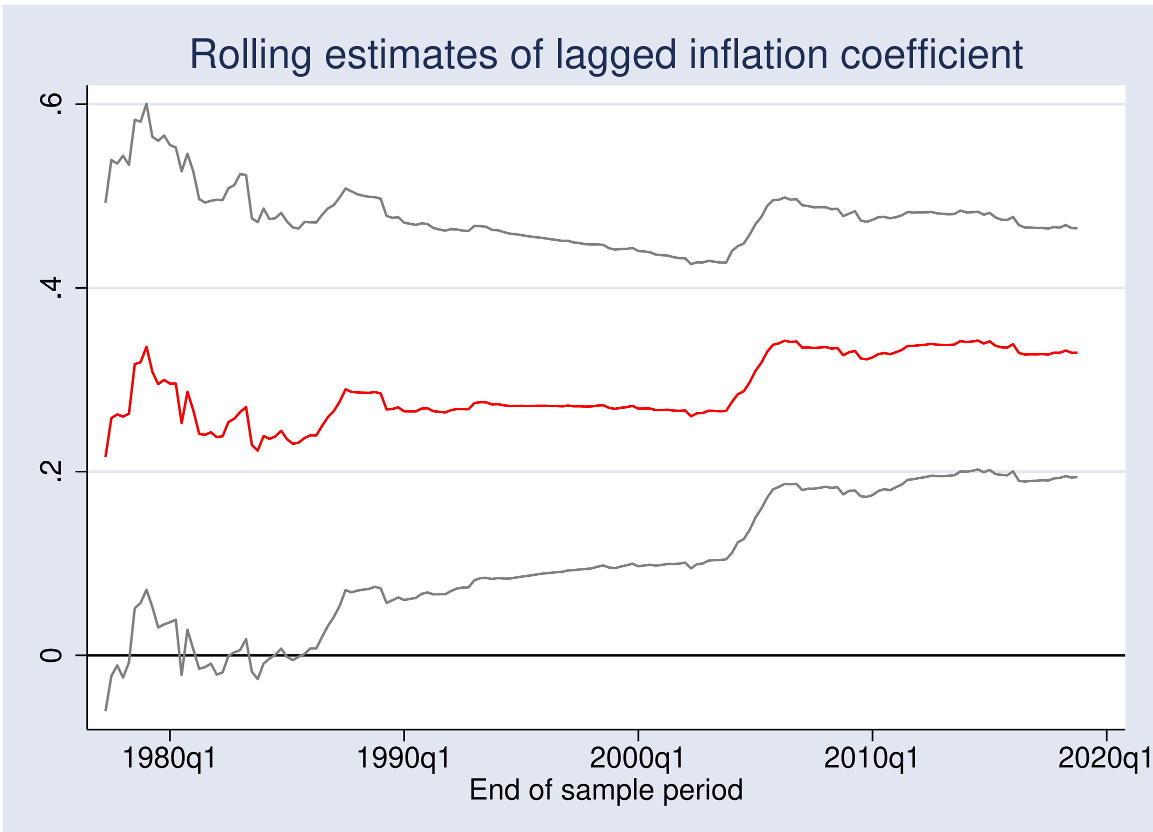


Figure 1.3: Estimates of Phillips Curve coefficients over time

Note: Gap variable is the filtered labor share. Coefficients are estimated from 1968 to quarter $t$ for $t$ = 1985 to 2015.

# Chapter 2

# Estimating Structural Breaks in Impulse Response Functions via the Local Projection Estimator

## 2.1 Introduction

Impulse response functions trace out the effects of a shock on the endogenous variables of a dynamic model. An important question is whether these functions are stable over time. Change in impulse response functions indicates change in the economy's dynamic response to shocks over time. Such change could be due to strctural change in private-sector organization and behavior, or due to changes in policymaking that affect the economy's response to shocks. Candidate times in the past have been suggested: after the early

1970s breakup of the Bretton Woods regime, after the 1973 and 1979 oil crises, due to improvements in policymaking during the early 1980s. The timing of the break is often unknown.

Impulse response functions are typically estimated after first estimating an underlying model, usually some form of a vector autoregression. Jordà (2005) introduced the local projection estimator for impulse response functions. This method produces estimates of the impulse response coefficients directly, without first specifying an underlying VAR. Estimates of impulse response coefficients over time and across impulse/response combinations can be estimated separately, or they may be estimated jointly. When the data–generating process is a VAR and the local projections are estimated with the appropriate lag length, the local projection estimates are consistent and only slightly less efficient than IRF coefficient estimates based on the VAR. When the data–generating process is not a VAR, the local projection coefficients provide additional flexibility that is not provided when estimating from a VAR of similar lag length. Regardless, the local projection coefficients are estimated directly via a simple regression procedure. Thus, instead of requiring delta-method standard errors, the standard errors of the local projection estimator can be obtained from the usual multiple-equation least squares method with a heteroskedasticity correction. The ability to directly estimate IRF coefficients by least squares opens up the possibility of applying structural break tests directly to the IRF coefficients.

Meanwhile, macroeconomists have long been interested in tests for structural change in time–series models. Andrews (1993) has provided tests for structural breaks in the coeficients of a single–equation time–series model in the presence of an unknown break point. His strategy involves testing for a break by taking the supremum of Wald, LR, and/or LM test statistics for each candidate break date; his paper derives the distribution of these supremum statistics under the null of no break, and provides critical values.

The present paper is concerned with tests of parameter instability in impulse–response functions. The IRFs are estimated directly via the local projection estimator, the break date is treated as unknown, and the unknown break date is found via a multiple–equation analogue to the tests in Andrews (1993). In particular, I focus on dynamic multipliers, that is, the response functions to an innovation in an exogenous variable in the proposed model. Since the local projection method generates estimates of the impulse–response and dynamic–multiplier coefficients directly, the tests for a structural break can also be applied directly. Calculation of standard errors is simplified and inference is more straightforward than the alternative, which would be to first estimate a vector autoregression, then compute estimated impulse–response coefficients from the estimated VAR, then compute the standard errors for the IRF coefficients through the delta method. By using local projections, instability in the parameters of direct interest – the impulse response and dynamic multiplier coefficients – are tested directly. I first show that the standard least–squares estimator provides adequate coverage of the dynamic–multiplier coefficients when estimated by local projections. I then turn to the question of structural change, deriving the multiple–equation analogue to the tests proposed in Andrews (1993).

I then turn to an examination of structural change in impulse response functions in three applied settings. First, I examine the response of real economic activity to estimated monetary shocks using the Romer and Romer (2004) Federal funds shocks. I find evidence of parameter instability in the dynamic multiplier coefficients, with the break date falling in the late 1970s or early 1980s and a point estimate of the first quarter of 1980. When estimating dynamic multipliers in each subperiod separately, I find that the result of large effects of monetary policy on real activity rests primarily on the responses experienced in the early period. Monetary policy shocks lead to a large, persistent decline in economic activity during the early sample, but there is little evidence of real effects of monetary policy using these shocks in the late period. Second, I investigate instability in the response of real economic activity to tax shocks using the Romer and Romer (2010) exogenous tax shock series. I find no evidence for structural change in these responses. Third, I investigate instabilit in the response of real economic activity to oil price shocks using the Hamilton (1996) series on net oil price changes. There is little evidence for structural change in the response of industrial production to oil shocks using this measure.

The rest of this paper is organized as follows. Section 2 reviews the construction of impulse response coefficients by both traditional VAR and local projection methods, then derives the local projection dynamic multiplier estimator that will be used in the following sections. Section 3 reviews the theory of tests of parameter instability in a single–equation context, then derives the test of parameter instability using local projection dynamic multipliers. Section 4 applies the parameter instability test to the monetary shocks of Romer and Romer (2004), the tax shocks of Romer and Romer (2010), and the oil price shocks of Hamilton

(1996). Section 5 concludes.

## 2.2 The local projection estimator

### 2.2.1 Local projection estimates of impulse response coefficients

This section reviews the local projection estimator of impulse response functions and contrasts it with the VAR estimates of impulse response functions.

The model consists of a $k \times 1$ vector endogenous variables $\mathbf{y}_t$, which are written as a function of the first $p$ lags,

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_{t-p} \mathbf{y}_{t-p} + \mathbf{u}_t \tag{2.1}$$

Estimates of the model's parameter matrices can be obtained by least squares. Assume that the roots of the autoregressive polynomial lie outside the unit cricle. Then we can invert the model, writing

$$\begin{aligned} \mathbf{y}_t &= (\mathbf{I} - \mathbf{A}_1 - \cdots - \mathbf{A}_p)^{-1} \mathbf{u}_t \\ &= \sum_{s=0}^{\infty} \mathbf{\Phi}_s \mathbf{u}_{t-s} \end{aligned}$$

The latter is the VMA($\infty$) form of the model. The $k \times k$ matrices $\mathbf{\Phi}_i$ are the vector moving average coefficients. These are the object of interest. The impulse response functions to innovations in $\mathbf{u}_t$ can be read off of the entries in the $\mathbf{\Phi}_i$. Each entry of $\mathbf{\Phi}_i$ is a function of the underlying model parameters in the $\mathbf{A}_i$ matrices. As such, asymptotically valid standard errors for the elements of $\mathbf{\Phi}_i$, and hence for the impulse response coefficients, can be calculated by the delta method.

The local projection estimator instead runs a collection of regressions. For each horizon $h = 1, 2, \dots, H$, estimate the $h-$period ahead regression

$$\mathbf{y}_{t+h} = \mathbf{A}_1^{h+1}\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p^{h+1}\mathbf{y}_{t-p} + \mathbf{v}_{t+h} \tag{2.2}$$

The timing convention is made so that the matrix $\mathbf{A}_1^h$ contains the estimated impulse response coefficients for $h$ steps ahead to shocks $\mathbf{u}_0$ . That is, $\mathbf{A}_1^h$ is an estimator for $\mathbf{\Phi}_s$. Local projection coefficients may be estimated equation–by–equation, step–by–step, or the system may be stacked together and the impulse response coefficients can be estimated jointly. For estimates of the impulse–response coefficients and their standard errors, the equation–by–equation, step–by–step approach is sufficient. For hypothesis testing involving coefficients across multiple equations, a joint estimation procedure is necessary.

The local projections are consistent for the true IRFs when the true model is a VAR and the lag length used in the local projections is the same as the data–generating process. Under these conditions, the local projection IRFs can be considered an alternative method for calculating the impulse response coefficients, one which offers parameter estimates and standard errors for joint inference that can be obtained through standard seemingly–unrelated regression procedures.

Even under the ideal conditions, it is known that the standard errors for the local projection impulse response functions that would appear from linear regression are not valid for inference. The error term at step $h$, $\mathbf{v}_{t+h}$, is a moving average function of the underlying errors between $t$ and $t+h$. Since $\mathbf{v}_{t+h}$ is a moving average of underlying errors, the standard errors on impulse response coefficients $\mathbf{A}_1^h$ are too small. Jordà (2005) recommends applying a Newey–West procedure to correct the standard errors. Once this correction is applied, the covariance matrix of the $\mathbf{A}_1^h$ is suitable for single and joint hypothesis testing. Inference on the impulse response coefficients can then be performed without resorting to delta–method, asymptotic estimates of the variance of the IRF coefficients that come with VAR estimation. I will exploit this fact extensively below.

One additional point of departure between the traditional impulse response functions and the local projection impulse response functions is the effective sample size available to each estimator. In the vector autoregression approach, the researcher loses $p$ periods of the data as a priming sample, leaving $T-p$ periods available for estmation of the VAR coefficients. By contrast, the collection of local projection regressions loses $H$ periods at the end of the sample in addition to the $p$ periods in the beginning of the sample, leaving only $T-p-H$ periods available for estimation. The loss of $H$ periods can be significant for datasets with small sample sizes, and increases with the number of periods ahead for which one wishes to construct the responses. If five years worth of responses are to be generated, one loses 20 periods in quarterly data and 60 periods in monthly data.

Local projection coefficients are estimators of the reduced-form impulse response functions. These coefficients are not always those that are of most interest to economists. We are often interested instead in orthogonalized impulse response functions, or more broadly in impulse responses as identified through structural modifications to the VAR. Jordà (2005) provides limited guidance on translating local projection coefficients into a structural analogue. The suggestion is to consider the structural shock of interest to be a linear combination of reduced-form shocks $\mathbf{u}_0$. These reduced-form shocks can be, for example, the columns of the Cholesky factor of $\boldsymbol{\Sigma}_u$, which are the impulses associated with the orthogonalized IRFs. However, he treats these impulses as fixed, leading to structural IRFs whose standard errors are too small. Rather than confront the important issue of structural shock extraction directly, this paper instead turns to the case in which shocks are observed. This case is not as limiting as it might seem, for recently much progress has been made in constructing observable series for many kinds of macroeconomic shocks.

### 2.2.2 Local projection estimates of dynamic multipliers

One strand of work has focused on producing time–series of observable macroeconomic shocks. Romer and Romer (2004) produced a series of monetary shocks measured as surprise movements in the Federal funds rate relative to Greenbook forecasts. Romer and Romer (2010) produced a series for tax shocks. Basu, Fernald, and Kimball (2006) produced a series for utilization-adjusted productivity shocks. Hamilton (1996) produced a series for net oil price increases, which he interprets as oil shocks. These shocks are used in applied settings to trace out the macroeconomic effects of fiscal, monetary, and technology shocks;

see Ramey (2016). Both VAR and local projection methods have been used to estimate the effects of these shocks on economic outcomes of interest. Impulse–response functions are typically calculated for these shocks in a single–equation context in which the shock series and its lags appear along with lags of the variable of interest. These shocks are designed to be independent of economic variables of interest. As such, there is room to consider the case of dynamic multipliers that trace out the effect of an innovation to one of these shocks on macroeconomic variables of interest.

This section describes the estimation of dynamic multipliers by local projections with a focus on the calculation of valid confidence intervals for the multipliers. As before, the model consists of endogenous variables $\mathbf{y}_t$; to these variables we add a $m \times 1$ vector of exogenous variables $\mathbf{x}_t$. The $\mathbf{y}_t$ process is driven by $\mathbf{x}_t$ and the vector of (possibly correlated) errors $\mathbf{u}_t$. By assumption $\{\mathbf{x}_t, \mathbf{u}_t\}$ are assumed to be independent. This assumption allows us to consider the $\mathbf{x}_t$ in isolation, and allows us to ignore any correlation among the $\mathbf{u}_t$.

We are interested in the dynamic multiplier coefficients. These are defined as the path of the endogenous variables $\mathbf{y}_t$ after a one–time, one–unit shock to one of the exogenous variables $\mathbf{x}_t$,

$$\mathbf{d}(s) = E(\mathbf{y}_{t+s}|\Delta\mathbf{x}_t = 1) - E(\mathbf{y}_{t+s}|\Delta\mathbf{x}_t = 0) \tag{2.3}$$

where $\Delta\mathbf{x}_t = 1$ denotess a unit vector in the direction of one of the components of $\mathbf{x}_t$. The impact period corresponds to $s = 0$ and the response on impact may be nonzero.

The traditional method follows a pattern similar to the VAR discussed above. A vector autoregression model augumented with exogenous variables is first estimated,

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}_0\mathbf{x}_t + \cdots + \mathbf{B}_l\mathbf{x}_{t-l} + \mathbf{u}_t \tag{2.4}$$

or more compactly,

$$\mathbf{y}_t = \mathbf{A}(L)\mathbf{y}_{t-1} + \mathbf{B}(L)\mathbf{x}_t + \mathbf{u}_t \tag{2.5}$$

As before, estimates of elements of the $k \times k$ coefficient matrices $\mathbf{A}_i$ and the $k \times m$ coefficient matrices $\mathbf{B}_j$ and their standard errors can be estimated by least squares. From the estimated model, the final form of the system can be derived

$$\mathbf{y}_t = \mathbf{D}(L)\mathbf{x}_t + \mathbf{\Phi}(L)\mathbf{u}_t \tag{2.6}$$

with $\mathbf{D}(L) = (\mathbf{I} - \mathbf{A}(L))^{-1}\mathbf{B}(L)$ and $\mathbf{\Phi}(L) = (\mathbf{I} - \mathbf{A}(L))^{-1}$. The dynamic multipliers of interest can be read off of elements of $\mathbf{D}(L)$. The elements of the $\mathbf{D}(L)$ matrices are functions of the underlying parameters in $\mathbf{A}_i$ and $\mathbf{B}_j$, so standard errors for elements of $\mathbf{D}(L)$ can be computed by the delta method. One estimates the VARX coefficients 2.5, then computes dynamic multipliers via 2.6, then computes standard errors either through the delta method or a bootstrap procedure.

The local–projection estimator with exogenous variables estimates the dynamic multiplier coefficients directly. Rather than running a single VARX and inverting its coefficients, consider running a collection of $H$ regressions, $h = 0, 1, 2, \ldots, H$:

$$\mathbf{y}_{t+h} = \mathbf{A}_1^{h+1}\mathbf{y}_{t-1} + \cdots + \mathbf{A}_k^{h+1}\mathbf{y}_{t-k} + \mathbf{B}_0^h\mathbf{x}_t + \cdots + \mathbf{B}_p^h\mathbf{x}_{t-p} + \mathbf{v}_{t+h} \tag{2.7}$$

The collection of matrices $\mathbf{B}_0^h$ contain the dynamic multiplier coefficients, just as the collection of matrices $\mathbf{A}_1^h$ contain the impulse response function coefficients. The remaining question is whether the standard errors on $\mathbf{B}_0^h$ have proper coverage. The impulse response coefficients are collected in $\mathbf{A}_1^h$, but their standard errors must be corrected with a HAC adjustment. The same is not true for the dynamic multiplier estimates; under strong exogeneity, their standard errors need not be corrected.

Consider the parameter estimates and their standard errors produced by estimating the local projection dynamic multipliers jointly via seemingly unrelated regression. The $h$-period-ahead regression involves an error term $\mathbf{v}_{t+h}$. This $\mathbf{v}_{t+h}$ error term is a function of the underlying period error terms from time $t$ to time $t+h$ and the exogenous variables from time $t$ to time $t+h$. When the exogenous variables are a white noise process, these elements are uncorrelated to the regressors, which are dated $t$ to $t-p$ and $t-k$. As such, unlike the impulse response coefficients, the moving–average error structure does not pose a problem for inference on dynamic multiplier coefficients by the strong exogeneity of $\mathbf{x}_t$. The SUR standard errors provide appropriate inference for the dynamic multiplier coefficients; in particular, they do not require a HAC correction. This feature is attractive, but relies heavily on strong exogeneity of the $\mathbf{x}_t$ process.

The above argument rests on the strong exogeneity of the variables $\mathbf{x}_t$. In the present context, this restriction is reasonable in two senses. First, if the process $\mathbf{x}_t$ actually follows an autoregressive process (for example), then it is no longer truly exogenous; the innovations to that autoregressive process are instead the exogenous variables, and $\{\mathbf{x}_t, \mathbf{y}_t\}$ are instead jointly determined in a restricted VAR where $\mathbf{x}_t$ is ordered first and the coefficients on lagged $\mathbf{y}_t$ in the $\mathbf{x}_t$ equation are set to zero. If that is the case, then the proper model is a joint VAR of $\{\mathbf{x}_t, \mathbf{y}_t\}$, not a model with exogenous $\mathbf{x}_t$. Second, as a practical matter, the main application considered here uses the shock series designed to provide observable counterparts to innovations in monetary policy, fiscal policy, productivity, and other macroeconomic shocks. By design, innovations to these series ought to be unforecastable. I explore the extent to which the shocks derived in the literature are forecastable below.

### 2.2.3 Monte Carlo evidence

This first set of experiments explores the coverage properties of the local– projection estimator with exogenous variables. The data–generating process is a VARX model with two lags in the endogenous variables and current and two lags of the exogenous variables. The true dynamic multiplier functions are calculated using the data–generating process. Then, a local projection is estimated using the same lag structure as the DGP. The local projection estimates of the dynamic multiplier functions are then compared to the true values. The local projection dynamic multipliers are calculated by estimating the full system jointly via seemingly unrelated regression, and standard errors are calculated using standard SUR formulas. I simulate 800 periods for each estimation, and perform 1,000 replicates. Of interest are the coverage properties of the local projections.

The data generating process consists of two endogenous variables $y_1$ and $y_2$, two exogenous variables $x_1$ and $x_2$, and two uncorrelated errors $e_1$ and $e_2$. The DGP is

$$
\begin{aligned}
\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} &= \begin{bmatrix} 0.7 & 0.1 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.2 & -0.1 \\ 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} \\
&+ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + \begin{bmatrix} -0.3 & 0.5 \\ 1.2 & -2 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ 0.3 & -0.4 \end{bmatrix} \begin{bmatrix} x_{1,t-2} \\ x_{2,t-2} \end{bmatrix} \\
&+ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}
\end{aligned}
$$

A sample length of 800 was chosen based on the typical length of U.S. monthly macroeconomic time series, which typically cover 65 years of monthly data (780 periods).

Of interest are the properties of the dynamic multiplier coefficients. First I compute the true dynamic multipliers from the data–generating process. Then, for each horizon $h = 0$ to $h = 16$, I run a local projection and store the resulting $\mathbf{B}_1^h$ and its standard errors. I run 1,000 such simulations.

For each outcome variable $y_{i,t}$, the dynamic multipliers are the coefficients $\{b_0^1, \ldots, b_0^h\}$ in the collection of regressions

$$\begin{aligned} y_{it} &= \mathbf{a}(L)\mathbf{y}_{t-1} + \mathbf{b}_0^0 x_t + \mathbf{b}(L)x_{t-1} + v_{t+h} \\ &\vdots \\ y_{i,t+H} &= \mathbf{a}(L)\mathbf{y}_{t-1} + \mathbf{b}_0^H x_t + \mathbf{b}(L)x_{t-1} + v_{t+h} \end{aligned} \tag{2.8}$$

Dynamic multipliers at horizon $h$ are the coefficients in the vector $\mathbf{b}_0^h$. Dynamic multipliers for multiple outcome variables can be calculated by stacking each collection of regressions together. Alternatively, dynamic multipliers can be estimated outcome–by–outcome, horizon–by–horizon; for each outcome–horizon pair, the $b_{i,h}^0$ and its estimated standard error are saved.

I provide first a summary image. Figure 2.1 provides true dynamic multipliers from the data–generating process, along with the dynamic multipliers calculated with local projections and two–standard-error bands associated with the local projection estimator. The mean local projection estimator is indistinguishable from the truth.

Tables 3.1 and 3.2 provide the true dynamic multiplier coefficients, the average local–projection dynamic multiplier coefficients, their standard errors, and the rejection rate across the 1000 simulations. From the table we see that the dynamic multiplier estimates are slightly attenuated in small samples. This result is not entirely unexpected, as attenuation bias is not uncommon for the least–squares estimator in autocorrelated data. The rejection rates are satisfactory, with nearly all coming close to the true nominal size of 0.05. In samples of this size, the coefficient variances reported by seemingly–unrelated regression provides an adequate basis for inference. Note again that no correction to the standard errors is needed. This is useful because the impulse response coefficients do require an adjustment due to the moving–average nature of the error term. This correction complicates inference for parameters across steps, which we will need in the next section. Such a concern is not an issue with the dynamic multiplier coefficients.

## 2.3 Tests for structural change

### 2.3.1 Tests for parameter instability in a single equation

Andrews (1993) introduced Lagrange multiplier, likelihood ratio, and Wald tests for parameter instability in a single–equation time–series model in which the break date was unknown. To fix ideas, consider a single–equation, parametric, linear model

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_k y_{t-k} + \beta_0 x_t + \cdots + \beta_p x_{t-p} + e_t \tag{2.9}$$

in which $\beta_0$ is suspected to change after some date.

$$\begin{aligned} H_0 : & \beta_0 = \overline{\beta} \quad \text{for all } t \\ H_1 : & \begin{cases} \beta_0 = \beta_{0,1} & \text{for } t \leq t_0 \\ \beta_0 = \beta_{0,2} & \text{for } t > t_0 \end{cases} \end{aligned}$$

The test for a break at a known date is standard.

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_k y_{t-k} + \beta_0 x_t + 1(t \leq t_0)\delta x_t + \cdots + \beta_p x_{t-p} + e_t \tag{2.10}$$

where 1() is the indicator function. The null hypothesis corresponds to $\delta = 0$. The Wald test statistic may be formed to test the null hypothesis.

The procedure for an unknown break date is to run the test for a known break at each candidate break date, then to take the supremum of the Wald test statistics. As such this will be a sup–Wald statistic. The distribution of the resulting test statistic is nonstandard. The test statistic follows a Brownian motion (see Andrews (1993), theorem 3) that depends on the number of parameters that are subject to a break and the fraction of the sample which is subject to the test.

### 2.3.2 Tests for parameter instability in multiple equations

In a multiple equation context, we stack the single-equation regressions. Consider the collection of $h$ local projections for a single response variable $y_t$ and a single exogenous variable $x_t$. These regressions are

$$\begin{aligned} y_t &= \alpha_1^h y_{t-1} + \cdots + \alpha_k^h y_{t-k} + \beta_0^o x_t + \cdots + \beta_p x_{t-p} + u_t \\ &\vdots \\ y_{t+h} &= \alpha_1^h y_{t-1} + \cdots + \alpha_k^h y_{t-k} + \beta_0^h x_t + \cdots + \beta_p^h x_{t-p} + u_{t+h} \end{aligned}$$

The method for multiple response variables and multiple exogenous variables is analogous: it simply increases the number of stacked equations.

To test for a structural break in the collection of dynamic multiplier coefficients, proceed as follows. At a candidate date $t_0$, test for a structural break by introducing coefficients on the interaction terms of $\{x_t, \ldots, x_{t-p}\}$ with the indicator for time $t > t_0$. The null hypothesis is of no break in any of the coefficients at any date; the alternative is aa break in any of the coefficients at any date. Since these parameters can be directly interpreted as the dynamic multiplier coefficients, this procedure provides a direct test for parameter instability in dynamic multipliers.

Perform this test at each candidate date and select the maximum of the Wald test statistics. The sup-Wald test described in Andrews (1993) provides valid inference for the existence of a break within the candidate date range.

In principle this procedure can also be used to test for parameter instability in the impulse response coefficients, but such an approach encounters two difficulties. First, the standard errors of the impulse response coefficients needs to be corrected, and corrected for a different degree of moving–average terms at each horizon. Second, these raw impulse response coefficients are not the orthogonalized impulse responses that are of most direct interest. There is little agreement so far as to how to compute orthogonalized local projections. As such, conducing valid inference for cross-equation tests of orthogonalized local projection impulse responses is an important extension, but one that is not considered here.

### 2.3.3 Monte Carlo evidence

This second set of experiments reports Monte Carlo results regarding the finite sample performance of the tests discussed above. It considers tests of parameter instability in a model with a single outcome variable $y_t$ and a single shock variable $x_t$. Details of the data–generating process are in the Appendix. I consider a test for a break in the collection of the first 8 dynamic multipliers, corresponding to responses out to two years ahead in quarterly data.

The object of interest is the stability of the dynamic multiplier function. The dynamic multiplier function traces out the effect of a shock on $x_t$ on $y_{t+s}$ for horizons $s = 1, 2, \ldots, H$. The collection of responses is estimated simultaneously and jointly by local projections. At any given date $t_0$, the collection of dynamic multipliers can be estimated for each subperiod $t \leq t_0$ and $t > t_0$ and a test can be run to assess the stability of the dynamic multiplier function across subperiods. The sup–Wald test for parameter stability performs this test for each candidate break date. The null hypothesis is that there is no break in the dynamic multiplier coefficients.

The data–generating process consists of a scalar exogenous variable $x_t$ and an endogenous response variable $y_t$. The exogenous variable is independently and identically distributed and drawn from the normal distribution. The endogenous variable follows

$$\begin{aligned} y_t = {} & 1.3y_{t-1} - 0.16y_{t-2} - 0.30y_{t-3} + 0.15y_{t-4} \\ & - 0.16x_t - 0.2x_{t-1} + 0.37x_{t-2} - 019x_{t-3} - 0.22x_{t-4} + u_t \end{aligned} \tag{2.11}$$

where $u_t$ is an i.i.d. shock. These coefficients were chosen to match those of a regression of log real GDP on its first four lags, and on the contemporaneous value and four lags of the Romer and Romer (2010) tax shocks. As such, the data–generating process is close to the processes that are tested in Section 4.

The test is performed under the null hypothesis of no structural break. For each potential break date $t_0$, I simultaneously estimate the dynamic multipliers before and after the break date,

$$\begin{aligned} y_t &= \mathbf{a}(L)\mathbf{y}_{t-1} + b_0^0\mathbf{x}_t + (t < t_0)\delta_0^0\mathbf{x}_t + \mathbf{b}(L)\mathbf{x}_{t-1} + (t < t_0)\boldsymbol{\delta}_L\mathbf{x}_{t-1} + v_t \\ &\vdots \\ y_{t+H} &= \mathbf{a}(L)\mathbf{y}_{t-1} + b_0^H\mathbf{x}_t + (t < t_0)\delta_0^H\mathbf{x}_t + \mathbf{b}(L)\mathbf{x}_{t-1} + (t < t_0)\boldsymbol{\delta}_L\mathbf{x}_{t-1} + v_{t+H} \end{aligned} \tag{2.12}$$

Compare (2.8) to (2.12). The difference is that (2.12) contains interaction terms for each coefficient on $\{x_t, \ldots, x_{t-p}$. Since the collection of coefficients $\{b_0^0, \ldots, b_0^h\}$ are the dynamic multipliers, the collection of coefficients $\{\delta_0^0, \ldots, \delta_0^h\}$ are the collection of coefficients indicating the presence of a structural break. If these coefficients are nonzero, then the dynamic multiplier coefficients differ across the two subperiods. The collection of coefficients $\boldsymbol{\delta}_L$ are nuisance parameters.

Since the test involves parameters across each outcome variable and horizon, the regressions must be estimated jointly. The collection of estimates $\{\delta_0^0, \delta_0^1, \ldots, \delta_0^H\}$ are the parameters of interest. Under the null all of the $\delta_0^h$ are equal to zero. For each candidate break date, the Wald test statistic is recorded. The test statistic for the test of parameter instability is the maximum of the period–by–period Wald tests. This test statistic is compared to the critical values in Andrews (1993).

To build intuition, Figure 2.2 displays five representative runs of the test for structural break under the null. For each period, a test for structural break is performed and the Wald test statistic is recorded. The resulting Wald test statistics are shown in the figure. When there is no break, the test statistic is approximately uniformly distributed over the sample period. The maximum of the Wald tests, which is the statistic of interest, is biased somewhat towards the endpoints because, near the endpoints, the sample sizes of the two subperiods will be unbalanced, the regression coefficients will be more variable in the subperiod with the smaller sample size, and the test statistic will tend to report a break. These biases are known and pose no difficulty for inference.

The sup-Wald test is an asymptotic test. For this reason, the following set of Monte Carlo results compare the nominal size of the test to its actual size for three sample sizes. One thousand replications were run at sample sizes $T = 240$, $T = 500$, and $T = 800$, in each case with the interior 70% of the sample period being considered a candidate for a break. Table

3.3 reports rejection rates at nominal sizes of 10%, 5%, and 1%. When the sample size is 240, the test over-rejects slightly, but the 5% and 10% rejection rates continue to fall within their own 95% confidence interval. At a sample size of 500, the tests do not substantially over– or under–reject. The sample size of 800 slightly over–rejects at the 1% level, but the correct nominal size remains within the 95% confidence interval.

## 2.4 Applications

I demonstrate the utility of this test by an investigation of parameter instability in three measured economic shock series. Over the past three decades, advances have been made in measuring shocks directly. These shocks to monetary policy, fiscal policy, technology, and energy prices can then be used directly as shocks in a VAR framework. In particular, dynamic multipliers to these shocks can be computed in a single-equation system involving only the desired response variable and the shock variable. In this section, I use the test for parameter instability for dynamic multipliers to revisit the monetary policy shocks described in Romer and Romer (2004), the tax shocks in Romer and Romer (2010), and the energy price shocks in Hamilton (1996).

Some prelimiary comments on the interpretation of breaks in this context are in order. The impulse response coefficients are amalgams of underlying deep parameters. Tests for a break in the impulse response coefficients can indicate the presence of structural change in the economy, but are not necessarily capable of distinguishing among competing theories as to the nature of the change. To do that requires imposing further structure on the problem, to find specify more fully the mapping between deep parameters and impulse response coefficients.

### 2.4.1 Monetary shocks

The Romer monetary shocks are calculated by regressing the Federal funds rate on the Greenbook forecast, and taking the residual of the regression to be the shock. The resulting shocks, as extended in Wieland and Yang (2016), are provided in figure 2.3. These are measured in units of the Federal funds rate, so a value of one indicates a shock of one hundred basis points.

First, I run a collection of local projection dynamic multipliers on the full sample. The regression specification is

$$y_{t+h} = \alpha_1^h y_{t-1} + \cdots + \alpha_p^h y_{t-p} + \beta_0^h r_t + \cdots + \beta_p^h r_{t-p} + u_t$$

where $y_t$ is an economic activity variable; as in Romer and Romer (2004), I use the logarithm of industrial production. Twenty–four lags are included in both variables. I consider $H = 48$ responses, corresponding to four years of monthly data. The collection of coefficients $\{\beta_0^h\}$ holds the full–sample estimate of the response are provided in figure 2.4. In the full sample, a 100 basis point Federal fund shock leads industrial production to fall by slightly more than two percent over the following two years. Over the subsequent two years, industrial production recovers, and by four years it returns to its pre–shock value. Note that industrial production rises slightly in the months immediately following a shock. This behavior is similarly documented in Ramey (2016) and is present across many specifications of the monetary shock.

I next test for parameter stability of the dynamic multiplier coefficients. Coibion (2012) and Ramey (2016) have noted that the output response to Romer–Romer monetary shocks is sensitive to the inclusion of the non–borrowed reserve targeting period of 1979–81, but do not test for parameter instability formally. I test for parameter instability in the first three years of dynamic multiplier coefficients, $h = 0$ to $h = 36$. This is enough to capture the trough at a horizon of two years. I allow all dates from 1974 to 1998, inclusive, to be candidates for a break; this corresponds to 70% of the available sample. I include twenty–four lags in each local projection, as in the full-sample specification.

The resulting test statistics for a null of no break in each period appear in figure 2.5. The test statistics are uniformly high in the late 1970s and early 1980s, and fall rapidly thereafter. The test statistic peaks in March 1980 and is statistically significant ($p < 0.00001$). I re–run the local projection estimator for the period 1969–80 and 1980–2007 based on these results. Note that the latter period contains most of the nonborrowed–reserve targeting period.

Figures 2.6 and 2.7 display dynamic multipliers up to a horizon of four years after the shock for the pre–1980 and post–1980 periods. In each figure, the shock is a 100 basis point Federal funds surprise, as above. In the early period, the monetary shock leads to a statistically significant decline in industrial production, with the effect reaching a trough of more than -5% after slightly over two years. The shock's effect dissipates only slowly, and after four years industrial production remains four percent below trend, though by that point the standard errors are wide and the effect is statistically indistinguishable from zero.

By contrast, the industrial production response to monetary shocks is muted in the post-1980 period. The same 100 basis point shock causes a decline of no more than 1% of industrial production after two years, and the response of industrial production is statistically indistinguishable from zero at nearly all time horizons. The rise in industrial production after a shock, first seen in the full–sample results, appears also in the second subperiod. This rise in economic activity after a contractionary shock is not present in the early subperiod.

To the extent that the Romer and Romer shocks indicate a strong response of real activity to monetary shocks, the effect is driven entirely by observations before 1980. As these estimated dynamic multipliers assume the same sized shock, the result is not attributable to smaller monetary shocks; it is attributable to a change in the response coefficients themselves. The full-sample responses are an average of the large negative effects seen in the 1969–80 sample and the small, statistically insignificant effects seen in the 1980-2007 sample.

### 2.4.2 Tax shocks

Romer and Romer (2010) study the effects of tax shocks on real GDP. I next investigate the possibility of structural change with respect to the response of output to tax shocks.

I begin by presenting the estimated response of output to a 1% of GDP tax shock using full-sample information. The local projection specification included twelve lags each of output, the contemporaneous value of the tax shock, and twelve lags of the tax shock, as in Romer and Romer (2010). The response variable is the logarithm of real GDP; the shock is the Romer and Romer series of exogenous tax changes. Local projection estimates of the effect of tax shocks on output for twenty–four periods (6 years) and using the full sample are presented in figure 2.8. A tax shock of one percentage point of GDP reduces output, with the effect attaining a trough of about -2% of GDP ten quarters after the shock. These results are qualitatively similar to the results in Romer and Romer (2010), whose VAR–based estimate indicates a trough of a 3% decline in real GDP ten quarters after the shock.

I test for structural change in the response coefficients. The test is flexible in that the researcher may choose the length of the horizon over which to test, and one could even test at specific portions of the response horizon. I run the test three times, looking for a change in the dynamic multiplier coefficients in the first 16, 20, and 24 periods after a shock, representing a test for stability in the first four, five, and six years worth of responses after a shock. I search for a break in each quarter from 1964 to 1996, representing 50% of the sample period. When testing for a break in the first sixteen response coefficients, there is no evidence for structural change. The Wald test statistic peaks at a value of 30 ($p = 0.23$) in the second quarter of 1981. When five years of response coefficients are tested, the Wald test statistic peaks in the second quarter of 1980, but remains statistically insignificant ($p = 0.54$). When six years of response coefficients are tested, the Wald test statistic peaks in the first quarter of 1964, and again is insignificant ($p = 0.5$). Inspection of the Wald test statistic throughout the potential break date range indicates that it has two peaks, one in the beginning of the sample and one in the 1980s.

Since the test for parameter instability did not indicate sufficient statistical evidence for a break, the dynamic response of output to a tax shock should be similar across the pre–1980 and post–1980 subperiods. Figures 2.9 and 2.10 plot the response of output to tax shocks in these two subsamples. Indeed, the response of output to a tax shock looks similar across the two subsamples, and both look similar to the full sample results. In the 1947-1981 subsample, output falls in response to a tax shock, with the effect reaching a trough approximately two years after the initial shock. In the 1981-2007 sample, the qualitative response is the same; the effect is somewhat attenuated, the standard errors are somewhat wider, and the responses are statistically indistinguishable from zero but are also statistically indistinguishable from the early–subsample responses. When estimated on the full sample, the point estimates are little changed and confidence intervals are narrower.

### 2.4.3 Oil shocks

The final application involves the resposne of economic activity to energy price shocks. Several options are available to measure oil price shocks. I will use the shocks described in Hamilton (1996). In each month, the oil shock is defined as the amount by which oil prices in month $t$ exceed their peak value over the previous 12 months; if they do not exceed the previous peak, then the oil shock is taken to be zero.

The oil price variable is the monthly price of Brent crude oil, and the shock to oil prices is the Hamilton (1996) net oil price increase. The response variable is industrial production.

I use one year of lags and test for parameter instability in the first two years of impulse responses. Figure 2.11 displays the response of output to a 1% net oil price increase. Output falls slightly in the first year after a shock, then falls considerably more during the second year and remains low for four years after the shock. A 1% net oil price increase is associated with output 0.2% below trend four years after the shock.

I next test for structural change in the response coefficients. I test for a break in the first 24 response coefficients, representing the response of industrial production to an oil shock up to two years after the onset of the shock. The candidate break dates consisted of every month from 1955 to 2000. The Wald test statistic attains a peak in 1971, but is statistically insignificant ($p = 0.8$). Nevertheless, I again split the sample at the most likely date to demonstrate the behavior of the responses in each period.

Figures 2.12 and 2.13 display the response of industrial production to an oil shock in the early and late subsamples, respectively. In both subsamples, industrial production falls in response to an oil price increase. The response is weaker in the early period, with output's response at many horizons being statistically indistinguishable from zero. The response coefficients in the 1971-2013 period closely mirror those in the full sample. In the 1971-2013 subsample, output falls slightly on impact, with a steeper decline one year after the onset of the shock. The response of output troughs after slightly more than one year, and output remains depressed four years after the onset of the shock.

## 2.5 Conclusion

This paper has extended the notion of local projection impulse response functions to include exogenous variables, proposed a test for parameter instability in impulse–response coefficients using the local projection estimator, and explored the parameter stability of the response of real output to identified monetary shocks.

Local projections provide an alternative to vector autoregressions in building impulse response functions. These projections can be obtained equation–by–equation, horizon–by–horizon, and can be calculated through simple linear regression. The standard errors produced by simple linear regression are inappropriate because the error term in the local projection estimator contains a moving–average term which necessitates a correction of the least–squares standard errors. However, in a local projection with strictly exogenous variables, resulting standard errors on the dynamic multiplier coefficients are sufficient for valid inference.

Tests for parameter stability in time–series linear regression have been previously described. I extend this to the case of multiple equations and, specifically, to local projections with exogenous variables. The LR–based test does not provide valid inference because it depends on the likelihood ratio, which is marred by the moving–average term in the residual of the local projection regressions. However, the test based on the Wald statistic continues to be valid for local projection dynamic multipliers.

I use the sup–Wald test to explore parameter instability in three leading examples of identified shocks: the Romer and Romer monetary shocks, the Romer and Romer tax shocks, and the Hamilton net oil price increases. I find evidence for a structural break in the dynamic response of real economic activity to monetary shocks. The candidate break dates cluster in the late 1970s to early 1980s, with the point estimate being early 1980. When running separate local projections for the two regimes, I find that the evidence for strong real effects of monetary policy shocks are driven primarily by the experience of the 1970s. There is little evidence of significant real effects of monetary policy in the post-1980 period. Meanwhile, there is little evidence for instability in the response of real activity to either tax shocks or oil price shocks. The sup–Wald test cannot reject the null hypothesis of parameter stability for either shock in any candidate break date.

## 2.6 Tables and Figures

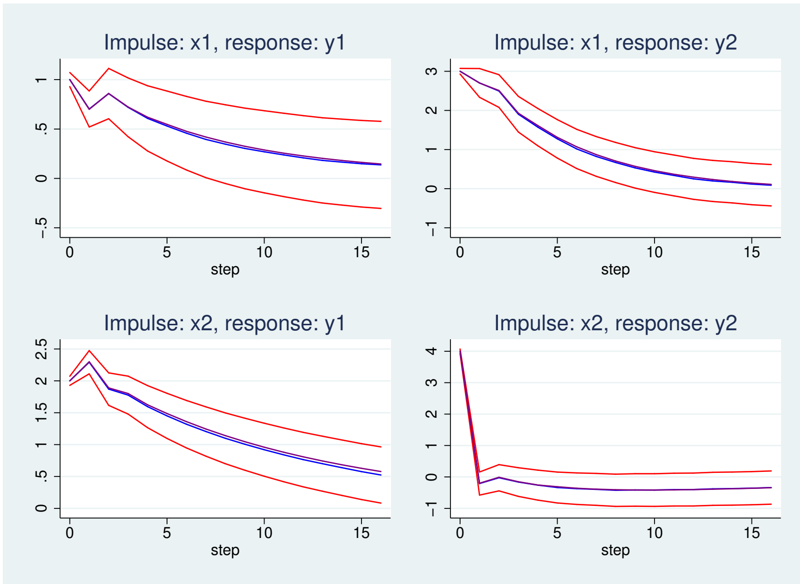


Figure 2.1: Local projection estimates of dynamic multipliers

Note: True values in purple; mean of local projection estimates in blue; 2 standard error bands in red.

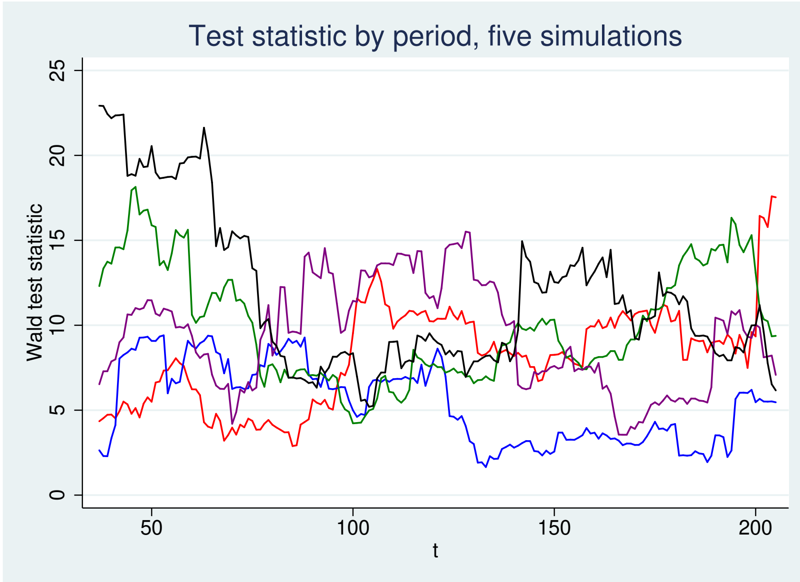


Figure 2.2: Wald test statistic for each period in a break test

Note: Five simulations are shown. Each is drawn from a data–generating process without a break. 240 periods are simulated; the test is performed in 170 (70%) of those periods. Horizontal axis denotes the simulation period.

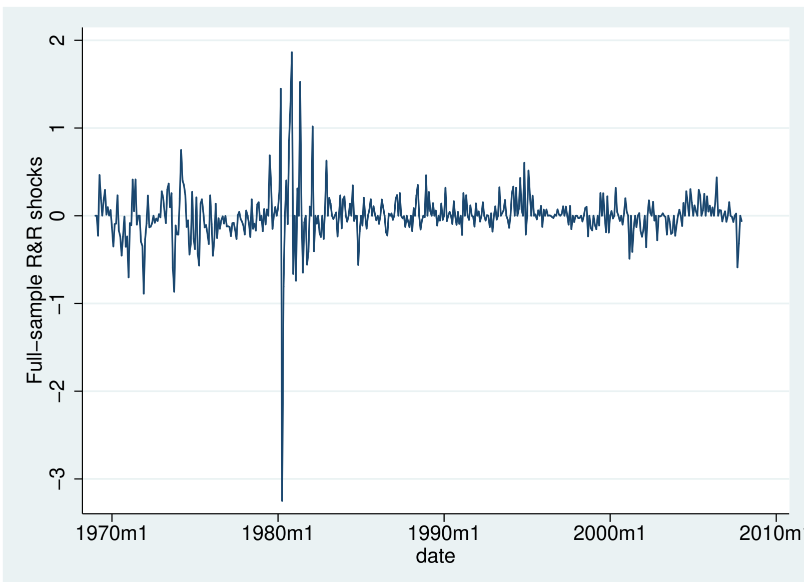


Figure 2.3: Romer and Romer monetary shocks, 1969–2007

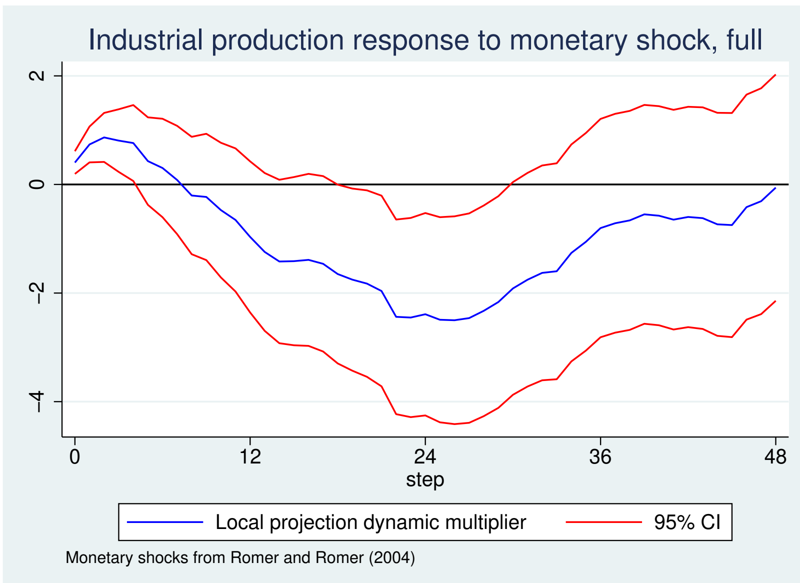


Figure 2.4: Response of industrial production to Romer–Romer monetary shock, 1969-2004
Note: Full sample. Estimated by local projections with 24 lags. Each step is one month.

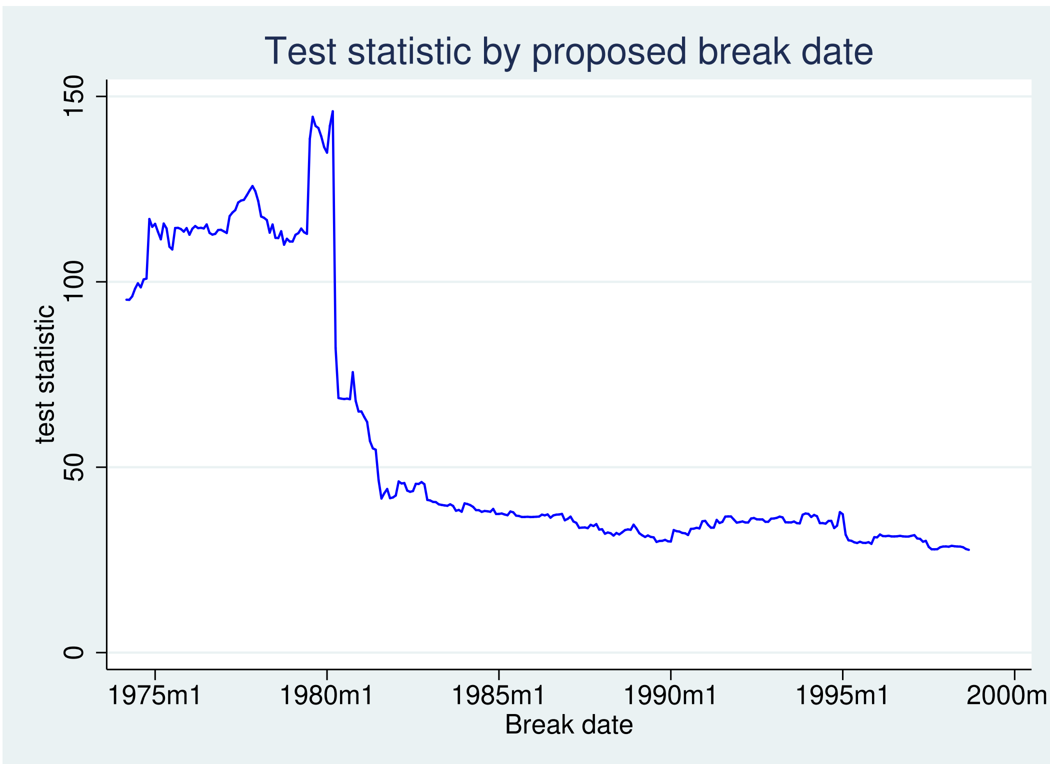


Figure 2.5: Test statistic associated with each break date; response of IP to monetary shock

Note: Romer and Romer (2004) industrial production–monetary shock model with 24 lags and considering the first 36 responses, 1974–1998

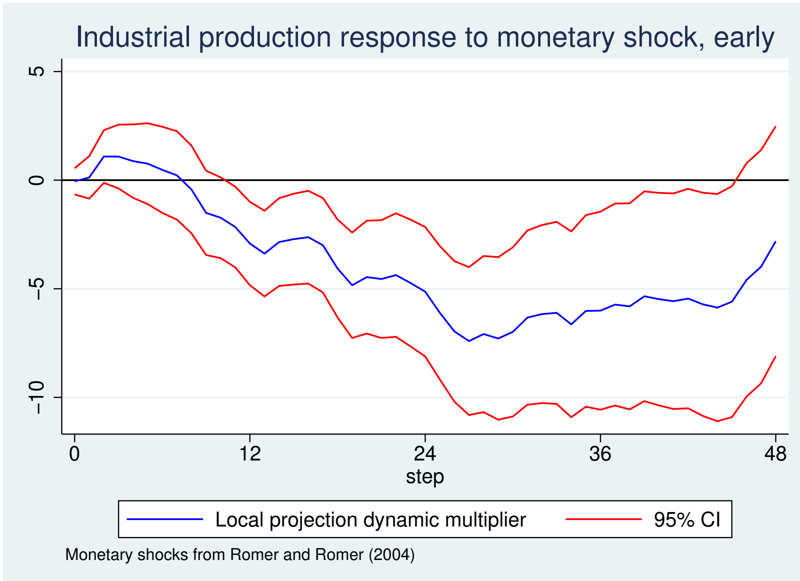


Figure 2.6: Response of industrial production to Romer–Romer monetary shock, 1969-1980
Note: Early sample as indicated by parameter instability test. Estimated by local projections with 24 lags. Each step is one month.

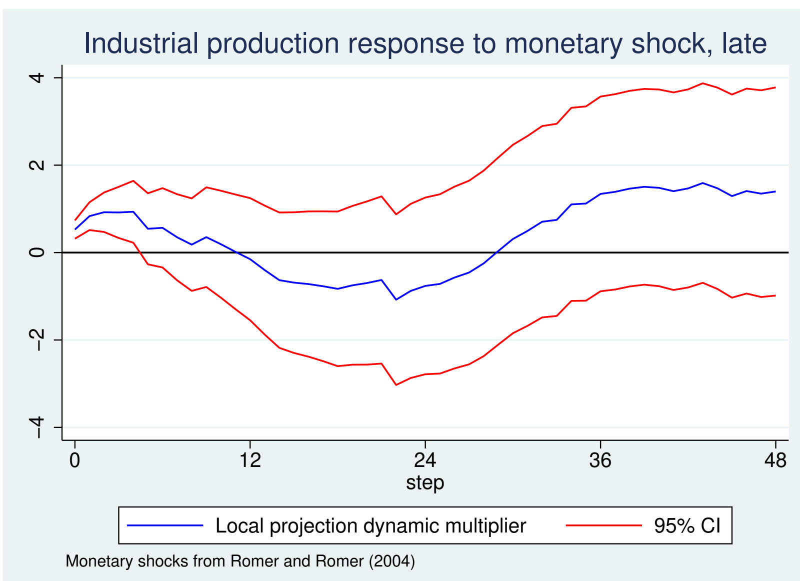


Figure 2.7: Response of industrial production to Romer–Romer monetary shock, 1980-2004
Note: Late sample as indicated by parameter instability test. Estimated by local projections with 24 lags. Each step is one month.

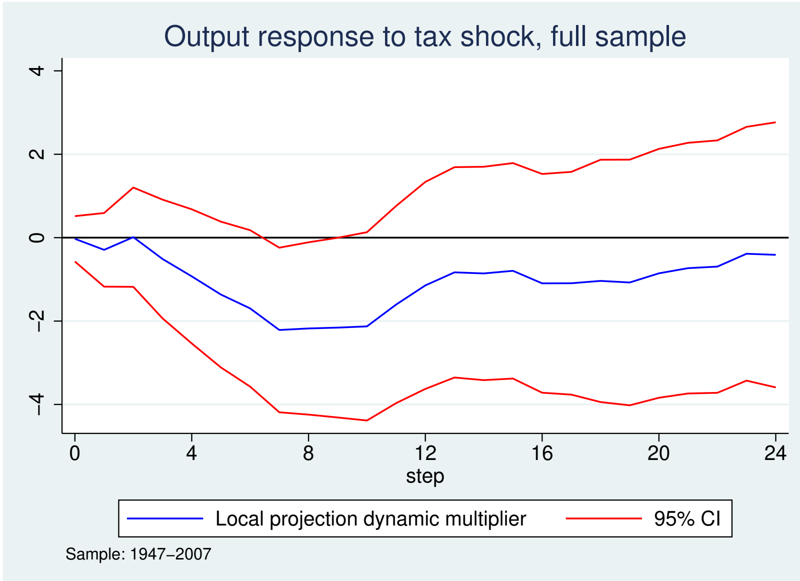


Figure 2.8: Response of output to Romer–Romer tax shock, 1947-2007
Note: Estimated by local projections with 12 lags. Each step is one quarter.

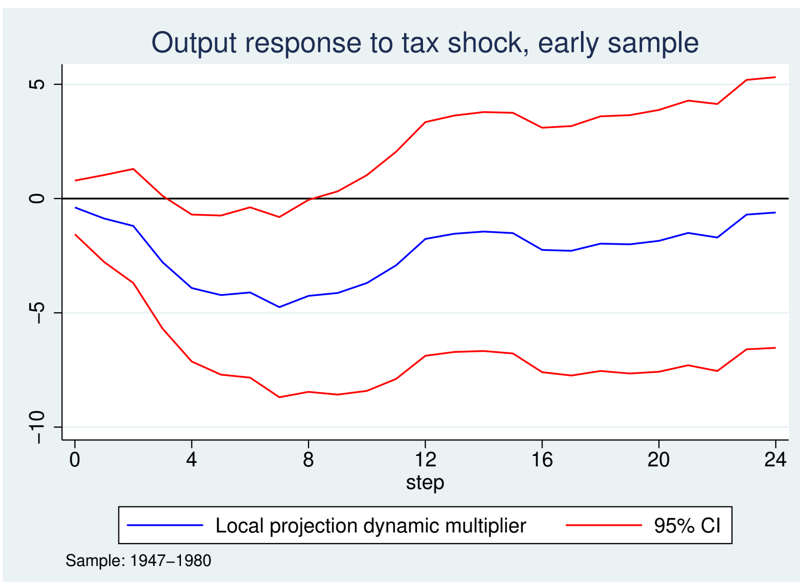


Figure 2.9: Response of output to Romer–Romer tax shock, 1947-1980
Note: Estimated by local projections with 12 lags. Each step is one quarter.

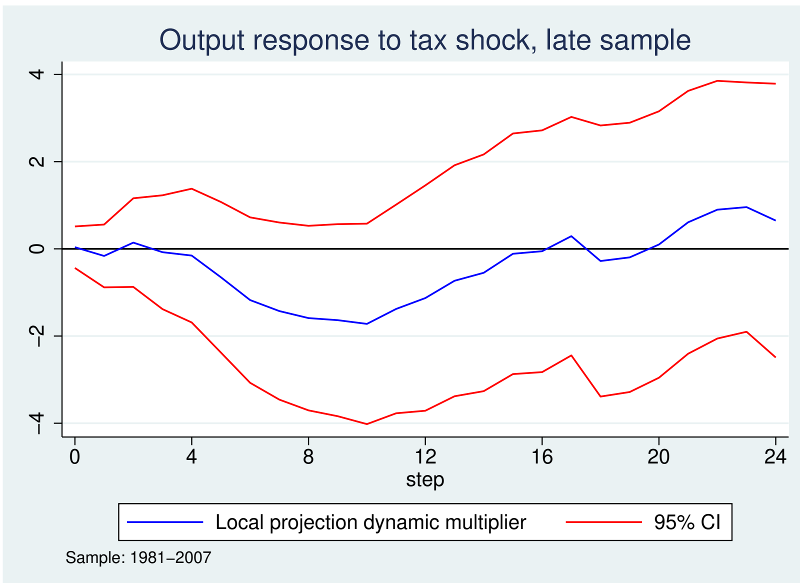


Figure 2.10: Response of output to Romer–Romer tax shock, 1981-2007

Note: Estimated by local projections with 12 lags. Each step is one quarter.

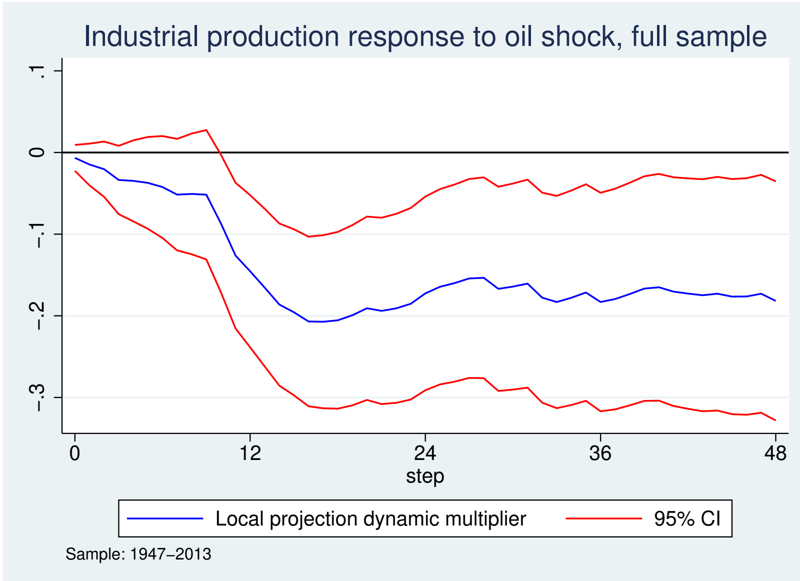


Figure 2.11: Response of industrial production to Hamilton oil shock, 1947-2013
Note: Estimated by local projections with 12 lags. Each step is one month.

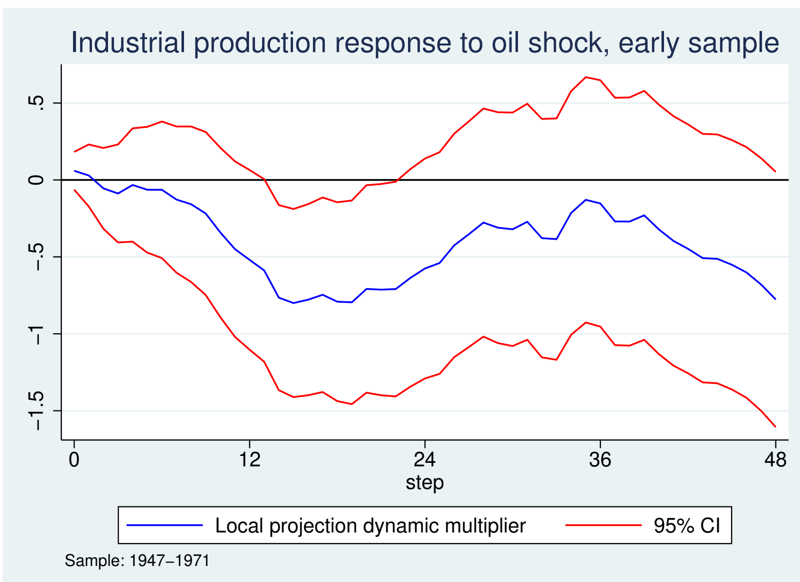


Figure 2.12: Response of industrial production to Hamilton oil shock, 1947-1971
Note: Estimated by local projections with 12 lags. Each step is one month.

Table 2.1: Dynamic multipliers to $x_1$ shock by local projections

| Horizon | True coeff. | Mean | Std. Dev. | Rejection rate |
|---|---|---|---|---|
| Impulse: $x_1$; Response: $y_1$ | | | | |
| 0 | 1.0000 | 1.0009 | 0.0357 | 0.0290 |
| 1 | 0.7000 | 0.7028 | 0.0911 | 0.0390 |
| 2 | 0.8600 | 0.8588 | 0.1274 | 0.0410 |
| 3 | 0.7230 | 0.7196 | 0.1488 | 0.0550 |
| 4 | 0.6199 | 0.6079 | 0.1651 | 0.0430 |
| 5 | 0.5471 | 0.5305 | 0.1770 | 0.0540 |
| 6 | 0.4768 | 0.4578 | 0.1863 | 0.0560 |
| 7 | 0.4190 | 0.3947 | 0.1935 | 0.0520 |
| 8 | 0.3687 | 0.3470 | 0.1993 | 0.0500 |
| 9 | 0.3254 | 0.3037 | 0.2040 | 0.0600 |
| 10 | 0.2880 | 0.2697 | 0.2079 | 0.0510 |
| 11 | 0.2556 | 0.2379 | 0.2110 | 0.0530 |
| 12 | 0.2274 | 0.2084 | 0.2136 | 0.0610 |
| 13 | 0.2028 | 0.1823 | 0.2158 | 0.0440 |
| 14 | 0.1814 | 0.1649 | 0.2176 | 0.0480 |
| 15 | 0.1625 | 0.1486 | 0.2191 | 0.0590 |
| 16 | 0.1460 | 0.1371 | 0.2203 | 0.0590 |
| Impulse: $x_1$; Response: $y_2$ | | | | |
| 0 | 3.0000 | 3.0023 | 0.0357 | 0.0570 |
| 1 | 2.7000 | 2.7033 | 0.1837 | 0.0430 |
| 2 | 2.5100 | 2.4956 | 0.2088 | 0.0410 |
| 3 | 1.9280 | 1.9032 | 0.2278 | 0.0440 |
| 4 | 1.6139 | 1.5726 | 0.2385 | 0.0560 |
| 5 | 1.3126 | 1.2737 | 0.2457 | 0.0630 |
| 6 | 1.0702 | 1.0158 | 0.2506 | 0.0550 |
| 7 | 0.8710 | 0.8242 | 0.2539 | 0.0610 |
| 8 | 0.7063 | 0.6692 | 0.2564 | 0.0560 |
| 9 | 0.5712 | 0.5320 | 0.2583 | 0.0510 |
| 10 | 0.4601 | 0.4242 | 0.2598 | 0.0530 |
| 11 | 0.3689 | 0.3382 | 0.2609 | 0.0460 |
| 12 | 0.2943 | 0.2512 | 0.2618 | 0.0410 |
| 13 | 0.2333 | 0.1970 | 0.2626 | 0.0440 |
| 14 | 0.1835 | 0.1629 | 0.2632 | 0.0390 |
| 15 | 0.1429 | 0.1176 | 0.2638 | 0.0550 |
| 16 | 0.1099 | 0.0881 | 0.2644 | 0.0450 |

True impulse response of $y_1$ and $y_2$ to a $x_2$ shock. True coefficient, mean calculated local projection response, standard deviation, and rejection rate are reported. 1,000 simulations were conducted at a sample size of 800. Sixteen steps after the initial shock are shown.

Table 2.2: Dynamic multipliers to $x_2$ shock by local projections

| Horizon | True coeff. | Mean | Std. Dev. | Rejection rate |
|---|---|---|---|---|
| Impulse: $x_2$; response: $y_1$ | | | | |
| 0 | 2.0000 | 2.0010 | 0.0357 | 0.0510 |
| 1 | 2.3000 | 2.2927 | 0.0911 | 0.0460 |
| 2 | 1.8900 | 1.8710 | 0.1273 | 0.0510 |
| 3 | 1.8020 | 1.7768 | 0.1487 | 0.0530 |
| 4 | 1.6251 | 1.5962 | 0.1650 | 0.0580 |
| 5 | 1.4876 | 1.4512 | 0.1769 | 0.0500 |
| 6 | 1.3608 | 1.3203 | 0.1862 | 0.0530 |
| 7 | 1.2453 | 1.2058 | 0.1935 | 0.0570 |
| 8 | 1.1410 | 1.0990 | 0.1993 | 0.0540 |
| 9 | 1.0460 | 1.0067 | 0.2040 | 0.0500 |
| 10 | 0.9595 | 0.9200 | 0.2078 | 0.0490 |
| 11 | 0.8808 | 0.8400 | 0.2110 | 0.0510 |
| 12 | 0.8089 | 0.7653 | 0.2136 | 0.0630 |
| 13 | 0.7433 | 0.7002 | 0.2157 | 0.0690 |
| 14 | 0.6833 | 0.6377 | 0.2175 | 0.0730 |
| 15 | 0.6284 | 0.5768 | 0.2190 | 0.0800 |
| 16 | 0.5781 | 0.5236 | 0.2203 | 0.0770 |
| Impulse: $x_2$; response: $y_2$ | | | | |
| 0 | 4.0000 | 4.0005 | 0.0357 | 0.0470 |
| 1 | -0.2000 | -0.2103 | 0.1837 | 0.0470 |
| 2 | -0.0100 | -0.0256 | 0.2088 | 0.0530 |
| 3 | -0.1530 | -0.1627 | 0.2279 | 0.0390 |
| 4 | -0.2564 | -0.2604 | 0.2385 | 0.0530 |
| 5 | -0.3116 | -0.3367 | 0.2458 | 0.0450 |
| 6 | -0.3595 | -0.3732 | 0.2507 | 0.0480 |
| 7 | -0.3887 | -0.3956 | 0.2541 | 0.0630 |
| 8 | -0.4066 | -0.4245 | 0.2565 | 0.0480 |
| 9 | -0.4149 | -0.4141 | 0.2584 | 0.0560 |
| 10 | -0.4159 | -0.4164 | 0.2599 | 0.0480 |
| 11 | -0.4112 | -0.4028 | 0.2610 | 0.0490 |
| 12 | -0.4022 | -0.4001 | 0.2619 | 0.0560 |
| 13 | -0.3901 | -0.3778 | 0.2627 | 0.0530 |
| 14 | -0.3757 | -0.3701 | 0.2633 | 0.0490 |
| 15 | -0.3598 | -0.3573 | 0.2640 | 0.0410 |
| 16 | -0.3429 | -0.3379 | 0.2645 | 0.0570 |

True impulse response of $y_1$ and $y_2$to a $x_2$ shock. True coefficient, mean calculated local projection response, standard deviation, and rejection rate are reported. 1,000 simulations were conducted at a sample size of 800. Sixteen steps after the initial shock are shown.

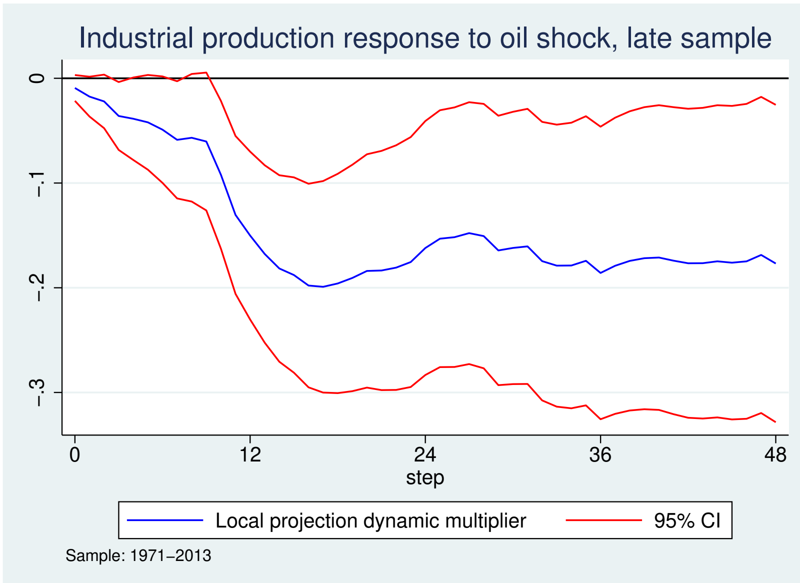


Figure 2.13: Response of industrial production to Hamilton oil shock, 1971-2013
Note: Estimated by local projections with 12 lags. Each step is one month.

Table 2.3: Simulated finite sample significance levels

| Sample Size | Nominal size | Simulated rejection rate | Binomial exact 95% CI | |
|---|---|---|---|---|
| $T = 240$ | 1% | 0.022 | 0.0138 | 0.0331 |
| | 5% | 0.064 | 0.049 | 0.081 |
| | 10% | 0.112 | 0.093 | 0.133 |
| $T = 500$ | 1% | 0.007 | 0.0028 | 0.0143 |
| | 5% | 0.043 | 0.031 | 0.057 |
| | 10% | 0.109 | 0.090 | 0.130 |
| $T = 800$ | 1% | 0.017 | 0.0099 | 0.0279 |
| | 5% | 0.058 | 0.044 | 0.074 |
| | 10% | 0.096 | 0.078 | 0.116 |

Size from simulated model under the null of no parameter instability.

# Chapter 3

# Standard Eerrors for Impulse Response Functions of Estimated DSGE Models

## 3.1 Introduction

A key object of interest in DSGE model analysis is the impulse response function, which traces out the effect of a shock on variables of interest in the model. In an estimated DSGE model, the impulse response functions are also estimated quantities, so point estimates of the impulse response function have associated standard errors. This paper derives asymptotically valid standard errors for impulse responses calculated from an estimated linear(ized) DSGE model. I then provide simulation evidence on the small-sample properties of these standard errors.

The results shown here are related to the literature on the small-sample properties of estimated DSGE models. Cho and Moreno (2006) provides a small–sample study of the parameters of a three–equation New Keynesian model with three shocks. Cho and Moreno (2006) consider the distribution of structural model parameters in settings with fairly small sample sizes. This paper provides complementary analysis of the distribution of impulse response coefficients in similarly data-constrained environments. Morris (2017) provides a study of the behavior of DSGE model in simulation, focusing on "pileups" in which estimated parameter values tend to cluster near the boundary of the acceptable parameter space.

A linear DSGE model consists of equations $\mathbf{A}(\theta)E_t\mathbf{z}_{t+1} = \mathbf{B}(\boldsymbol{\theta})\mathbf{z}_t$ where $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$ are matrices whose entries depend on an underlying structural parameter vector $\boldsymbol{\theta}$ and where $\mathbf{z}_t$ collects all the variables in the model. Using the methods of Klein (2000) or Schmitt-Grohé and Uribe (2004), the model can be solved to obtain the state–space form

$$\mathbf{y}_t = \mathbf{G}(\boldsymbol{\theta})\mathbf{x}_t \tag{3.1}$$

$$\mathbf{x}_{t+1} = \mathbf{H}(\boldsymbol{\theta})\mathbf{x}_t + \boldsymbol{\eta}(\boldsymbol{\theta})\mathbf{e}_{t+1} \tag{3.2}$$

where $\mathbf{x}_t$ contains the model's predetermined variables and $\mathbf{y}_t$ contrains the model's control variables. The vector of structual shocks $\mathbf{e}_{t+1}$ has mean zero and variance matrix equal to the identity matrix. The $n_y \times n_x$ matrix $\mathbf{G}(\boldsymbol{\theta})$ reports the impact effect of change in the state variables $x_t$ on the control variables $\mathbf{y}_t$. The $n_x \times n_x$ matrix $\mathbf{H}(\boldsymbol{\theta})$ reports the vector autoregressive process governing the evolution of the state variables. The $n_x \times n_e$ matrix

$\boldsymbol{\eta}(\boldsymbol{\theta})$ holds the standard deviations of the shocks along its diagonal.

An impulse response vector is the value the variables take when the model is subjected to a shock $\mathbf{e}_0$. Denote these responses by the vector $\mathbf{r}$. The impulse response at horizon $t$, $\mathbf{r}_t$, is

$$\begin{aligned}
\mathbf{r}_t^y &= \mathbf{G}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})^t\boldsymbol{\eta}(\boldsymbol{\theta})\mathbf{e}_0 \\
\mathbf{r}_t^x &= \mathbf{H}(\boldsymbol{\theta})^t\boldsymbol{\eta}(\boldsymbol{\theta})\mathbf{e}_0
\end{aligned}$$

for the control variables and state variables, respectively, and for $t = 0, 1, \ldots, T$.

## 3.2 Parameter estimation, IRF estimation, and IRF standard errors

Let a subset of the control variables be observed and used for estimation,

$$\mathbf{d}_t = \mathbf{D}\mathbf{y}_t \tag{3.3}$$

where $\mathbf{D}$ is a selection matrix which picks out rows of $\mathbf{G}(\boldsymbol{\theta})$ corresponding to variables observed by the econometrician. Then (3.1)–(3.3) forms a state space model and the parameters $\boldsymbol{\theta}$ can be estimated by maximum likelihood. An estimate of the variance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$ may be computed as the inverse of the negative Hessian at the maximized $\hat{\boldsymbol{\theta}}$.

The impulse response function at a given time horizon is a column vector, each entry of which is a function of $\boldsymbol{\theta}$. Point estimates are found by plugging in, $\hat{\mathbf{r}}_t = \mathbf{r}(\hat{\boldsymbol{\theta}})$. Asymptotic standard errors are obtained with the delta method. At a given horizon,

$$\sqrt{T}\,(\hat{\mathbf{r}}_t^y - \mathbf{r}) \to^d N(\mathbf{0}, \mathbf{R}_t^y \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \mathbf{R}_t^{y\prime}) \tag{3.4}$$

where

$$\mathbf{R}_t^y = \frac{\partial \mathbf{r}_t^y}{\partial \boldsymbol{\theta}}$$

Similarly, for the impulse response of state variables,

$$\sqrt{T}\,(\hat{\mathbf{r}}_t^x - \mathbf{r}) \to^d N\left(\mathbf{0}, \mathbf{R}_t^x \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \mathbf{R}_t^{x\prime}\right) \tag{3.5}$$

where

$$\mathbf{R}_t^x = \frac{\partial \mathbf{r}_t^y}{\partial \boldsymbol{\theta}}$$

The only remaining difficulty is working out $\partial \mathbf{r}_t^y / \partial \boldsymbol{\theta}$ and $\partial \mathbf{r}_t^x / \partial \boldsymbol{\theta}$.

Each entry in the solution matrices $\mathbf{G}(\boldsymbol{\theta})$, $\mathbf{H}(\boldsymbol{\theta})$, and $\boldsymbol{\eta}(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$. Each matrix can be vectorized and the derivative of each entry can be taken with respect to $\boldsymbol{\theta}$. Denote the matrix derivatives of the vectorized solution matrices by

$$\mathbf{G}_\theta = \frac{\partial \operatorname{vec}(\mathbf{G})}{\partial \boldsymbol{\theta}}$$
$$\mathbf{H}_\theta = \frac{\partial \operatorname{vec}(\mathbf{H})}{\partial \boldsymbol{\theta}}$$
$$\mathbf{N}_\theta = \frac{\partial \operatorname{vec}(\boldsymbol{\eta})}{\partial \boldsymbol{\theta}}$$

These matrices are $n_y n_x \times n_\theta$, $n_x n_x \times n_\theta$, and $n_x n_e \times n_\theta$ respectively.

Consider first the horizon $t = 1$. The derivative of $r_1^y$ can be found through an application of the product rule for matrix derivatives, where the elements of each matrix are functions of a vector. Extending the results in (Lütkepohl, 1996, 10.5.5), we can write

$$\begin{aligned} \mathbf{R}_1^y = \frac{\partial \mathbf{r}_1^y}{\partial \theta} &= \frac{\partial (\mathbf{GH}\boldsymbol{\eta}\mathbf{e}_0)}{\partial \theta} \\ &= \mathbf{e}_0 \otimes \left[ (\mathbf{I}_e \otimes \mathbf{GH})\mathbf{N}_\theta + (\boldsymbol{\eta}' \otimes \mathbf{G})\mathbf{H}_\theta + (\mathbf{H}'\boldsymbol{\eta}' \otimes \mathbf{I}_y)\mathbf{G}_\theta \right] \end{aligned} \tag{3.6}$$

The derivative of $\mathbf{r}_1^x$ is, similarly,

$$\begin{aligned} \mathbf{R}_1^x = \frac{\partial \mathbf{r}_1^x}{\partial \boldsymbol{\theta}} &= \frac{\partial (\mathbf{H}\boldsymbol{\eta}\mathbf{e}_0)}{\partial \boldsymbol{\theta}} \\ &= \mathbf{e}_0 \otimes \left[ (\mathbf{I}_e \otimes \mathbf{H})\mathbf{N}_\theta + (\boldsymbol{\eta}' \otimes \mathbf{I}_x)\mathbf{H}_\theta \right] \end{aligned} \tag{3.7}$$

The derivative at an arbitrary horizon $t$ may be found if we have an expression for the derivative of $\mathbf{H}^t$. This derivative is determined by (Lütkepohl, 1996, 10.5.5)

$$\frac{\partial \operatorname{vec}(\mathbf{H}(\theta)^t)}{\partial \boldsymbol{\theta}} = \left[\sum_{i=1}^{t-1} (\mathbf{H}')^{t-i-1} \otimes \mathbf{H}^t\right] \mathbf{H}_\theta \tag{3.8}$$

Combining (3.6) with (3.8) produces $\mathbf{R}_t^y = \partial \mathbf{r}_t^y / \partial \boldsymbol{\theta}$ at any desired horizon, conditional on $\mathbf{e}_0$. Combining (3.7) with (3.8) produces $\mathbf{R}_t^x = \partial \mathbf{r}_t^x / \partial \boldsymbol{\theta}$ at any desired horizon, conditional on $\mathbf{e}_0$. Inserting these matrices into (3.4) and using the estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ generates asymptotic standard errors for impulse response functions at any horizon after an estimated DSGE model.

## 3.3 Small sample performance

Standard errors derived from the delta method are valid asymptotically. I next provide some evidence on their coverage properties in small samples. I set up a Monte Carlo experiment. The data-generating process is a three-equation New Keynesian model with two autoregressive forcing processes. Its six parameters $(\beta, \kappa, \rho_z, \rho_m, \sigma_z^2, \sigma_m^2)$ are all identified.

$$
\begin{aligned}
x_t &= E_t x_{t+1} - (r_t - E_t \pi_{t+1} - z_t) \\
\pi_t &= \beta E_t \pi_{t+1} + \kappa x_t \\
r_t &= \frac{1}{\beta}\pi_t + m_t \\
z_{t+1} &= \rho_z z_t + e^z_{t+1} \\
m_{t+1} &= \rho_m m_t + e^m_{t+1}
\end{aligned}
$$

The true parameter vector is set to $\theta_0 = (\beta = 0.7, \kappa = 0.2, \rho_z = 0.9, \rho_m = 0.3, \sigma_z^2 = 1, \sigma_m^2 = 1)$. The chosen value for $\beta$ in these simulations is smaller than values near unity that one typically sees in applied macroeconomic studies. These parameter values were chosen to lie well within the parameter space for which this model admits a solution; this choice improves convergence of the estimated model parameters considerably in small samples. I solve the model at the true values, then draw a simulation of $T$ observations from the solution. I then estimate the model parameters on the simulated data, and estimate impulse response functions along with standard errors. I run $N = 10,000$ replications for each of three sample sizes: $T = 120$, to represent a quarterly dataset over a 30-year period; $T = 250$, to represent a quarterly dataset over a 60-year period; and $T = 850$, to represent a monthly dataset over a 70-year period. The data-generating process was the same thoughout all three sample sizes to maintain comparability.

Table 3.1 provides results for the model parameters: mean of simulation point estimates

Table 3.1: Parameter estimates and rejection rate

| | | $T = 120$ | | $T = 250$ | | $T = 850$ | |
|---|---|---|---|---|---|---|---|
| Parameter | True value | mean | rejection rate | mean | rejection rate | mean | rejection rate |
| $\beta$ | 0.7 | 0.702 | 0.064 | 0.701 | .054 | 0.70 | .050 |
| $\kappa$ | 0.2 | 0.240 | 0.062 | 0.219 | .049 | 0.208 | .048 |
| $\rho_m$ | 0.3 | 0.276 | 0.057 | 0.287 | .054 | 0.295 | .051 |
| $\rho_z$ | 0.8 | 0.774 | 0.059 | 0.787 | .057 | 0.796 | .052 |
| $\sigma_m$ | 1.0 | 0.990 | 0.064 | 0.995 | .054 | 0.998 | .054 |
| $\sigma_z$ | 1.0 | 1.14 | 0.059 | 1.119 | .048 | 1.020 | .045 |

Note: the Table summarizes parameter estimates across 10,000 simulations for three sample sizes indicated. The rejection rate is the proportion of runs in which the Wald test of the parameter value against the true value was rejected.

and rejection rate. For each replicate, I record the estimated parameter value and a Wald test of the parameter's value against the true value. For the sample size of 120, the simulated rejection rates are somewhat higher than the nominal size: most reject about six percent of the time in a nominal 5% test. For the simulations with sample sizes of $T = 250$ and $T = 850$, the Wald test rejection rate is statistically indistinguishable from 0.05. That is, the asymptotic standard errors taken from the inverse of the negative Hessian provide good coverage for samples of this length.

To construct my estimated impulse responses, I simulate a shock in period 0 to each state variable separately and trace out the response of each control and state variable for $t = 0$ to $t = 8$. Hence there are ninety point estimates: the response of five variables, each to two shocks, for nine periods. The true impulse response can be calculated for each from the model's solution matrices, so there are ninety Wald tests against the true value.

Two representative results are shown. Table 3.2 provides results for an impulse to the $z$ shock on itself. The first column records the horizon. The second column records the true impulse response. Since $z$ shock is autoregressive with $\rho_z = 0.8$, a unit shock in period 0 implies that the true impulse response is a geometrically decaying sequence $0.8^s$ for $s = 0, 1, \ldots, 8$. The third column records the mean of estimated impulse responses. The fourth column records the rejection rate of a Wald test of nominal size 5% against the true value. The the final two columns provide a 95% confidence interval for the rejection rate across simulations. At a sample size of 120, the Wald test over–rejects somewhat at every horizon, with the rejection rate increasing by horizon. The test with nominal size 5% rejects between 7% and 12% of the time, depending on horizon. Reported asymptotic confidence intervals for impulse response functions will be somewhat thinner than appropriate. At a sample size of 250, typical for quarterly macroeconomic data, the rejection rate is somewhat higher than 0.05 and increases somewhat with the horizon. At later horizons, the empirical rejection rate is statistically significantly different from 0.05, the nominal size of the test. Reported asymptotic confidence intervals continue to be somewhat thinner than appropriate, but the distortion is lessened considerably. At 850 observations, a typical dataset for monthly macroeconomic data, the rejection rate is nearly its nominal size and is statistically indistinguishable from 0.05 at all horizons.

Table 3.3 provides results for an impulse to the $z$ shock on $\pi$. The response of $\pi$ to an impulse $z$ at horizon $s$ is the policy matrix coefficient multiplied by $\rho_z^s$. The empirical rejection rate for $T = 120$ is larger than its 5% nominal size, ranging from 7.5% to over 17% as the horizon increases. The empirical rejection rate for $T = 250$ is closer to the nominal size, but still over-rejects considerably as the horizon increases. The empirical rejection rate for $T = 850$ is closer to the nominal size. As before, reported confidence intervals based on asymptotic arguments will be somewhat thinner than is appropriate for datasets of the size typically seen in macroeconomics.

The delta-method standard errors provide inference that is only appropriate asymptotically. While the standard errors for parameters themselves are valid at conventional sample sizes seen in macro time–series, and also provide appropriate inference for the first few periods of the impulse response function, the reported standard errors for impulse responses may be somewhat too thin at longer horizons with sample sizes that are typical for quarterly macro time–series. Performance is better for a sample size that is conventional for the monthly frequency. For small samples, it may be worthwhile to compute robust standard errors. "Robust" here refers to the Huber/White/sandwich estimator of the covariance matrix, which is appropriate for state-space models with non-normal errors and tends to be more conservative than the usual estimator, which directly uses the inverse of the negative Hessian of the likelihood function. The Huber/White/sandwich estimator can be applied to linearized DSGE models as estimated by maximum likelihood. Robust standard errors will will be overly conservative for the parameter estimates themselves but provide more realistic inference for the impulse response functions, even in the case of normally distributed errors. An alternative, explored in Cho and Moreno (2006), is to employ parametric bootstrap techniques to improve coverage in small sample situations.

Table 3.2: Response of $z$ to $z$ shock

| Horizon | True value | Mean | Rejection rate | [95% conf inter] | |
|---|---|---|---|---|---|
| *sample size:* $T = 120$ | | | | | |
| 0 | 1.000 | 1.010 | 0.070 | 0.064 | 0.075 |
| 1 | 0.800 | 0.836 | 0.086 | 0.081 | 0.092 |
| 2 | 0.640 | 0.670 | 0.110 | 0.103 | 0.116 |
| 3 | 0.512 | 0.531 | 0.108 | 0.102 | 0.115 |
| 4 | 0.410 | 0.419 | 0.087 | 0.082 | 0.093 |
| 5 | 0.328 | 0.332 | 0.084 | 0.079 | 0.090 |
| 6 | 0.262 | 0.265 | 0.092 | 0.086 | 0.098 |
| 7 | 0.210 | 0.210 | 0.102 | 0.096 | 0.108 |
| 8 | 0.168 | 0.167 | 0.116 | 0.110 | 0.123 |
| *sample size:* $T = 250$ | | | | | |
| 0 | 1.000 | 1.028 | 0.052 | 0.048 | 0.057 |
| 1 | 0.800 | 0.837 | 0.064 | 0.059 | 0.069 |
| 2 | 0.640 | 0.666 | 0.085 | 0.080 | 0.091 |
| 3 | 0.512 | 0.526 | 0.086 | 0.081 | 0.092 |
| 4 | 0.410 | 0.416 | 0.073 | 0.068 | 0.078 |
| 5 | 0.328 | 0.330 | 0.065 | 0.060 | 0.070 |
| 6 | 0.262 | 0.262 | 0.068 | 0.064 | 0.074 |
| 7 | 0.210 | 0.208 | 0.076 | 0.071 | 0.081 |
| 8 | 0.168 | 0.165 | 0.084 | 0.078 | 0.089 |
| *sample size:* $T = 850$ | | | | | |
| 0 | 1.000 | 1.023 | 0.045 | 0.041 | 0.049 |
| 1 | 0.800 | 0.815 | 0.051 | 0.047 | 0.055 |
| 2 | 0.640 | 0.648 | 0.060 | 0.056 | 0.065 |
| 3 | 0.512 | 0.516 | 0.061 | 0.056 | 0.065 |
| 4 | 0.410 | 0.410 | 0.053 | 0.048 | 0.057 |
| 5 | 0.328 | 0.327 | 0.046 | 0.042 | 0.051 |
| 6 | 0.262 | 0.261 | 0.048 | 0.044 | 0.052 |
| 7 | 0.210 | 0.208 | 0.052 | 0.048 | 0.057 |
| 8 | 0.168 | 0.166 | 0.055 | 0.051 | 0.060 |

True impulse response of $z_t$ to a $e_z$ shock at $\theta_0$, mean of calculated impulse responses, empirical rejection rate, and 95% confidence interval for rejection rate. 10,000 simulations were conducted at three sample sizes each, as listed.

Table 3.3: Response of $\pi$ to $z$ shock

| Horizon | True value | Mean | Rejection rate | [95% conf inter] | |
|---|---|---|---|---|---|
| *sample size:* $T = 120$ | | | | | |
| 0 | 0.936 | 0.915 | 0.075 | 0.069 | 0.081 |
| 1 | 0.749 | 0.713 | 0.098 | 0.092 | 0.104 |
| 2 | 0.599 | 0.556 | 0.099 | 0.093 | 0.105 |
| 3 | 0.479 | 0.435 | 0.110 | 0.104 | 0.116 |
| 4 | 0.383 | 0.343 | 0.121 | 0.114 | 0.127 |
| 5 | 0.307 | 0.271 | 0.136 | 0.129 | 0.143 |
| 6 | 0.245 | 0.215 | 0.147 | 0.140 | 0.154 |
| 7 | 0.196 | 0.171 | 0.161 | 0.154 | 0.168 |
| 8 | 0.157 | 0.137 | 0.173 | 0.165 | 0.180 |
| *sample size:* $T = 250$ | | | | | |
| 0 | 0.936 | 0.926 | 0.062 | 0.057 | 0.067 |
| 1 | 0.749 | 0.730 | 0.077 | 0.072 | 0.083 |
| 2 | 0.599 | 0.576 | 0.077 | 0.072 | 0.082 |
| 3 | 0.479 | 0.456 | 0.084 | 0.079 | 0.090 |
| 4 | 0.383 | 0.362 | 0.091 | 0.085 | 0.096 |
| 5 | 0.307 | 0.287 | 0.099 | 0.093 | 0.105 |
| 6 | 0.245 | 0.229 | 0.106 | 0.100 | 0.113 |
| 7 | 0.196 | 0.183 | 0.115 | 0.109 | 0.122 |
| 8 | 0.157 | 0.146 | 0.123 | 0.117 | 0.130 |
| *sample size:* $T = 850$ | | | | | |
| 0 | 0.936 | 0.934 | 0.053 | 0.049 | 0.058 |
| 1 | 0.749 | 0.743 | 0.055 | 0.051 | 0.060 |
| 2 | 0.599 | 0.592 | 0.055 | 0.051 | 0.060 |
| 3 | 0.479 | 0.472 | 0.058 | 0.053 | 0.062 |
| 4 | 0.383 | 0.376 | 0.061 | 0.056 | 0.066 |
| 5 | 0.307 | 0.301 | 0.063 | 0.058 | 0.068 |
| 6 | 0.245 | 0.240 | 0.067 | 0.062 | 0.072 |
| 7 | 0.196 | 0.192 | 0.071 | 0.066 | 0.076 |
| 8 | 0.157 | 0.153 | 0.074 | 0.069 | 0.080 |

True impulse response of $\pi_t$ to a $e_z$ shock at $\theta_0$, mean of calculated impulse responses, empirical rejection rate, and 95% confidence interval for rejection rate. 1,000 simulations were conducted at three sample sizes each, as listed.

# Bibliography


Andrews, Donald WK. 1993. "Tests for parameter instability and structural change with unknown change point." *Econometrica* :821–856.

Basu, Susanto, John G Fernald, and Miles S Kimball. 2006. "Are technology improvements contractionary?" *American Economic Review* 96 (5):1418–1448.

Benati, Luca. 2008. "Investigating inflation persistence across monetary regimes." *The Quarterly Journal of Economics* 123 (3):1005–1060.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3):383–398.

Cho, Seonghoon and Antonio Moreno. 2006. "A Small-Sample Study of the New-Keynesian Macro Model." *Journal of Money, Credit and Banking* 38 (6):1461–1481.

Christiano, L.J., M. Eichenbaum, and C.L. Evans. 2005. "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of political Economy* 113 (1):1–45.

Coibion, Olivier. 2012. "Are the effects of monetary policy shocks big or small?" *American Economic Journal: Macroeconomics* 4 (2):1–32.

Coibion, Olivier and Yuriy Gorodnichenko. 2015. "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics* 7 (1):197–232.

Coibion, Olivier, Yuriy Gorodnichenko, and Rupal Kamdar. 2018. "The formation of expectations, inflation, and the phillips curve." *Journal of Economic Literature* 56 (4):1447–91.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar. 2018. "How do firms form their expectations? new survey evidence." *American Economic Review* 108 (9):2671–2713.

Elsby, Michael, Bart Hobijn, and Ayşegül Şahin. 2013. "The decline of the US labor share." *Brookings Papers on Economic Activity* 2013 (2):1–63.

Fuhrer, Jeffrey C. 2013. "Real Expectations: Replacing Rational Expectations with Survey Expectations in Dynamic Macro Models." *Boston Fed Working Paper 12-19* Most recent version dated 2013-01-04.

Galı, J. and M. Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44 (2):195–222.

Hamilton, James D. 1996. "This is what happened to the oil price-macroeconomy relationship." *Journal of Monetary Economics* 38 (2):215–220.

———. 2018. "Why you should never use the Hodrick-Prescott filter." *Review of Economics and Statistics* 100 (5):831–843.

Ireland, Peter N. 2004. "Technology Shocks in the New Keynesian Model." *Review of Economics and Statistics* 86 (4):923–936.

Jordà, Òscar. 2005. "Estimation and inference of impulse responses by local projections." *American Economic Review* 95 (1):161–182.

Karabarbounis, Loukas and Brent Neiman. 2013. "The global decline of the labor share." *The Quarterly journal of economics* 129 (1):61–103.

King, Robert G and Mark W Watson. 2012. "Inflation and Unit Labor Cost." *Journal of Money, credit and Banking* 44:111–149.

Klein, Paul. 2000. "Using the generalized Schur form to solve a multivariate linear rational expectations model." *Journal of Economic Dynamics and Control* 24 (10):1405–1423.

Lütkepohl, Helmut. 1996. *Handbook of Matrices.* New York: John Wiley and Sons.

Mineyama, Tomohide. 2018. "Downward Nominal Wage Rigidity and Inflation Dynamics during and after the Great Recession." .

Morris, Stephen D. 2017. "DSGE pileups." *Journal of Economic Dynamics and Control* 74 (C):56–86.

Nunes, Ricardo. 2010. "Inflation dynamics: the role of expectations." *Journal of Money, Credit and Banking* 42 (6):1161–1172.

Ramey, Valerie A. 2016. "Macroeconomic shocks and their propagation." In *Handbook of macroeconomics*, vol. 2. Elsevier, 71–162.

Roberts, John M. 1995. "New Keynesian Economics and the Phillips Curve." *Journal of Money, Credit and Banking* 27 (4):975–84.

Romer, Christina D and David H Romer. 2004. "A new measure of monetary shocks: Derivation and implications." *American Economic Review* 94 (4):1055–1084.

———. 2010. "The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks." *American Economic Review* 100 (3):763–801.

Rotemberg, Julio. 1982. "Sticky Prices in the United States." *Journal of Political Economy* 90 (6):1187–1211.

Sbordone, Argia M. 2002. "Prices and unit labor costs: a new test of price stickiness." *Journal of Monetary Economics* 49 (2):265–292.

Schmitt-Grohé, Stephanie and Martın Uribe. 2004. "Solving dynamic general equilibrium models using a second-order approximation to the policy function." *Journal of Economic Dynamics and Control* 28 (4):755–775.

Smets, F. and R. Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3):586–606.

Wieland, Johannes F and Mu-Jeung Yang. 2016. "Financial dampening." Tech. rep., National Bureau of Economic Research working paper 22141.

Yun, Tack. 1996. "Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles." *Journal of Monetary Economics* 37 (2):345–370.