The Dissertation Committee for Julian Felix Ludwig
certifies that this is the approved version of the following dissertation:

# Essays in Empirical Macroeconomics

Committee:

Olivier Coibion, Supervisor

Saroj Bhattarai, Co-Supervisor

Haiqing Xu

Yuriy Gorodnichenko

# Essays in Empirical Macroeconomics

by

**Julian Felix Ludwig**

**DISSERTATION**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2019

Dedicated to my parents.

# Acknowledgments

I am grateful to Olivier Coibion, Saroj Bhattarai and Haiqing Xu for their invaluable guidance and support. I also benefited from conversations with Stefano Eusepi, Christoph Boehm, Tatevik Sekhposyan, Daniel Ackerberg, Chi Zhang, Jessie Coe, Kevin Kuruc, Cooper Howes, Choongryul Yang, Andrés Méndez Ruiz, and Shaofei Jiang as well as with many seminar participants from the University of Texas at Austin, Texas A&M University, and the Network of Swiss Economists Abroad.

# Essays in Empirical Macroeconomics

Publication No. ___________________

Julian Felix Ludwig, Ph.D.
The University of Texas at Austin, 2019

Supervisor: Olivier Coibion
Co-Supervisor: Saroj Bhattarai

This dissertation examines how expectations are formed and how they interact with economic activities. Beliefs about economic outcomes vary with timing and accuracy of information, which have important implications for macroeconomic dynamics. The importance of expectations has long been emphasized in rational expectations (RE) models (see e.g. Lucas 1972, 1976; Kydland and Prescott 1982), and diffusion of information has been modeled in many ways (see e.g. Beaudry and Portier 2004, 2006; Mankiw and Reis 2002; Woodford 2003; Sims 2003). My work builds on this literature and aims to improve the understanding of information structure, formation of beliefs, and decision-making, and how they contribute to macro business cycles.

In the first chapter, I point out how identification of full information rational expectations (FIRE) models suffers from Manski's (1993) reflection problem. I extend the standard rational expectations (RE) model to allow for a more general information structure and introduce a new framework to identify the generalized model with forecaster data. Identification is no longer subject to the reflection problem when two changes are made to the information structure: the addition

of news shocks and imperfect information. News shocks provide additional variation in expectations about the future. Imperfect information provides changes in beliefs about past states, through which the feedback between expectations and decisions goes only in one direction. Expectations data are consistent with both. An application to Greenbook forecasts illustrates the importance of both news shocks and learning about the past. When I apply this framework to a Blanchard and Quah (1989) decomposition, I reach qualitatively new results. For example, expansionary supply shocks decrease unemployment. Supply shocks are also particularly subject to both news and information rigidities, so relaxing the information structure is key to correctly identifying these shocks.

In the second chapter, I discover how both good and bad news shocks coincide with higher uncertainty on impact. This new stylized fact is robust to different empirical models of the news shocks literature and different proxies for U.S. macro uncertainty. The new stylized fact has implications in three fields. First, bad news shocks produce the dynamics discovered in the uncertainty literature: spikes in uncertainty are followed by drops in output. I show that there is indeed some overlap between bad news and uncertainty shocks, as the effect of an uncertainty shock gets weaker when controlling for bad news shocks. Second, I show that the close relationship between news shocks and uncertainty seems to be also responsible for the close relationship between quarterly stock returns and stock market volatility - a proxy for uncertainty. This contributes to the finance literature that works on this relationship. Third, introducing a non-linear empirical model, I find additional asymmetries in the responses to news shocks due to the asymmetric response of uncertainty. This contributes directly to the news shocks literature.

An important conclusion of chapters one and two is that economic shocks vary with availability of information. The third chapter deals with such heterogeneity. I relax the assumption that

economic shocks of the same type are homogeneous, respectively, always have the same effect. Instead, I argue that economists identify a shock that consists of a variety of heterogeneous components. For example, a technology shock is the sum of all disaggregate technology shocks, from innovations in marketing up to inventions in the manufacturing process, which all have different effects on the economy. I discuss how standard identification methods can identify the shocks of interest despite this heterogeneity. I find that the weights on the shock components depend on the identification strategy so that different identification strategies produce different effects. This could explain why different macro papers often identify different responses to the same shock, in the same country, and over the same time period.

# List of Tables

# List of Figures

# Table of Contents

# Chapter 1

# Identification of Rational Expectations Models Under Information Frictions

## 1.1 Introduction

In modern micro-founded macro models, the decisions of economic agents are inherently forward-looking and therefore depend on their expectations about the future. But if their expectations are also formed based on the current state, as is generally assumed in full-information models, it becomes difficult to determine to what extent expectations affect actions and vice-versa. I relate this simultaneity to Charles F. Manski's (1993) reflection problem and introduce a new way to deal with its implications for identification.

The importance of expectations has long been emphasized in rational expectations (RE) models (see e.g. Robert E Lucas, 1972; Robert E. Lucas, 1976; Finn E. Kydland and Edward C. Prescott, 1982). This paper provides a new methodology to identify the parameters of RE models using data on the expectations of economic agents at multiple horizons. Specifically, I relax the full-information assumption and allow for information to diffuse to agents both before and after shocks are realized, which allows me to match both forecasts and backcasts in the data. The resulting variation in expectations is no longer proportional to current actions so that identification is possible. This flexible information structure brings together the literature on news shocks (see Paul Beaudry and Franck Portier, 2004, 2006), where information arrives before impact, and the literature on information frictions (see N. Gregory Mankiw and Ricardo Reis, 2002; Michael Woodford, 2003; Christopher A. Sims, 2003), where agents gather relevant information only after the period is realized. Combining both dimensions is the key to identifying the parameters of the model.

I relate the identification issue to Manski (1993), who describes the role of expectations about competitors and peers in social interaction models. His well-known *reflection problem* states that when expectations are modelled simply as the average among individuals, the researcher cannot distinguish whether expectations change individual behaviour or if they simply reflect behaviour without causing it. I extend his proposition to show that full information rational expectations (FIRE) models suffer from the same type of reflection problem as well: since information is only gathered in one period, expectations about the future are proportional to realizations today. The researcher can therefore not distinguish the direct effect of a shock on the economy from the indirect effect of observing that shock. The FIRE literature successfully bypasses this reflection problem by imposing a set of parameter restrictions on the FIRE model typically derived from a micro-founded model. I provide a new way to deal with this identification issue by incorporating data on expectations at multiple horizons and by relaxing the full information assumption.

There are many ways to relax full information (FI). Mankiw and Reis (2002) introduce a *sticky information* approach where agents update their expectations infrequently, but when they update their expectations, they fully observe the state. Woodford (2003) models information rigidities with a *noisy information* approach where agents receive noisy signals about the state of the economy. Sims (2003) and Bartosz Maćkowiak and Mirko Wiederholt (2009) make acquiring information costly so that agents need to choose whether they want to pay the cost to get information. The solution of this problem is characterized by *rational inattention*, where agents choose to deviate from full information. Models with such information rigidities allow for information to arrive *after* the shocks are realized so that agents no longer fully observe the current state of the economy. In contrast, news shock models allow for information to arrive *before* the shock hits the economy. Beaudry and Portier (2004) introduce news shocks as noisy signals about future developments of the economy, while Beaudry and Portier (2006), Joshua Mark Davis (2007), and Lawrence J. Christiano, Cosmin Ilut, Roberto Motto and Massimo Rostagno (2010) model news shocks as future shocks that are fully observed today. My paper combines the two branches of the literature to formalize a much more general information structure than what is allowed in standard macroeconomic models.

The core identification property is that future outcomes only have an effect on today's economy if agents observe these outcomes in advance. Hence, any fluctuations that are unobserved cannot be caused by future outcomes. These unobserved fluctuations are therefore fully backward-looking and can be used to identify the effect of current on future outcomes. Changes in expectations about the previous period, defined as *backcast revisions*, collect these unobserved fluctuations. Hence, data on expectations about both the contemporaneous and the previous period are sufficient to identify one direction of the relationship between outcomes and expectations. The other direction is identified when data on expectations about the next period is available. Given the identified effect of current outcomes on expectations of future outcomes, the remaining variation in expectations of the next period must come from additional information that is obtained today about the future. The effect of the remainder on current outcomes thus identifies the effect of expectations so that the forward- and backward-looking components of the RE model are identified. Key for identification is the timing assumption that expectations can only depend on information obtained up until today, but not on information obtained tomorrow. Timing restrictions on information sets are common in the production function literature (see G. Steven Olley and Ariel Pakes, 1996; Richard Blundell and Stephen Bond, 2000; James Levinsohn and Amil Petrin, 2003). To my knowledge, this is the first paper that uses the variation generated by relaxing full information to separately identify the parameters governing the backward- and forward-looking dynamics of RE models.

The proposed strategy identifies the forward- and backward-looking components simultaneously, without imposing additional restrictions on the model equations. Instead of imposing a particular structural model, this approach nests all models that have the form of RE models with a flexible information structure. Hence, estimation is less subject to model specifications other than the choice which variables to include, the number of lags and leads, and how the shocks are orthogonalized. Identification instead relies on the assumption that data on expectations across horizons, both future and past, is available and that this data correctly captures beliefs of agents. Moreover, identification requires a positive variance of backcast revisions, and noncollinearity between now- and forecasts, features which appear to be consistent with the data. Hence, I impose a completely different set of assumptions to identify the RE model than what is common in the literature.

I implement the identification strategy using data on expectations from the Greenbook, a collection of forecasts from the U.S. Federal Reserve. In the first application, I estimate an unrestricted multivariate RE model with output, consumption, and investment growth, as well as inflation. I find that about half the information about shocks is gathered before and during realization, while the other half is obtained only after the shock materializes. This provides evidence for information rigidities, where agents don't fully observe the current state of the economy. Moreover, around one fourth of the information is collected before realization, which provides support for the news shock literature. I then produce counterfactuals by shifting arrival of information to simulate a full information environment, where everything is observed on impact. In this counterfactual, persistence of all variables significantly declines. Hence, the information structure seems to be a significant driver of the persistence in macro variables.

In the second application, I estimate a RE model with output growth and changes in unemployment, and orthogonalize the shocks following Olivier Jean Blanchard and Danny Quah (1989): demand shocks are assumed to be transitory shocks, while supply shocks are the only shocks with a long run impact on output. The results indicate that demand shocks are much better observed than supply shocks, hence, agents seem to be better informed about the demand side of the economy, and less informed about the production side. Moreover, I find on average a significant response to supply shocks before impact, which is line with the news shock literature modelling anticipated supply side shocks. Overall, both demand and supply shocks increase output and decrease unemployment which is consistent with standard real business cycles (RBC) models. The finding that supply shocks decrease unemployment is different from the conclusion of Blanchard and Quah (1989), who find increased unemployment in response to supply shocks. This illustrates the importance of properly controlling for timing of information arrival and processing when identifying economic shocks.

This paper relates to the literature on structural vector autoregressions (SVARs), where economic shocks are identified with reduced form models (see Christopher A Sims, 1980). Similar to SVARs, my identification strategy does not require one to specify the structural equations of the RE model. However, while SVARs identify the reduced form version of the RE model, this paper identifies the structural version directly, where expectations about the future and dependence on past are identified separately.

My paper builds on and contributes to the literature on news shocks. Beaudry and Portier (2004, 2006) introduce the notion of TFP news shocks as changes in current expectations about future productivity, and they find that these news shocks have real effects today, even though they only materialize in the future. Beaudry and Portier (2006); Paul Beaudry and Franck Portier (2014), Barsky and Sims (2011); Robert B. Barsky and Eric R. Sims (2012), Paul Beaudry, Martial Dupaigne and Franck Portier (2011), Robert B. Barsky, Susanto Basu and Keyoung Lee (2015) and others identify TFP news shocks in SVARs as changes in current expectations, measured for example in terms of stock prices or consumer confidence, that are orthogonal to current but contribute to future TFP. Today's effects of the news shocks are then considered as effects of expectations, while the shock only realizes in the future when TFP increases. These papers therefore rely on structural assumptions about the dynamic effects of news shocks on economic variables. My approach relies on a different set of assumptions, yet yields results that confirm the importance of news shocks for macroeconomic dynamics.

This paper also contributes to the literature on VAR invertibility, which refers to the ability to rewrite the RE model as a reduced form VAR of observables (see Jesús Fernández-Villaverde, Juan F. Rubio-Ramírez, Thomas J. Sargent and Mark W. Watson, 2007). Non-invertibility occurs if there are unobserved state variables causing a bias in VAR estimation (see Mark W. Watson, 1986). Models with news shocks are particularly vulnerable to non-invertibility, because VAR variables might not be able to capture the information used to predict news shocks (see Eric M. Leeper, Todd B. Walker and Shu-Chun Susan Yang, 2013). Watson (1986), Christopher A. Sims and Tao Zha (2006), and Eric R. Sims (2012) address the invertibility issue by including forward-looking variables. I introduce an alternative way to avoid non-invertibility by estimating the reduced form model (VAR) with backcast revisions, as described above.

An essential feature in Manski (1993) is that beliefs are modelled rather than observed. Whether true expectations are indeed equal to modelled expectations cannot be tested using data on realizations alone. Charles F. Manski (2004) thus recommends usage of data on expectations, namely: "econometric analysis of decision making with information cannot prosper on choice data alone" (see Manski, 2004, p. 1330). Macroeconomists have access to rich data on expectations on macro indicators at multiple horizons, which is provided by the Federal Reserve Bank of Philadelphia, the University of Michigan and other institutions, and there is a growing literature that makes use of this data.

Olivier Coibion and Yuriy Gorodnichenko (2015) use survey data on expectations to test for both, full information (FI) and rational expectations (RE). They reject FIRE, and find that rejection most likely reflects deviations from FI rather than RE. This is in line with Klaus Adam and Mario Padula (2011), who show that survey expectations satisfy the law of iterated expectations, a law that is satisfied under RE. Based on these findings and in line with the literature on information frictions and news shocks, I only relax FI and keep the RE assumption.

There is already a growing literature using data on expectations in RE models. Marco Del Negro and Stefano Eusepi (2011) and Marco Del Negro and Frank Schorfheide (2013) use data on both realizations as well as on expectations to model inflation expectations in DSGE models, while Thuy Lan Nguyen and Wataru Miyamoto (2014), Yasuo Hirose and Takushi Kurozumi (2012), and Fabio Milani and Ashish Rajrhandari (2012) include data on expectations in DSGE models to measure the effect of TFP news shocks, which are anticipated productivity shocks. Jeff Fuhrer (2017) embeds survey forecasts in a standard DSGE model and demonstrates that much of the persistence in aggregate data can be attributed to slow moving expectations. Precursors of this literature estimate univariate models, for example, John M Roberts (1995) uses survey measures to estimate the Philips curve. Instead of using survey data directly, Jordi Galí and Mark Gertler (1999) first regress future variables on information sets to model expectations, and then they estimate the Philips curve based on these projections. While I relax the information structure to match data on expectations, these models either rely on measurement errors, or deviations from rationality (RE), as these models recognize that observed expectations do not always match model predictions.

Section 1.2 illustrates the reflection problem in rational expectations models. Section 1.3 provides intuition for the new identification method with a univariate model for inflation. Section 1.4 generalizes the information structure and identifies the model. Section 1.5 discusses identifying restrictions to extract economic shocks. Section 1.6 examines forecaster data from the Federal

Reserve and uses this data to estimate the model. Section 3.6 concludes.

## 1.2 Manski's Reflection Problem in RE Models

Consider a standard linear rational expectations model:

$$\mathbf{y}_t = \mathbf{\Gamma}\mathbf{y}_{t-1} + \mathbf{\Phi} E[\mathbf{y}_{t+1}|\mathcal{F}_t] + \mathbf{\Pi}\mathbf{u}_t, \tag{1.1}$$

where $\mathbf{y}_t$ is an $N$-dimensional vector of variables and $\mathbf{u}_t$ is an $N$-dimensional vector of mutually exclusive shocks with zero-mean, i.e. $E[\mathbf{u}_t] = 0$, identity variance-covariance matrix, i.e. $E[\mathbf{u}_t\mathbf{u}_t'] = I$, and no correlation over time, i.e. $E[\mathbf{u}_t\mathbf{u}_{t+k}'] = 0$, $k \neq 0$. Moreover, $\mathcal{F}_t$ is all the information the agent obtains up to period $t$.[1] $\mathbf{\Gamma}$, $\mathbf{\Phi}$, and $\mathbf{\Pi}$ are $N \times N$ matrices of parameters, which are typically derived by linearising a micro-founded model.

The trademark of rational expectations models is the role of expectations about the future, which makes the model not just backward-looking, but also forward-looking. The workhorse macroeconomic models (e.g. RBC or New Keynesian models) would satisfy the following assumptions:

**Rational Expectations (RE).** *Expectations are consistent with model.*

The notion of rational expectations is originally defined by John F. Muth (1961), who suggests to model expectations the same way as predictions of the underlying model. This notion is introduced to macroeconomics by Lucas (1972), promoting RE in his later called Lucas's (1976) critique. In particular, he points out that expectations should not be modelled arbitrarily in a static way, instead,

[1] Formally, $\mathcal{F}_t$ is denoted by a collection of events (i.e. $\sigma$-algebra) that can be assigned probabilities by using all the information from the past. If the information represented by a random variable $w_s$ belongs to $\mathcal{F}_t$ (i.e. the $\sigma$-algebra generated by the random variable $w_s$), then we have $E[w_s|\mathcal{F}_t] = w_s$.

macro models should allow for expectations to change over time, and they should be consistent within the model.

**Full Information (FI).** *Shocks are only and fully revealed at realization:*

$$E[\mathbf{u}_t|\mathcal{F}_{t+s}] = \begin{cases} E[\mathbf{u}_t], & s < 0, \qquad (i) \\ \mathbf{u}_t, & s \geq 0, \qquad (ii) \end{cases}$$

*and initial conditions are known:* $E[\mathbf{y}_0|\mathcal{F}_0] = \mathbf{y}_0$.

FI means *all* the information about shocks $\mathbf{u}_t$ is gathered at time of realization. Specifically, FI.i states that no information is gathered before realization, and FI.ii states that no information is gathered after realization, as realized shocks are fully observed at the moment they are realized. Recall that expectations of shocks are normalized to be zero, i.e. $E[\mathbf{u}_t] = 0$. Thus, FI.i is also denoted as the zero conditional mean (ZCM) condition. FI implies that the agent observes all realized variables and shocks, $\mathbf{y}_{t-s}$ and $\mathbf{u}_{t-s}$, for all $s \geq 0$, and predictions about the future are only based on these realizations. Models that satisfy both FI and RE are standard and are sometimes referred to as full information rational expectations (FIRE) models (see e.g. Olivier Coibion and Yuriy Gorodnichenko, 2012).

Under RE and FI.i, the shocks disappear when taking expectations of future variables so that for all $h \geq 1$:

$$E[\mathbf{y}_{t+h}|\mathcal{F}_t] = \mathbf{\Gamma} E[\mathbf{y}_{t+h-1}|\mathcal{F}_t] + \mathbf{\Phi} E[\mathbf{y}_{t+h+1}|\mathcal{F}_t]. \tag{1.2}$$

Inspection of (1.2) reveals the following:

**Lemma 1** (Invertibility of Future)**.** *Under RE and FI.i, expectations about the future are invertible,* $E[\mathbf{y}_{t+h}|\mathcal{F}_t] = \mathbf{A}E[\mathbf{y}_{t+h-1}|\mathcal{F}_t]$, $h \geq 1$, *where* $\mathbf{A}$ *is the fixed point solving* $\mathbf{A} = (I - \mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}$.

**Proof**: see appendix 1.8.1.

Lemma 1 states that it is sufficient to know expectations about realized variables, $E[\mathbf{y}_t|\mathcal{F}_t]$, to infer expectations about the future, $E[\mathbf{y}_{t+h}|\mathcal{F}_t]$, for all $h \geq 1$. Namely, $E[\mathbf{y}_{t+h}|\mathcal{F}_t] = \mathbf{A}^h E[\mathbf{y}_t|\mathcal{F}_t]$. Under RE and FI.i, lemma 1 is a sufficient condition to identify $\mathbf{A}$ and residuals that are independent of past information, $\mathcal{F}_{t-1}$, which is referred to invertability or fundamentalness in the VAR literature (see Watson, 1986; Fernández-Villaverde et al., 2007; Leeper, Walker and Yang, 2013; Sims, 2012, and discussion of proposition 1.2.1).

Lemma 1 means in particular that for $h = 1$ and under FI, $E[\mathbf{y}_{t+1}|\mathcal{F}_t] = \mathbf{A}\mathbf{y}_t$. There is therefore no independent variation in expectations in system (1.1), which prevents the separate identification of forward- and backward-looking components in (1.1). More formally, the rational expectations model suffers from Manski's reflection problem, which is a wel-known issue for identification in social interaction models (see Manski, 1993).[2] Proposition 1.2.1 uses a similar econometric structure as Manski (1993) and extends the reflection problem from a cross-sectional environment to rational expectations models:

**Proposition 1.2.1** (Reflection Problem)**.** *(a) Under RE and FI, parameters* $\mathbf{\Phi}$ *and* $\mathbf{\Gamma}$ *are not identified separately. Moreover, (b) under RE and FI.i, the composite parameter* $\mathbf{A} = (I - \mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}$ *and residuals* $E[\mathbf{y}_{t+k}|\mathcal{F}_t] - E[\mathbf{y}_{t+k}|\mathcal{F}_{t-1}]$ *are identified by regressing* $E[\mathbf{y}_{t+k}|\mathcal{F}_t]$ *on* $E[\mathbf{y}_{t+k-1}|\mathcal{F}_{t-1}]$ *for any* $k \geq 0$.

**Proof**: see appendix 1.8.2.

[2]Manski (1993) discovers the reflection problem in models with endogenous social effects, which describe how individuals' behaviour is influenced by the behaviour of a group. The reflection problem arises in Manski (1993) when data cannot identify whether average behaviour affects the behaviour of individuals, or whether the average simply reflects individuals' behaviour without causing it. Manski (1993) compares the problem with observing the simultaneous movements of a person with her reflection in the mirror, as this observation does not reveal whether the mirror causes the movements or reflects them.

Proposition 1.2.1 shows that in the FIRE model, data reveals how the economy propagates (i.e. $\mathbf{A}$) but cannot tell whether this persistence is caused by forward-looking or backward-looking behaviour. Macroeconomists therefore rely on a series of parameter restrictions to extract both $\mathbf{\Gamma}$ and $\mathbf{\Phi}$. For instance, if one assumes $\mathbf{\Phi} = 0$, then $\mathbf{\Gamma} = \mathbf{A}$. This identifying restriction implies that expectations don't matter, and dynamics in the model are entirely backward-looking. While unusual in modern macro models, this is precisely the environment used in empirical models like Thomas Laubach and John C. Williams (2003) which is the workhorse approach used by central banks to estimate the natural rate of interest (see e.g. James Bullard, 2018). The opposite approach is to set $\mathbf{\Gamma} = 0$ so that $\mathbf{\Phi} = \mathbf{A}^{-1}$. In this case, dynamics are entirely forward-looking, as is the case for example in the workhorse New Keynesian model (see e.g. Richard Clarida, Jordi Galí and Mark Gertler, 2000). This implies that information is very important, as without it, the economy is simply a white noise process. More generally, structural macroeconomic models imply a set of zero restrictions on $\mathbf{\Gamma}$ and $\mathbf{\Phi}$. However, different models imply different restrictions and there is little a priori reason to favor one approach over another.

A different branch of literature started by Sims (1980) identifies economic shocks and their responses with vector autoregressions (VARs), which are reduced-form models. Under FI, the regression in proposition 1.2.1.b is a VAR estimated with realizations $\mathbf{y}_t$. Proposition 1.2.1.b shows that this VAR identifies composite parameter $\mathbf{A}$, as well as residuals, i.e. $E[\mathbf{y}_t|\mathcal{F}_t] - E[\mathbf{y}_t|\mathcal{F}_{t-1}] = \mathbf{\Pi}\mathbf{u}_t$. If these residuals can be linked to economic shocks with restrictions on the impact matrix $\mathbf{\Pi}$, the VAR can produce impulse responses thereof. The property that these residuals are not contaminated by past shocks so that they are only function of current shocks makes the VAR invertible or fundamental (see Watson, 1986; Fernández-Villaverde et al., 2007; Leeper, Walker and Yang, 2013; Sims, 2012).

Olivier J. Blanchard, Jean-Paul L'Huillier and Guido Lorenzoni (2013) discuss VARs when relaxing full information FI.ii. To illustrate their point, consider the VAR estimated with real-time data, $E[\mathbf{y}_t|\mathcal{F}_t]$, across $t$. This VAR identifies composite parameter $\mathbf{A}$, as well as residuals $E[\mathbf{y}_t|\mathcal{F}_t] - E[\mathbf{y}_t|\mathcal{F}_{t-1}]$ according to proposition 1.2.1. Note that these residuals are uncorrelated to past information, $\mathcal{F}_{t-1}$, by law of iterated expectations, which means the VAR satisfies the invertibility condition. Blanchard, L'Huillier and Lorenzoni (2013) however show that these residuals can only be linked to economic shocks under FI, where VAR residuals are equal to the residuals of the model, $E[\mathbf{y}_t|\mathcal{F}_t] - E[\mathbf{y}_t|\mathcal{F}_{t-1}] = \mathbf{\Pi u}_t$. Once FI is relaxed, true economic shocks can no longer be separated from noise. Hence, when relaxing FI, identification of composite parameter $\mathbf{A}$ and residuals $E[\mathbf{y}_t|\mathcal{F}_t] - E[\mathbf{y}_t|\mathcal{F}_{t-1}]$ might not provide economic meaning per se.

Estimation of regression in proposition 1.2.1.b does not require data on expectations at multiple horizons, i.e. $E[\mathbf{y}_{t+h}|\mathcal{F}_t]$, for several $h$. Even when data on expectations is available at multiple horizons, it is unclear how to incorporate this data in FIRE models. In particular, proposition 1.2.1.a shows that a researcher could still only identify a combination of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$.

Full information is inconsistent with data on expectations. In particular, data rejects FI.ii as expectations are not equal to realized values, $E[\mathbf{y}_{t-k}|\mathcal{F}_t] \neq \mathbf{y}_{t-k}$, for some $t$ and $k \geq 0$. Moreover, FI.i does not hold because forecasts are noncollinear: there is no matrix that satisfies lemma 1, $E[\mathbf{y}_{t+h}|\mathcal{F}_t] = \mathbf{A}E[\mathbf{y}_{t+h-1}|\mathcal{F}_t]$, for all $t$ and $h \geq 1$. FI.ii is relaxed in models with information frictions such as sticky and noisy information models (see Mankiw and Reis, 2002; Woodford, 2003; Sims, 2003). FI.i is relaxed in models with news shocks (see Beaudry and Portier, 2004, 2006). This paper uses the variation generated by relaxing FI to match data on expectations at all horizons, and to separately identify $\mathbf{\Gamma}$ and $\mathbf{\Phi}$.

## 1.3 A Simple Example

Before I formally introduce the model with flexible information structure, I use a simple example to illustrate the new framework. Consider the following univariate version of model (1.1) governing inflation:

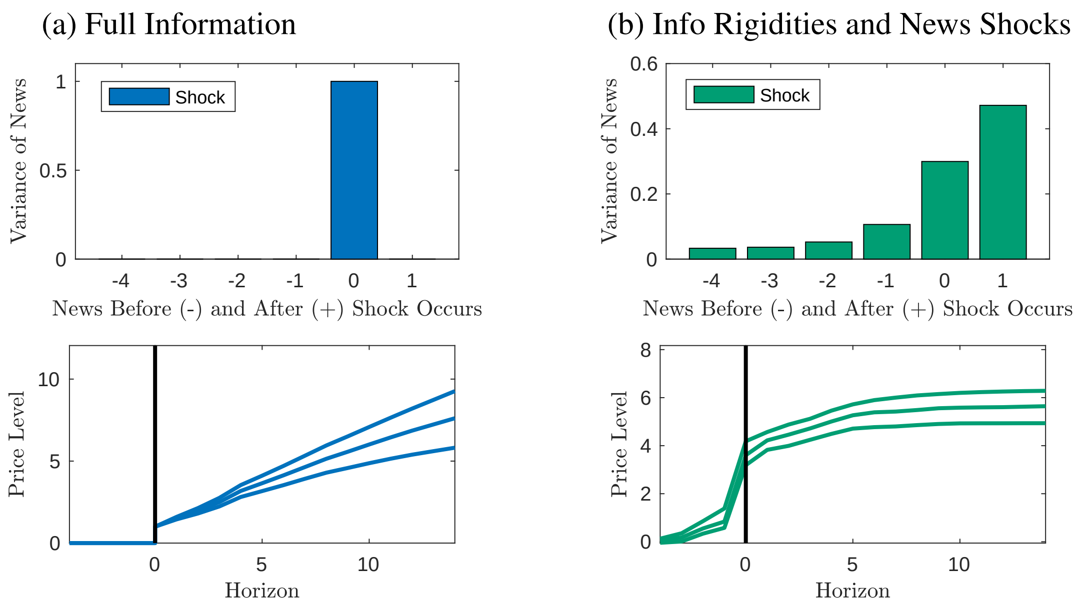
$$\pi_t = \gamma\pi_{t-1} + \phi E[\pi_{t+1}|\mathcal{F}_t] + \sigma u_t, \tag{1.3}$$

with $iid$ shocks $u_t \sim N(0,1)$, a forward-looking and a backward-looking component, $\phi$ and $\gamma$ respectively. Model 1.3 represents the solution to a central bank's problem (see appendix 1.9.1).

Under FI, expected future inflation fully reflects current inflation, $E[\pi_{t+1}|\mathcal{F}_t] = a\pi_t$, where $a$ is the solution to the fixed point in $a = (1-\phi a)^{-1}\gamma$, according to lemma 1. By proposition 1.2.1, additional restrictions on parameters are necessary to identify the model. For example, Coibion and Gorodnichenko (2015) describe inflation with a backward-looking AR(1) process and thus implicitly assume $\phi = 0$ so that $\gamma = a$ is identified by assumption. In this environment, expectations have no impact on inflation, which means that whether agents are well informed or not does not affect the inflation rate.

Note that FI imposes strong model restrictions that can be rejected when data on expectations is observed. The Federal Reserve Bank of Philadelphia provides data on Fed's expectations about past, current, and future inflation. This Greenbook data shows that about 14% of inflation's variance is not observed in real time, which rejects the full information assumption FI.ii (see figure 1.2 of section 1.6). Moreover, the forecasts of inflation are noncollinear, which rejects FI.i according to lemma 1. Motivated by these observations, I relax the full information assumption and allow for information to diffuse slowly over time. Specifically, the agent obtains news about future shocks ahead of time, while some information about the shock is delayed. The top left plot of figure 1.1 illustrates how information is concentrated under FI, and the top right plot shows an example of how information about the shock diffuses slowly starting before realization ($t-4$) up until one period after realization ($t+1$).

Figure 1.1: Information Diffusion in RE Model




*Notes:* This figure compares (a) the full information rational expectations model (1.3) with (b) the same model when FI is relaxed. The top panel displays the share of variance of the shock that is learned on average $h$ quarters after realization, respectively, -$h$ quarters before realization, where $h$ is plotted on the x-axis (see (1.28) for a formal definition of the metric). The lower panel displays the average impulse response to a shock. (a) The FI impulse response is the estimated response of an AR(5) process using real-time data. (b) The impulse response of the model without FI is estimated based on a backcast, a nowcast, and five forecasts, where the RE model in (1.3) is extended to include five lag and four lead terms. 68% confidence bands are estimated using bootstrap method. The model is estimated with quarterly Greenbook data, 1967Q2 - 2011Q4, and price level is defined as GDP price deflator. Units of the bottom plots are cumulated annualized percentage points.

Consider the regression of proposition 1.2.1.b when relaxing FI.i:

$$
\begin{aligned}
E[\pi_t|\mathcal{F}_t] &= \gamma E[\pi_{t-1}|\mathcal{F}_t] + \phi E[\pi_{t+1}|\mathcal{F}_t] + \sigma E[u_t|\mathcal{F}_t], \\
&= aE[\pi_{t-1}|\mathcal{F}_t] + \sum_{s=0}^{\infty}(b_1)^s b_0 \sigma E[u_{t+s}|\mathcal{F}_t], && (1.4) \\
&= aE[\pi_{t-1}|\mathcal{F}_{t-1}] + a(E[\pi_{t-1}|\mathcal{F}_t] - E[\pi_{t-1}|\mathcal{F}_{t-1}]) + \sum_{s=0}^{\infty}(b_1)^s b_0 \sigma E[u_{t+s}|\mathcal{F}_t], && (1.5)
\end{aligned}
$$

where $a$ is the fixed point in $a = (1-\phi a)^{-1}\gamma$, $b_1 = (1-\phi a)^{-1}\phi$, and $b_0 = (1-\phi a)^{-1}$. Regression of $E[\pi_t|\mathcal{F}_t]$ on $E[\pi_{t-1}|\mathcal{F}_{t-1}]$ is biased, as for example yesterday's beliefs about tomorrow's shock, $E[u_{t+1}|\mathcal{F}_{t-1}]$, shift the left- and right-hand side variable simultaneously: $E[\pi_{t-1}|\mathcal{F}_{t-1}]$ shifts by $(b_1)^2 b_0 \sigma$, and $E[\pi_t|\mathcal{F}_t]$ shifts by $b_1 b_0 \sigma$. This inability to identify reduced form parameters under foresight is known as non-invertibility problem (see Fernández-Villaverde et al., 2007; Leeper, Walker and Yang, 2013; Sims, 2012).

Data on expectations at multiple horizons solve this identification issue. The idea is that inflation does not *directly* depend on the future in the RE model, it only depends *indirectly* on the future through current expectations. Hence, any variation that is unobserved cannot depend on the future, because expectations are formed based on observed variation, only, i.e. variation in $\mathcal{F}_t$. Variation that is unobserved today, but it is observed tomorrow is captured in tomorrow's *backcast revision*, which is the change in expectations about the previous period. Again, this backcast revision cannot depend on the future, because the variation that causes a backcast revision is unobserved when it is realized. Formally, the forward-looking term drops out when calculating the backcast revision:

$$
\begin{aligned}
&E[\pi_{t-1}|\mathcal{F}_t] - E[\pi_{t-1}|\mathcal{F}_{t-1}] \\
&\quad = \gamma(E[\pi_{t-2}|\mathcal{F}_t] - E[\pi_{t-2}|\mathcal{F}_{t-1}]) + \phi\left(E[E[\pi_t|\mathcal{F}_{t-1}]|\mathcal{F}_t] - E[E[\pi_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-1}]\right) \\
&\qquad + \sigma\left(E[u_{t-1}|\mathcal{F}_t] - E[u_{t-1}|\mathcal{F}_{t-1}]\right), \\
&\quad = \gamma(E[\pi_{t-2}|\mathcal{F}_t] - E[\pi_{t-2}|\mathcal{F}_{t-1}]) + \sigma\left(E[u_{t-1}|\mathcal{F}_t] - E[u_{t-1}|\mathcal{F}_{t-1}]\right),
\end{aligned} \tag{1.6}
$$

as by law of iterated expectations, $E[E[\pi_t|\mathcal{F}_{t-1}]|\mathcal{F}_t] = E[\pi_t|\mathcal{F}_{t-1}]$. Note how the backcast revision in 1.6 only depends on beliefs about past shocks, but not on beliefs about current and future shocks, which are responsible for the omitted variable bias (OVB) in estimating $a$. In particular, equation (1.5) together with (1.6) show that regression of the nowcast $E[\pi_t|\mathcal{F}_t]$ on backcast revision $(E[\pi_{t-1}|\mathcal{F}_t] - E[\pi_{t-1}|\mathcal{F}_{t-1}])$ does not suffer from OVB, as the left- and right-hand side variables depend on different unobservables. This regression requires expectations at two horizons, expectations about today and expectations about the previous period. This is different from the regression of proposition 1.2.1 of today's nowcast $E[\pi_t|\mathcal{F}_t]$ on yesterday's nowcast $E[\pi_{t-1}|\mathcal{F}_{t-1}]$, where only one horizon is necessary.

Is the above regression unbiased? There is one more catch, while past shocks are uncorrelated with future shocks, it is still possible that *beliefs* about shocks are correlated across time, i.e. $Corr(E[u_{t+k}|\mathcal{F}_t], E[u_{t+h}|\mathcal{F}_t]) \neq Corr(u_{t+k}, u_{t+h}) = 0$, for some $k \neq h$. This is not an issue under FI, where expectations are either equal to zero or equal to the true shock so that correlations of beliefs are always equal to zero, anyways. Intuitively, beliefs about uncorrelated shocks are correlated when there is uncertainty about timing. For example, the central bank receives information about a future tax cut, but it doesn't know whether it will occur in one or in two years. The belief of experiencing a tax cut is thus positively correlated between one and two years from now, even though tax cuts are uncorrelated over time. Identification of feature $a$ needs the following restriction: there is no more uncertainty about timing, once the shock is realized. Hence, once the tax cut occurs, the central bank is no longer uncertain whether it occurs today or another time, while there might still be uncertainty about the size of the tax cut. This restriction on uncertainty about timing is denoted as *revealed timing at realization* (RTR) assumption: $Cov(E[u_t|\mathfrak{F}_{t+k}], E[u_{t+h}|\mathfrak{F}_{t+k}]) = Cov(u_t, u_{t+h}) = 0, \forall h \geq 1, \forall k \geq 0$.

The above exercise shows how we can still identify the composite parameter $a$, despite having relaxed FI.i and FI.ii. It turns out, data on expectations can do even more than that. The reason is that the model is no longer subject to Manski's (1993) reflection problem, as relaxing FI.i provides additional variation in expectations that breaks the reflection. Data on expectations about the future can make use of this variation to separately identify $\phi$ and $\gamma$. In particular, shift equation (1.4) forward:

$$E[\pi_{t+1}|\mathfrak{F}_t] = aE[\pi_t|\mathfrak{F}_t] + \sum_{s=0}^{\infty}(b_1)^s b_0 \sigma E[u_{t+1+s}|\mathfrak{F}_t], \tag{1.7}$$

and use identified $a$ to extract the remainder, i.e. $E[\pi_{t+1}|\mathfrak{F}_t] - aE[\pi_t|\mathfrak{F}_t]$, which is a weighted sum of all future shocks. Moreover, extract the remainder of the nowcast in (1.4), as well, i.e. $E[\pi_t|\mathfrak{F}_t] - aE[\pi_{t-1}|\mathfrak{F}_t]$, and then regress it on the remainder of (1.7) to identify $b_1$. This is apparent in equation (1.4), after taking the current shock out of the sum:

$$\begin{aligned} E[\pi_t|\mathfrak{F}_t] - aE[\pi_{t-1}|\mathfrak{F}_t] &= \sum_{s=0}^{\infty}(b_1)^s b_0 \sigma E[u_{t+s}|\mathfrak{F}_t], \\ &= b_1 \left( \sum_{s=0}^{\infty}(b_1)^s b_0 \sigma E[u_{t+1+s}|\mathfrak{F}_t] \right) + b_0 \sigma E[u_t|\mathfrak{F}_t], \\ &= b_1 \left( E[\pi_{t+1}|\mathfrak{F}_t] - aE[\pi_t|\mathfrak{F}_t] \right) . + b_0 \sigma E[u_t|\mathfrak{F}_t]. \end{aligned} \tag{1.8}$$

This regression gives an unbiased estimate of $b_1$ under RTR. RTR is required because the weighted sum of beliefs of future shocks should not be correlated with the belief of the current shock, i.e. $E[u_t|\mathcal{F}_t]$, as it is the error term in regression (1.8). Remember that composite parameters are invertible functions of the model parameters, $a = (1-\phi a)^{-1}\gamma$, and $b_1 = (1-\phi a)^{-1}\phi$, hence, once $a$ and $b_1$ are identified, $\phi$ and $\gamma$ are identified, as well as $b_0 = (1-\phi a)^{-1}$.

To summarize so far, if data on backcasts, nowcasts, and forecasts are available, the parameters $\gamma$ and $\phi$ of the rational expectations model (1.3) are identified. Identification is possible without a parametric structure for how agents gather information. This flexible information structure allows to match data on expectations at different horizons, without relying on measurement errors or deviations from rationality. The necessary rank conditions are that backcast revisions, nowcasts, and forecasts are noncollinear. These conditions are testable and are satisfied for inflation expectations. The necessary zero conditional mean (ZCM) conditions are that beliefs about realized shocks are uncorrelated with beliefs about future and past shocks (RTR), while beliefs across unrealized shocks can still be correlated.

Under FI, parameters $\gamma$ and $\phi$ of model (1.3) and the underlying process $\pi_t$ are sufficient to extract the shocks, i.e. $\sigma u_t = \frac{1}{b_0}(\pi_t - a\pi_{t-1})$, to identify the persistence of the process, i.e. $\rho(\pi_t, \pi_{t-1}) = a$, and to identify predictions of the agents at all horizons, i.e. $E[\pi_{t+h}|\mathcal{F}_t] = a^h\pi_t$, $\forall h > 0$, and $E[\pi_{t-k}|\mathcal{F}_t] = \pi_{t-k}$, $\forall k \geq 0$. Without FI, identified parameters $\gamma$ and $\phi$ quantify economic relationships, but they say nothing about realized shocks, $u_t$, nothing about how inflation propagates over time, and they do not reveal predictions of the agents beyond the expectations used for estimation. To identify these characteristics, the researcher needs to identify how much information is gathered by the agents before, during, and after the shocks are realized. One extreme is that agents receive most information about the shocks far ahead of time so that agents respond to this information in advance through $\phi$, generating a high persistence in inflation. Another extreme is when most information is gathered after realization so that once the shocks are observed, they are no longer relevant for today's economy so that expectations barely matter, despite a large $\phi$.

I introduce a non-parametric way to identify the information structure, in order to capture the characteristics described above. For this I need to assume a finite *information diffusion interval* (IDI): all the information about a shock diffuses within the interval of $H$ periods before until $K$ periods after the shock is realized. This means that shocks are fully revealed after $K$ periods, i.e. $E[u_t|\mathcal{F}_{t+K}] = u_t$, and there is no information available $H+1$ periods before impact, i.e. $E[u_t|\mathcal{F}_{t-H-1}] = E[u_t] = 0$. The shocks can then be decomposed into $H+1+K$ revisions:

$$u_t = E[u_t|\mathcal{F}_{t+K}] = \sum_{k=-H}^{K} \Big( E[u_t|\mathcal{F}_{t+k}] - E[u_t|\mathcal{F}_{t+k-1}] \Big). \tag{1.9}$$

These revisions are uncorrelated by law of iterated expectations. The variances of the different shock components therefore add up to one:

$$Var(u_t) = \sum_{k=-H}^{K} Var\Big( E[u_t|\mathcal{F}_{t+k}] - E[u_t|\mathcal{F}_{t+k-1}] \Big),$$

where $Var(u_t) = 1$. The variance of a shock component, i.e. $Var\left(E[u_t|\mathcal{F}_{t+k}] - E[u_t|\mathcal{F}_{t+k-1}]\right)$, thus reflects how much of total variance the agent learns $k$ periods after the shock is realized, respectively, $-k$ periods before it is realized. The top right plot of figure 1.1 shows this measure for shocks to inflation expectations under the assumption of $K = 1$ and $H = 4$. About half of the shock's variance is learned before or when it is realized, and the remaining half after. Identification of this result will be discussed shortly. The top left panel shows that under FI, all the variance is learned on impact.

The variances of the components don't fully capture the information structure, yet. Remember there is uncertainty about timing so that beliefs about different shocks can be correlated. This is captured in the covariances, which add up to zero for $h > 0$:

$$Cov(u_t, u_{t+h}) = \sum_{k=-H}^{K+h} Cov\Big(E[u_t|\mathcal{F}_{t+k}] - E[u_t|\mathcal{F}_{t+k-1}], E[u_{t+h}|\mathcal{F}_{t+k}] - E[u_{t+h}|\mathcal{F}_{t+k-1}]\Big),$$

where $Cov(u_t, u_{t+h}) = 0$. RTR states that the expression is also equal to zero when taking the sum only from $k = -H$ to $k = 0$. Again, let's consider the example where the agent receives information about a future tax cut but it is unclear whether it occurs in one or in two years. The agent revises her expectations of both shocks in the same direction so that the revisions about $t+1$ and $t+2$ shocks are positively correlated. Let's say the tax cut occurs in one year at $t+1$. When the tax cut occurs, the agent observes that a tax cut is realized and thus reinforces her previous prediction. Moreover, she revises her incorrect prediction of a $t+2$ tax cut, as it turns out the previous information did not refer to that period, after all. On average, the agent readjusts her beliefs of $t+1$ and $t+2$ tax cuts to an extent where they become uncorrelated, because the agent knows that there is no correlation over time. Hence, the covariances of time $t$ and time $t+1$ revisions add up to zero.

The variance-covariance matrix of the shock components describes how information arrives, which together with the model parameters provide sufficient information to characterize the first and second moments of the data, as well as predictions of the agents. Identification of the different shock components requires data on expectations for the horizons for which agents receive information about shocks. Hence, the IDI assumptions on $H$ and $K$ dictate the necessary forecast horizons for identification. Formally, the shock components can be collected with backcast, nowcast, and forecast revisions according to model (1.3):

$$
\begin{aligned}
&E[\pi_{t-k}|\mathfrak{F}_t] - E[\pi_{t-1}|\mathfrak{F}_{t-1}] \\
&\qquad = \gamma(E[\pi_{t-k-1}|\mathfrak{F}_t] - E[\pi_{t-k-1}|\mathfrak{F}_{t-1}]) + \sigma\left(E[u_{t-k}|\mathfrak{F}_t] - E[u_{t-k}|\mathfrak{F}_{t-1}]\right),\ k \geq 1, \qquad (1.10)\\
&E[\pi_{t+h}|\mathfrak{F}_t] - E[\pi_{t+h}|\mathfrak{F}_{t-1}] \\
&\qquad = \gamma(E[\pi_{t+h-1}|\mathfrak{F}_t] - E[\pi_{t+h-1}|\mathfrak{F}_{t-1}]) + \phi(E[\pi_{t+h+1}|\mathfrak{F}_t] - E[\pi_{t+h+1}|\mathfrak{F}_{t-1}]) \\
&\qquad\quad + \sigma\left(E[u_{t+h}|\mathfrak{F}_t] - E[u_{t+h}|\mathfrak{F}_{t-1}]\right),\ h \geq 0. \qquad (1.11)
\end{aligned}
$$

After extracting the shock revisions, they can be summed up according to (1.9) which identifies the shock $\sigma u_t$, where $\sigma$ is the standard deviation of that expression. The variance-covariance matrix of the shock components then identifies the information structure so that the model is identified.

Let's assume that information about a shock is gathered four periods ahead of time, up to one period after realization, i.e. $H = 4$ and $K = 1$. I include additional lag and lead terms, as well as a trend in model (1.3) and estimate the model using quarterly Greenbook data. For identification, I need $K = 1$ backcast, as well as a nowcast and $H + 1 = 5$ forecasts of inflation. The top right panel of figure 1.1 plots the variances of the components, which are informative about how much the agent learns on average before and after the shock is realized. Note that the agent learns more and more about the variance with time which is not an assumption: it is possible for example that the agent learns more at $t-1$ than at $t+1$. The low amount of learning four quarters ahead of time supports the IDI assumption that there is no learning about shocks far ahead of time. Strikingly, the high variance of the revision one quarter after realization shows that a significant share of shocks to inflation are only observed in retrospect. Remember that this is assumed to be zero in standard FIRE models. Hence, the top right panel of figure 1.1 provides economically significant evidence that the assumption of full information does not hold for modelling inflation;

a finding that is consistent with the empirical literature on information frictions (see e.g. Coibion and Gorodnichenko, 2015).

In an environment with flexible information structure, impulse responses to shocks change over time, as they depend on what information the agent receives about each shock. The lower right panel of figure 1.1 therefore plots the *average* impulse response. Interestingly, on average, inflation responds to a shock even before it is realized. Since the future can only affect today's economy through expectations, this means that expectations about the future matter. This is exactly the story of the news shock literature, where today's economy is affected by a future TFP shock, because expectations thereof change current behaviour (see Beaudry and Portier, 2004, 2006).

This section shows how data on backcasts, nowcasts, and forecasts fully identify a univariate rational expectations model, when full information is relaxed. The findings of this section might be a result of having a highly stylized model, which is why the next section extends the analysis to a multivariate rational expectations model.

## 1.4 RE Model with Generalized Information Structure

This section identifies model (1.1) and its information structure. The foundation for identification is data on expectations at multiple horizons. Building on the literature on information frictions (see Mankiw and Reis, 2002; Woodford, 2003; Sims, 2003) and news shocks (see Beaudry and Portier, 2004, 2006), I relax the full information (FI) assumption to an extent where the model can replicate data on expectations without relying on measurement errors or deviations from rational expectations (RE). Specifically, relaxing FI allows to generate all the moments of the data on expectations other than the ones restricted by the law of iterated expectations, a law that holds under RE. This generalized information structure permits rich variation in expectations that is used to separately identify the forward-looking and backward-looking components of the model.

Section 1.4.1 relaxes FI, by allowing for information to be spread out across time so that information arrives before, at, and after shocks are realized. In line with the news shocks literature, and the literature on sticky and noisy information, I model information exogenously, as opposed to models with rational inattention, where information is chosen endogenously. Instead of introducing a parametric model for the information process, I characterize arrival of information with the moments of the data. In particular, I use the moments of conditional expectations of the shocks to describe the information structure.

Section 1.4.2 solves model (1.1), without having a parametric expression for diffusion of information. Solving models without FI is challenging, as when expectations matter, the way shocks propagate depends on the available information. The solution thus needs to keep track of the state of the information set, $\mathcal{F}_t$, which increases the number of state variables significantly. However, since information is exogenous, I can solve the model without keeping track of information. The idea is that the rational expectations model only depends on two different information sets, the information set when the economy is realized, $\mathcal{F}_t$, through $E[\mathbf{y}_{t+1}|\mathcal{F}_t]$, and the information set when everything is revealed, through $\mathbf{y}_{t-1}$ and $\mathbf{u}_t$. I thus split up the solution into two parts, the part that is observed at realization, $E[\mathbf{y}_t|\mathcal{F}_t]$, and the part that is observed after realization, $\mathbf{y}_t - E[\mathbf{y}_t|\mathcal{F}_t]$. These two parts can then be solved separately, where solution can be found by conditioning on the same information. As information is fix for each part, solving the two parts is similar to solving perfect foresight models. Solution is then expressed in terms of conditional expectations of shocks, which are exogenously determined by the information structure according to section 1.4.1.

Section 1.4.3 identifies model parameters $\mathbf{\Gamma}$ and $\mathbf{\Phi}$, the shocks of the model and the information structure described in section 1.4.1. Identification is possible as the model is no longer subject to Manski's (1993) reflection problem, which occurs when the econometrician cannot separate the direct effect of a shock from the indirect effect of observing that shock (see section 1.2). The variation coming from information about the future breaks Manski's (1993) reflection problem, as expectations no longer solely reflect changes in the current economy. New information about the past identifies the backward-looking component, which then helps to extract the variation that solves Manski's (1993) reflection problem. The necessary assumption for identification is that there is no uncertainty about when a shock hits the economy, once it is realized. Identification of shocks and information structure requires the assumptions that information about shocks diffuses within a finite interval. Information structure is characterized by the variance-covariance matrix of changes in expectations about past, current, and future shocks. Identification of shocks also requires identifying restrictions on the impact matrix of the shocks, i.e. $\mathbf{\Pi}$, which are discussed in detail in section 1.5.

Changing information structure might alter the form of the rational expectations model (1.1), depending on how it is micro-founded. Two generalizations are considered. First, if there is missing information about the past, beliefs about the past, $E[\mathbf{y}_{t-1}|\mathcal{F}_t]$, might have different effects than the actual past, $\mathbf{y}_{t-1}$, a case discussed in appendix 1.10.1. Second, if there is information about future shocks, expectations of those shocks, $E[\mathbf{u}_{t+h}|\mathcal{F}_t]$, might enter the system separately, a case discussed in appendix 1.10.2. The findings of the paper are robust to both generalizations. The model estimated in section 1.6.2 extends the number of lag and lead terms of model (1.1), and introduces a trend.

### 1.4.1 Information Structure

Consider a standard linear rational expectations model as in (1.1) under rational expectations (RE), but without imposing full information (FI). In this generalized framework, a representative agent forms expectations about current, future, and past variables $\mathbf{y}_{t+s}$, as well as shocks $\mathbf{u}_{t+s}$, $\forall s$. The information structure is expressed in terms of shocks only, $\mathbf{u}_{t+s}$, without loss of generality. The assumption of full information (FI) is abandoned among the following dimensions:

**Information Diffusion Interval (IDI).** *Information about shocks arrive $H$ periods before until $K$ periods after realization:*

$$E[\mathbf{u}_t|\mathcal{F}_{t+s}] = \begin{cases} E[\mathbf{u}_t], & s < -H, \qquad (i) \\ \mathbf{u}_t, & s \geq K. \qquad (ii) \end{cases}$$

IDI states that new information may arrive before, during, and after shocks are realized, but information is only gathered within $H + K + 1$ periods. FI is a degenerated case of IDI, where all the information arrives on impact. Specifically, FI.i implies $H = 0$, and FI.ii implies $K = 0$. Identification of forward- and backward-looking parameters $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ does not require $K$ and $H$ to be finite. Finite $K$ and $H$ are necessary for identification of shocks and the underlying information structure.

**Incomplete Information and News about the Future (IIN).** *Agents update beliefs about shocks before, at, and after realization: $Var(E[\mathbf{u}_t|\mathcal{F}_{t+s}] - E[\mathbf{u}_t|\mathcal{F}_{t+s-1}]) > 0$, for $s = 0$, (i) for some $s < 0$, and (ii) for some $s > 0$.*

IIN is a rank condition necessary for identification of parameters. IIN violates FI, where agents only update on impact, i.e. $Var(E[\mathbf{u}_t|\mathcal{F}_t] - E[\mathbf{u}_t|\mathcal{F}_{t-1}]) = Var(\mathbf{u}_t)$, and $Var(E[\mathbf{u}_t|\mathcal{F}_{t+s}] - E[\mathbf{u}_t|\mathcal{F}_{t+s-1}]) = 0$, for $s \neq 0$. IIN is testable if data on expectations is available at multiple horizons. In particular, agents update their beliefs about past shocks if expectations are not equal to realized values, $E[\mathbf{y}_{t-k}|\mathcal{F}_t] \neq \mathbf{y}_{t-k}$, for some $t$ and $k \geq 0$. Agents update their beliefs about future shocks if there is no matrix that satisfies lemma 1, $E[\mathbf{y}_{t+h}|\mathcal{F}_t] = \mathbf{A}E[\mathbf{y}_{t+h-1}|\mathcal{F}_t]$, for all $t$ and $h \geq 1$, respectively, if now- and forecasts are noncollinear. IIN.i is satisfied in models with news shocks (see Beaudry and Portier, 2004, 2006), while IIN.ii is satisfied in models with information frictions (see Mankiw and Reis, 2002; Woodford, 2003; Sims, 2003).

As discussed in section 1.3, relaxing FI may cause *beliefs* about shocks to be correlated across time, because of *uncertainty about timing*. The following zero conditional mean (ZCM) assumption restricts this uncertainty for realized shocks.

**Revealed Timing at Realization (RTR).** *There is no uncertainty about timing after shocks are realized:* $Cov(E[\mathbf{u}_t|\mathcal{F}_{t+k}], E[\mathbf{u}_{t+h}|\mathcal{F}_{t+k}]) = Cov(\mathbf{u}_t, \mathbf{u}_{t+h}) = 0$, $\forall h \geq 1$, $\forall k \geq 0$.

RTR states that uncertainty about timing, $Cov(E[\mathbf{u}_t|\mathcal{F}_{t+k}], E[\mathbf{u}_{t+h}|\mathcal{F}_{t+k}]) \neq 0$, is still allowed among future shocks, but it is no longer allowed for current and past shocks. This means that uncertainty about *when* the shock occurs disappears the moment the shock is realized. Note that there might still be uncertainty about sign, magnitude, and type of shock. In other words, current information on $\mathbf{u}_t$ is no longer correlated with information that is used to predict future and past shocks, $\mathbf{u}_{t+h}$, $h \neq 0$, but information on $\mathbf{u}_t$ might still be incomplete. For example, a tax cut occurs a year from now, but people don't know whether it occurs in one or in two years. Once the tax cut is realized, there is no more uncertainty on whether tax rate changes that same year or the following year. There is however still uncertainty by how much the tax rate changes, and it might even be unclear whether the shock is indeed a tax cut.

RTR is the ZCM assumption necessary to identify parameters $\mathbf{\Gamma}$ and $\mathbf{\Phi}$. RTR is satisfied in models with news shocks as in Beaudry and Portier (2004, 2006); Davis (2007); Christiano et al. (2010), as these models don't allow for uncertainty about timing, at all. In this literature, changes in expectations of future shocks are modelled as independent to changes in expectations about other shocks.

Relaxing FI together with IDI, IIN and RTR produce the following five properties:

(1) *News shocks about the future:* $E[\mathbf{u}_t|\mathcal{F}_{t-h}] \neq E[\mathbf{u}_t]$*, for some* $t$ *and* $1 \leq h \leq H$. This is the counterpart of FI.i and follows from IIN.i, and IDI.i restricts $H$. News shocks are shocks that are partially or fully observed before they are realized. This means the agent receives information not just about today but also about the future. Information about the future is available, for example, when the government announces future military spending in advance.

(2) *Incomplete information:* $E[\mathbf{u}_t|\mathcal{F}_{t+k}] \neq \mathbf{u}_t$*, for some* $t$ *and* $0 \leq k < K$. This is the counterpart of FI.ii and follows from IIN.ii, and IDI.ii restricts $H$. Without FI.ii, information might be missing about shocks, even after they are realized. Information is imperfect, for example, when the government lacks some information about current GDP, because it does not collect taxes and therefore does not have the necessary information until after the calender year ends.

Property (1) is standard in the news shock literature, while property (2) is standard in the literature on information frictions. The next three properties are non-standard, and describe how *beliefs* about shocks can be correlated, even though realized shocks are uncorrelated. Note that properties (3), (4), and (5) describe possible relationships, but they are not assumptions necessary for identification.

(3) *Uncertainty about timing among future shocks:* $Cov(E[\mathbf{u}_{t+k}|\mathcal{F}_t], E[\mathbf{u}_{t+h}|\mathcal{F}_t]) \neq Cov(\mathbf{u}_{t+k}, \mathbf{u}_{t+h}) = 0$*, is possible for* $h > k \geq 1$. Note that under FI, this property does not exist, because $Cov(E[\mathbf{u}_{t+k}|\mathcal{F}_t], E[\mathbf{u}_{t+h}|\mathcal{F}_t])$ can either be equal to $Cov(\mathbf{u}_{t+k}, \mathbf{u}_{t+h})$, $Cov(0, \mathbf{u}_{t+h})$, $Cov(\mathbf{u}_{t+k}, 0)$, or $Cov(0, 0)$, which are all equal to zero. Timing is uncertain, for example, when information about a future tax cut is available, but it is unclear whether the tax cut occurs in one or in two years. The belief that taxes will be reduced in one year is thus correlated with the belief that taxes will be cut in two years, even though tax reductions themselves are uncorrelated over time. This property is an explanation why expected variables have different persistence than realized variables. For example, Monica Jain (2017) and Jane Ryngaert (2017) find evidence that expected inflation has different persistence than actual inflation. While they argue that perceived inflation persistence might deviate due to missperception of forecasters, this could instead be rationalized with uncertain timing.

(4) *Uncertainty about type:* $Cov(E[u_{it}|\mathcal{F}_{t+s}], E[u_{jt}|\mathcal{F}_{t+s}]) \neq Cov(u_{it}, u_{jt}) = 0$*, is possible for* $i \neq j$. Type is uncertain, for example, when there is information about a shock, but it is unclear whether it is a fiscal policy shock or a monetary policy shock. This property is an explanation why expected variables have different correlations among each other than realized variables.

(5) *Uncertainty about both timing and type:* $Cov(E[u_{it+k}|\mathcal{F}_{t+s}], E[u_{jt+h}|\mathcal{F}_{t+s}]) \neq Cov(u_{it+k}, u_{jt+h}) = 0$*, is possible for* $i \neq j$*, and* $h > k \geq 1$. This property occurs when there is both types of uncertainty, for example, when it is unclear whether a signal refers to a monetary policy shock in two years or to a fiscal policy shock in one year.

I don't impose a parametric structure on the information structure. This is desirable because of two reasons. First, it keeps the information structure as general as possible so that the structure nests existing models of the literature. Second, additional assumptions would restrict the relationships of conditional expectations, $E[E[\mathbf{y}_{t+k}|\mathcal{F}_t]E[\mathbf{y}_{t+h}|\mathcal{F}_t]]$ across $k$ and $h$. Deviations in observed relationships from these restrictions could then only be explained by either relaxing the RE assumption, or by introducing mistakes in data collection. Instead, this paper matches the data by adjusting information sets only. Specifically, the estimated model of section 1.6.2 will be exactly identified. In particular, multiple lag and lead terms as well as the generalized information structure replicate all the first and second moments of revisions in expectations at all horizons.

Literature on information frictions often express information processing in terms of variables $\mathbf{y}_t$, instead of the shocks $\mathbf{u}_t$. Properties (1)-(5) can be expressed in variables, as well, by just replacing $\mathbf{u}_t$ with $\mathbf{y}_t$. The advantage of expressing information structure with shocks is that the shocks are uncorrelated and have a unit variance, i.e. $E[\mathbf{u}_{t+k}\mathbf{u}_t] = 0$, and $E[\mathbf{u}_t\mathbf{u}_t'] = I$, for $k \neq 0$, so that deviations in conditional expectations thereof can be simply attributed to the information structure.

The different properties of the information structure can be quantified by decomposing structural shocks $\mathbf{u}_t$ into the following components by IDI:

$$\mathbf{u}_t = \sum_{k=-H}^{K} \Big( E[\mathbf{u}_t|\mathcal{F}_{t+k}] - E[\mathbf{u}_t|\mathcal{F}_{t+k-1}] \Big), \tag{1.12}$$

where each component reflects what the agent learned about shock $\mathbf{u}_t$, $k$ periods after, respectively $-k$ periods before realization. The shock $\mathbf{u}_t$ is fully revealed at time $t + K$ to the agent, and there is no information available before time $t - H$. This decomposition characterizes the agent's cumulative learning about the shock, which ultimately is fully revealed at $t + H$. More specifically, the components can be interpreted as follows, conditional on today's information set:

- *News components*, $E[\mathbf{u}_{t+h}|\mathfrak{F}_t] - E[\mathbf{u}_{t+h}|\mathfrak{F}_{t-1}]$, $h \geq 1$, are changes in expectations about a future shock. These will be non-zero whenever agents receive new information about future values of the shock.
- *Surprise components*, $E[\mathbf{u}_t|\mathfrak{F}_t] - E[\mathbf{u}_t|\mathfrak{F}_{t-1}]$, are changes in expectations about a current shock. This is the only non-zero component in FIRE models, where all the information is learned on impact.
- *Revise components*, $E[\mathbf{u}_{t-k}|\mathfrak{F}_t] - E[\mathbf{u}_{t-k}|\mathfrak{F}_{t-1}]$, $k \geq 1$, are changes in expectations about a past shock. These revisions will occur whenever agents have imperfect information about current or future shocks and only gradually learn about their values over time.

The information structure can then be quantified by the variance-covariance matrix of the shock components:

$$\mathfrak{D} \equiv E\Big[(E[\mathbf{U}_t|\mathfrak{F}_t] - E[\mathbf{U}_t|\mathfrak{F}_{t-1}])(E[\mathbf{U}_t|\mathfrak{F}_t] - E[\mathbf{U}_t|\mathfrak{F}_{t-1}])'\Big], \tag{1.13}$$

where $\mathbf{U}_t \equiv [\mathbf{u}'_{t-K} \ \cdots \ \mathbf{u}'_{t+H}]'$. For illustrative purpose, let $K = 1$, and $H = 2$ so that there is no uncertainty about $t-1$, and no information on shocks beyond $t+2$. The covariance matrix $\mathfrak{D}$ can then be decomposed into $(K+1+H)^2 = 16$ matrices, each with dimension $N \times N$ as follows:

$$\mathfrak{D} = \begin{pmatrix} \mathfrak{D}_{-1,-1} & \mathfrak{D}_{-1,0} & \mathfrak{D}_{-1,1} & \mathfrak{D}_{-1,2} \\ \mathfrak{D}_{-1,0} & \mathfrak{D}_{00} & \mathfrak{D}_{01} & \mathfrak{D}_{02} \\ \mathfrak{D}_{-1,1} & \mathfrak{D}_{01} & \mathfrak{D}_{11} & \mathfrak{D}_{12} \\ \mathfrak{D}_{-1,2} & \mathfrak{D}_{02} & \mathfrak{D}_{12} & \mathfrak{D}_{22} \end{pmatrix}.$$

This variance-covariance matrix captures how information arrives in the economy. For example, under FI, shocks $\mathbf{u}_t$ are fully and only revealed at the time of realization so that $E[\mathbf{u}_t|\mathfrak{F}_t] - E[\mathbf{u}_t|\mathfrak{F}_{t-1}] = I\mathbf{u}_t$, which implies $\mathfrak{D}_{00} = I$, and $\mathfrak{D}_{hk} = 0$ for all other $h$ and $k$. Relaxing FI means other matrices than $\mathfrak{D}_{00}$ might be non-zero. More specifically, the above properties can be quantified as follows:

(1) *Imperfect information:* $\sum_{h=0}^{H} \mathfrak{D}_{hh} \neq I$.

(2) *News shocks about the future:* $\sum_{h=1}^{H} \mathfrak{D}_{hh} \neq 0$.

(3) *Uncertainty about timing among future shocks:* $diag(\mathfrak{D}_{hk}) \neq 0$ *for some* $h > k \geq 0$*, and by RTR,* $\mathfrak{D}_{h,k} = 0$ *for all* $h \leq -1$ *and for all* $k$.

(4) *Uncertainty about type:* $\mathfrak{D}_{hh}$ *is not diagonal for some* $h$.

(5) *Uncertainty about timing and type:* $\mathfrak{D}_{hk} - diag(\mathfrak{D}_{hk}) \neq 0$ *for some* $h > k \geq 0$.

If information has no real effects, $\mathfrak{D}$ can be altered arbitrarily without affecting $\mathbf{y}_t$, as long as the properties of the shocks remain satisfied: $E[\mathbf{u}_t \mathbf{u}'_{t+h}] = 0$, $\forall h \neq 0$, and $E[\mathbf{u}_t \mathbf{u}'_t] = I$, which imply $\sum_{h=-K}^{H} \mathfrak{D}_{h,h+s} = 0$, for all $s \neq 0$, and $\sum_{h=-K}^{H} \mathfrak{D}_{hh} = I$. One extreme is to change $\mathfrak{D}$ so that everything is only observed ex post, by setting all matrices equal to zero except $\mathfrak{D}_{-1-1} = I$. Another extreme is to put the identity at the bottom right so that everything is observed $H$ periods in advance, $\mathfrak{D}_{HH} = I$. If $\mathbf{y}_t$ is different for these two cases, the timing of arrival of information affects economic dynamics, which means expectations must matter for economic decisions. If $\mathbf{y}_t$ remains the same, information is irrelevant meaning that forward-looking behaviour has no real impact. Section 1.6.3 calculates these two counterfactuals, as well as other manipulations of $\mathfrak{D}$.

### 1.4.2 Solution

Model (1.1) depends on variables and shocks conditional on two different information sets, the information set of the period when $\mathbf{y}_t$ is realized, $\mathcal{F}_t$, and the information set when $\mathbf{y}_t$ is fully revealed. The economy thus propagates differently for information that arrives before time $t$, than for information that arrives after time $t$:

$$\mathbf{y}_t = \underbrace{E[\mathbf{y}_t|\mathcal{F}_t]}_{\text{info before } t} + \underbrace{(\mathbf{y}_t - E[\mathbf{y}_t|\mathcal{F}_t])}_{\text{info after } t} \tag{1.14}$$

Let's solve for the part of the model that is based on information that arrives before time $t$. The solution of $E[\mathbf{y}_t|\mathcal{F}_t]$ can be expressed as in Michael Binder and M. Hashem Pesaran (1997):

$$E[\mathbf{y}_t|\mathcal{F}_t] = \mathbf{A}E[\mathbf{y}_{t-1}|\mathcal{F}_t] + E[\mathbf{z}_t|\mathcal{F}_t], \tag{1.15}$$

$$E[\mathbf{z}_t|\mathcal{F}_t] = \mathbf{B}_1 E[\mathbf{z}_{t+1}|\mathcal{F}_t] + \mathbf{B}_0 \mathbf{\Pi} E[\mathbf{u}_t|\mathcal{F}_t]. \tag{1.16}$$

Appendix 1.9.2 derives $\mathbf{A} = (I - \mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}$, $\mathbf{B}_1 = (I - \mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Phi}$, and $\mathbf{B}_0 = (I - \mathbf{\Phi}\mathbf{A})^{-1}$. Hence, once $\mathbf{A}$ and $\mathbf{B}_1$ are identified, it is possible to solve for $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ without parameter restrictions. This solution nests the solution of the previous section where FI.i is imposed. FI.i implies $E[\mathbf{z}_{t+1}|\mathcal{F}_t] = 0$ so that no variation goes through $\mathbf{B}_1$, and therefore only $\mathbf{A}$ shows up.

The second part of the solution, $\mathbf{y}_t - E[\mathbf{y}_t|\mathcal{F}_t]$, is based on information that arrives after time $t$. From the rational expectations model (1.1) it is apparent that fluctuations in $\mathbf{y}_t$ are either caused directly by fundamental changes through $\{\mathbf{y}_{t-1}, \mathbf{u}_t\}$, or by changes in beliefs through $E[\mathbf{y}_{t+1}|\mathcal{F}_t]$. Note that beliefs are fully observed when the period is realized. There are therefore no revisions in $E[\mathbf{y}_{t+1}|\mathcal{F}_t]$ in the future, only in $\{\mathbf{y}_{t-1}, \mathbf{u}_t\}$. Mathematically, this is just the law of iterated expectations, which holds under rational expectations: $E[E[\mathbf{y}_{t+1}|\mathcal{F}_t]|\mathcal{F}_{t+h}] = E[\mathbf{y}_{t+1}|\mathcal{F}_t]$, $\forall h \geq 0$. Future revisions in the representative agent's belief about $\mathbf{y}_t$ can thus only be caused directly by unobserved fluctuations in fundamentals $\{\mathbf{y}_{t-1}, \mathbf{u}_t\}$, which agents learn about only after time $t$, and not by changes in beliefs, since the relevant beliefs which affected economic outcomes at time $t$ are known in subsequent periods. Hence, the part of the model that depends on beliefs about the future cancels out:

$$(\mathbf{y}_t - E[\mathbf{y}_t|\mathcal{F}_t]) = \tilde{\mathbf{\Gamma}} (\mathbf{y}_{t-1} - E[\mathbf{y}_{t-1}|\mathcal{F}_t]) + \mathbf{\Pi} (\mathbf{u}_t - E[\mathbf{u}_t|\mathcal{F}_t]), \tag{1.17}$$

where $\tilde{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$ is possible if beliefs about the past matter, which is further discussed in appendix

1.10.1. When agents revise their beliefs about past outcomes, it can only be because they learned since then something about past fundamentals.

Combining both parts of $\mathbf{y}_t$ according to systems (1.14), (1.15), (1.16), and (1.17):

$$\begin{aligned}\mathbf{y}_t =& \mathbf{A}\mathbf{y}_{t-1} + \sum_{h=1}^{H}(\mathbf{B}_1)^h\mathbf{B}_0\mathbf{\Pi}E[\mathbf{u}_{t+h}|\mathcal{F}_t] + \mathbf{B}_0\mathbf{\Pi}\mathbf{u}_t \\ &+ (\tilde{\mathbf{\Gamma}} - \mathbf{A})\left(\mathbf{y}_{t-1} - E[\mathbf{y}_{t-1}|\mathcal{F}_t]\right) + (I - \mathbf{B}_0)\mathbf{\Pi}\left(\mathbf{u}_t - E[\mathbf{u}_t|\mathcal{F}_t]\right). \end{aligned} \tag{1.18}$$

The first line illustrates how the economy depends on realizations $\{\mathbf{y}_{t-1}, \mathbf{u}_t\}$, but also on expectations about the future $E[\mathbf{u}_{t+h}|\mathcal{F}_t]$, $h \geq 1$. The second line adjusts for information that is gathered after $\mathbf{y}_t$ is realized. The impact of these unobserved fluctuations differ, as they don't experience any feedback from expectations.

This section solved the model without imposing assumptions on the information process. The next section shows how the parameters of the model can be identified by a researcher who has access to data on the expectations of the representative agent.

### 1.4.3 Identification of the Model

This section identifies $\mathbf{A}$ and $\mathbf{B}_1$, which solve $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{B}_0$. The identification strategy of section 1.2 no longer works to get composite parameter $\mathbf{A}$. The regression of $E[\mathbf{y}_t|\mathcal{F}_t]$ on $E[\mathbf{y}_{t-1}|\mathcal{F}_{t-1}]$ is biased, because of the confounding factor $\mathbf{z}_t$: $E[\mathbf{y}_t|\mathcal{F}_t]$ depends on $E[\mathbf{z}_t|\mathcal{F}_t]$ through $\mathbf{B}_0\mathbf{\Pi} \neq 0$, while $E[\mathbf{y}_{t-1}|\mathcal{F}_{t-1}]$ depends on $E[\mathbf{z}_t|\mathcal{F}_{t-1}]$ through $\mathbf{B}_1\mathbf{B}_0\mathbf{\Pi} \neq 0$.

A new identification strategy is introduced that makes use of fluctuation that is only observed in retrospect. Unobserved fluctuations have no impact on expectations and thus do not depend on the confounding factor. This unobserved variation is captured with *backcast revisions*, which are defined as changes in expectations about the past, $E[\mathbf{y}_{t-k}|\mathcal{F}_t] - E[\mathbf{y}_{t-k}|\mathcal{F}_{t-1}]$, $k \geq 1$.

Formally, system (1.15) shows nowcast revisions depend on backcast revisions through $\mathbf{A}$:

$$E[\mathbf{y}_t|\mathcal{F}_t] - E[\mathbf{y}_t|\mathcal{F}_{t-1}] = \mathbf{A}\left(E[\mathbf{y}_{t-1}|\mathcal{F}_t] - E[\mathbf{y}_{t-1}|\mathcal{F}_{t-1}]\right) + \left(E[\mathbf{z}_t|\mathcal{F}_t] - E[\mathbf{z}_t|\mathcal{F}_{t-1}]\right), \tag{1.19}$$

where backcast revisions don't depend on $\mathbf{z}_t$ according to system (1.17). Hence, there is no correlation between backcast revisions and the error term coming from the shocks directly, as they are uncorrelated over time, $E[\mathbf{u}_{t+h}\mathbf{u}_t] = 0$, $h \neq 0$, implies $E[\mathbf{z}_t\mathbf{u}_{t-k}] = 0$, $k \geq 1$. Section 1.4.1 shows however that *beliefs* about the shocks might still be correlated so that identification of $\mathbf{A}$ requires the RTR assumption. The revealed timing at realization (RTR) assumption makes regression (1.19) unbiased, because the error term, $E[\mathbf{z}_t|\mathcal{F}_t] - E[\mathbf{z}_t|\mathcal{F}_{t-1}]$, depends on today's and yesterday's beliefs about current and future shocks, while backcast revisions depend on today's and yesterday's beliefs about all past shocks. They are not correlated because there is no more uncertainty about timing between past shocks versus current and future shocks according to RTR. Formally, inspection of (1.19) shows that the regressor is uncorrelated with the error term under RTR:

$$\begin{aligned}
&E\left[\left(E[\mathbf{y}_{t-1}|\mathcal{F}_t] - E[\mathbf{y}_{t-1}|\mathcal{F}_{t-1}]\right)\left(E[\mathbf{z}_t|\mathcal{F}_t] - E[\mathbf{z}_t|\mathcal{F}_{t-1}]\right)\right] \\
&= E\left[\left(\sum_{s=1}^{\infty}(\tilde{\mathbf{\Gamma}})^s\left(E[\mathbf{u}_{t-s}|\mathcal{F}_t] - E[\mathbf{u}_{t-s}|\mathcal{F}_{t-1}]\right)\right)\left(\sum_{h=0}^{\infty}(\tilde{\mathbf{B}}_1)^h\tilde{\mathbf{B}}_0\tilde{\mathbf{\Pi}}\left(E[\mathbf{u}_{t+h}|\mathcal{F}_t] - E[\mathbf{u}_{t+h}|\mathcal{F}_{t-1}]\right)\right)\right] \\
&= \sum_{s=1}^{\infty}\sum_{h=0}^{\infty}(\tilde{\mathbf{\Gamma}})^s(\tilde{\mathbf{B}}_1)^h\tilde{\mathbf{B}}_0\tilde{\mathbf{\Pi}}\Big(Cov\left(E[\mathbf{u}_{t-s}|\mathcal{F}_t], E[\mathbf{u}_{t+h}|\mathcal{F}_t]\right)\big) - Cov\left(E[\mathbf{u}_{t-s}|\mathcal{F}_{t-1}], E[\mathbf{u}_{t+h}|\mathcal{F}_{t-1}]\right)\big)\Big) \\
&= 0.
\end{aligned}$$

Once $\mathbf{A}$ is known, identification of $\mathbf{B}_1$ can be done in two steps. First, use $\mathbf{A}$ to extract $E[\mathbf{z}_{t+h}|\mathcal{F}_t] = E[\mathbf{y}_{t+h}|\mathcal{F}_t] - \mathbf{A}E[\mathbf{y}_{t+h-1}|\mathcal{F}_t]$ according to system (1.15). Second, regress $E[\mathbf{z}_t|\mathcal{F}_t]$ on $E[\mathbf{z}_{t+1}|\mathcal{F}_t]$ to get $\mathbf{B}_1$ of system (1.16). $\mathbf{A}$ and $\mathbf{B}_1$ are then sufficient to solve for $\mathbf{\Gamma}$ and $\mathbf{\Phi}$, and $\mathbf{B}_0 = (I - \mathbf{\Phi}\mathbf{A})^{-1}$ so that the parameters are identified. Regression (1.16) is unbiased because regressor and error terms are uncorrelated under RTR:

$$\begin{aligned} E\bigg[E[\mathbf{z}_{t+1}|\mathcal{F}_t]E[\mathbf{u}_t|\mathcal{F}_t]\bigg] &= E\bigg[\sum_{h=1}^{\infty}(\tilde{\mathbf{B}}_1)^h\tilde{\mathbf{B}}_0\tilde{\mathbf{\Pi}}E[\mathbf{u}_{t+h}|\mathcal{F}_t]E[\mathbf{u}_t|\mathcal{F}_t]\bigg] \\ &= \sum_{h=1}^{\infty}(\tilde{\mathbf{B}}_1)^h\tilde{\mathbf{B}}_0\tilde{\mathbf{\Pi}}Cov\left(E[\mathbf{u}_{t+h}|\mathcal{F}_t], E[\mathbf{u}_t|\mathcal{F}_t]\right) = 0. \end{aligned}$$

Next, I impose IDI in addition to RTR to identify the shocks of the model as well as the shock components. The variance-covariance matrix of the shock components then identifies the information structure.

IDI implies that structural shocks $\mathbf{u}_t$ can be decomposed into $K+H+1$ news, surprise, and revise components according to system (1.12). Define *residual* components as linear combinations of these components: $\mathbf{\Pi}(E[\mathbf{u}_{t+s}|\mathcal{F}_t] - E[\mathbf{u}_{t+s}|\mathcal{F}_{t-1}])$, guided by the impact matrix $\mathbf{\Pi}$. Systems (1.16) and (1.17) show that revise, surprise, and news *residual* components can be extracted as follows:

$$\begin{aligned} E[\mathbf{y}_{t-k}|\mathcal{F}_t] - E[\mathbf{y}_{t-k}|\mathcal{F}_{t-1}] &= \tilde{\mathbf{\Gamma}}\left(E[\mathbf{y}_{t-k-1}|\mathcal{F}_t] - E[\mathbf{y}_{t-k-1}|\mathcal{F}_{t-1}]\right) \\ &\quad + \mathbf{\Pi}\left(E[\mathbf{u}_{t-k}|\mathcal{F}_t] - E[\mathbf{u}_{t-k}|\mathcal{F}_{t-1}]\right),\ 1 \le k \le K, \qquad (1.20) \\ E[\mathbf{z}_{t+h}|\mathcal{F}_t] - E[\mathbf{z}_{t+h}|\mathcal{F}_{t-1}] &= \mathbf{B}_1(E[\mathbf{z}_{t+h+1}|\mathcal{F}_t] - E[\mathbf{z}_{t+h+1}|\mathcal{F}_{t-1}]) \\ &\quad + \mathbf{B}_0\mathbf{\Pi}(E[\mathbf{u}_{t+h}|\mathcal{F}_t] - E[\mathbf{u}_{t+h}|\mathcal{F}_{t-1}]),\ 0 \le h \le H. \qquad (1.21) \end{aligned}$$

Parameters $\mathbf{B}_1$, $\mathbf{B}_0$, and $\mathbf{\Gamma}$, as well as unobservable $E[\mathbf{z}_{t+h}|\mathcal{F}_t]$ are identified in the previous section. If $\tilde{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$, which is the case when beliefs about the past matter (see appendix 1.10.1), system (1.20) can be used to identify $\tilde{\mathbf{\Gamma}}$ under RTR.

Residuals, $\mathbf{\Pi}\mathbf{u}_t$, are identified by summing up all identified residual components:

$$\mathbf{\Pi}\mathbf{u}_t = \sum_{k=-H}^{K}\mathbf{\Pi}\Big(E[\mathbf{u}_t|\mathcal{F}_{t+k}] - E[\mathbf{u}_t|\mathcal{F}_{t+k-1}]\Big). \qquad (1.22)$$

Since $E[\mathbf{u}_t\mathbf{u}_t'] = I$, the covariance matrix is identified: $E[(\mathbf{\Pi u}_t)(\mathbf{\Pi u}_t)'] = \mathbf{\Pi\Pi}'$.

Identification of shocks $\mathbf{u}_t$ requires restrictions on the impact matrix $\mathbf{\Pi}$. Identified covariance matrix $\mathbf{\Pi\Pi}'$ provides $\frac{(N+1)N}{2}$ out of $N^2$ parameters so that $\frac{(N-1)N}{2}$ additional restrictions are necessary. Constraints can be derived from dynamic stochastic general equilibrium (DSGE) models, directly, or from restrictions on impulse response functions. Section 1.5 discusses the different strategies to get appropriate identifying restrictions. Identification of $\mathbf{\Pi\Pi}'$ is sufficient to quantify the importance of availability of information (see section 1.6.3), while identification of the impact matrix $\mathbf{\Pi}$ is necessary for impulse response analyses (see section 1.6.4).

The information structure is characterized by the $N(H+K+1) \times N(H+K+1)$ variance-covariance matrix $\mathfrak{D}$ according to system (1.13), which is derived using the identified shock components. Remember that there is uncertainty about type so that the shock components, $E[u_{i,t+h}|\mathcal{F}_t] - E[u_{i,t+h}|\mathcal{F}_{t-1}]$, might be correlated across $i$. In order to orthogonalize across $i$, one needs to identify information matrix $\mathbf{D}$, where $\mathbf{DD}' = \mathfrak{D}$. Identifying restrictions on $\mathbf{D}$ are, for example, whether agents confuse demand shocks with supply shocks, the other way around, or if they confuse demand shocks with supply shocks as often as they confuse supply shocks with demand shocks.

To summarize, section 1.4 relaxes FI by introducing a non-parametric description of how information diffuses over time, then it solves the model despite not having a parametric expression for the information process, and finally, section 1.4 uses data on expectations to identify parameters, shocks, and information structure of the rational expectations model.

## 1.5 Identifying Restrictions on Shocks

The previous section shows how one can identify rational expectations models under more general information structures than normally considered in standard macroeconomic models. The assumptions of section 1.4 are however not sufficient to identify impact matrix $\mathbf{\Pi}$, which is required to identify the shocks of the model. Identified variance-covariance matrix $\mathbf{\Pi\Pi}'$ provides some information so that only $\frac{(N-1)N}{2}$ additional restrictions are necessary to get a unique impact matrix $\mathbf{\Pi}$.

The next two sections provide strategies to extract meaningful shocks $\mathbf{u}_t$ from identified residuals $\mathbf{\Pi u}_t$. Section 1.5.1 uses DSGE models to find restrictions on $\mathbf{\Pi}$, while section 1.5.2 uses restrictions on impulse responses as it is common for structural VARs.

### 1.5.1 Restrictions from DSGE Models

The rational expectations model (1.1) can be written as follows:

$$\mathbf{\Pi}^{-1}\mathbf{y}_t = \mathbf{\Pi}^{-1}\mathbf{\Gamma}\mathbf{y}_{t-1} + \mathbf{\Pi}^{-1}\mathbf{\Phi}E[\mathbf{y}_{t+1}|\mathcal{F}_t] + \mathbf{u}_t, \tag{1.23}$$

which can be related to a system of equations derived by a DSGE model. Theory might provide zero or sign restrictions on $\mathbf{\Pi}^{-1}$, $\mathbf{\Pi}^{-1}\mathbf{\Gamma}$, or $\mathbf{\Pi}^{-1}\mathbf{\Phi}$. Note that only a few zero restrictions are necessary, for example, in a bivariate model only $\frac{(2-1)2}{2} = 1$ restriction, and in a model with three equations only $\frac{(3-1)3}{2} = 3$ restrictions.

For illustrative purpose, consider a standard New-Keynesian DSGE model with a Philips curve (1.24), a Taylor rule (1.25), and an IS-curve 1.26:

$$\pi_t = \omega E_t \pi_{t+1} + (1-\omega)\pi_{t-1} + \lambda y_t + \sigma_1 u_{1t}, \tag{1.24}$$

$$y_t = E_t y_{t+1} - \frac{1}{\sigma}(r_t - E_t \pi_{t+1}) + \sigma_2 u_{2t}, \tag{1.25}$$

$$i_t = \rho i_{t-1} + (1-\rho)(\gamma \pi_t + \eta y_t + \psi(y_t - y_{t-1})) + \sigma_3 u_{3t}. \tag{1.26}$$

This NK-model imposes thirteen zero restrictions on the rational expectations model:

$$\begin{aligned} &\mathbf{\Pi}^{-1}\mathbf{y}_t = \mathbf{\Pi}^{-1}\mathbf{\Gamma}\mathbf{y}_{t-1} + \mathbf{\Pi}^{-1}\mathbf{\Phi}E[\mathbf{y}_{t+1}|\mathcal{F}_t] + \mathbf{u}_t, \\ &\begin{pmatrix} 1 & \kappa_{12} & 0 \\ 0 & 1 & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & 1 \end{pmatrix} \begin{pmatrix} \pi_t \\ y_t \\ i_t \end{pmatrix} = \begin{pmatrix} \rho_{11} & 0 & 0 \\ 0 & \rho_{22} & 0 \\ 0 & \rho_{32} & \rho_{33} \end{pmatrix} \begin{pmatrix} \pi_{t-1} \\ y_{t-1} \\ i_{t-1} \end{pmatrix} \\ &\qquad + \begin{pmatrix} \psi_{11} & 0 & 0 \\ \psi_{21} & \psi_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} E_t[\pi_{t+1}] \\ E_t[y_{t+1}] \\ E_t[i_{t+1}] \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}. \end{aligned} \tag{1.27}$$

Identification of impact matrix $\mathbf{\Pi}$ only requires three restrictions out of the thirteen so that the NK model provides more than enough restrictions to identify the shocks. Hence, the challenge is not finding restrictions, but rather choosing the right restrictions that clearly separate the different equations.

### 1.5.2 Restrictions from Impulse Response Functions

An alternative strategy is to impose restrictions on the impulse response function as it is common in the structural VAR literature. Once the model is identified, solution 1.18 can be used to calculate IRFs for different $\mathbf{\Pi}$ until the desired short- or long-run, and zero- or sign-restrictions are satisfied. While shocks in structural VARs always produce the same impulse responses, the responses of the rational expectations model depend on how much information is available about the shock before, at, and after it is realized. This can be accommodated by imposing impulse response restrictions on the average response, which is the average response to shocks across time, where information sometimes arrive earlier and sometimes later (see definition in appendix 1.8.3). Appendix 1.8.3 shows that the average response is equal to the response to the *average shock*, which is defined as the shock which realization is learned with the same speed as the agent learns on average. Calculation of impulse response to the average shock does not require simulation, which makes identification straight-forward.

The model does not only identify shocks $\mathbf{u}_t$, but it also identifies the shocks' revise, surprise, and news components according to system (1.12). Instead of imposing restrictions on the response to the sum of all components, $\mathbf{u}_t$, one could also impose restrictions on the response to specific components, $E[\mathbf{u}_t|\mathcal{F}_{t-k}] - E[\mathbf{u}_t|\mathcal{F}_{t-k-1}]$, for a given $k$. For example, if $u_{it}$ is a monetary policy shock, one could impose that when information about the shock comes as surprise, $E[u_{it}|\mathcal{F}_{t-k}]-E[u_{it}|\mathcal{F}_{t-k-1}] = u_{it}$ for $k = 0$ and zero otherwise, the impact on output is zero within the quarter. Another example is to set the impact of a TFP shock $u_{jt}$ on TFP itself equal to zero before it is realized. Hence, the impact of a TFP news component, $E[u_{jt}|\mathcal{F}_{t-k}]-E[u_{it}|\mathcal{F}_{t-k-1}] = u_{it}$ for $k = 1$ and zero otherwise, can be restricted to have no impact on TFP before time $t$. These timing restrictions are more precise so that the impact matrix $\mathbf{\Pi}$ can be identified using weaker assumptions.

The examples of this section show how identifying restrictions can be found in many places. The application of section 1.6.4 will use a traditional approach following Blanchard and Quah (1989), by separating demand and supply shocks with a zero long-run restriction on the response of real GDP to an average demand shock.

## 1.6 Application

This section discusses data on expectations with a new metric of information gain in section 1.6.1, estimates models in section 1.6.2, and uses estimates for the following two applications. Section 1.6.3 quantifies the effect of availability of information with the moments of the data. I find that early arrival of information makes data more persistent and more pro-cyclical. Section 1.6.4 identifies impulse responses to demand and supply shocks and identifies how well these shocks are observed on average. I find that almost half of the variance of supply shocks are only observed after the shocks are realized, and the effect of information is stronger for supply than for demand shocks.

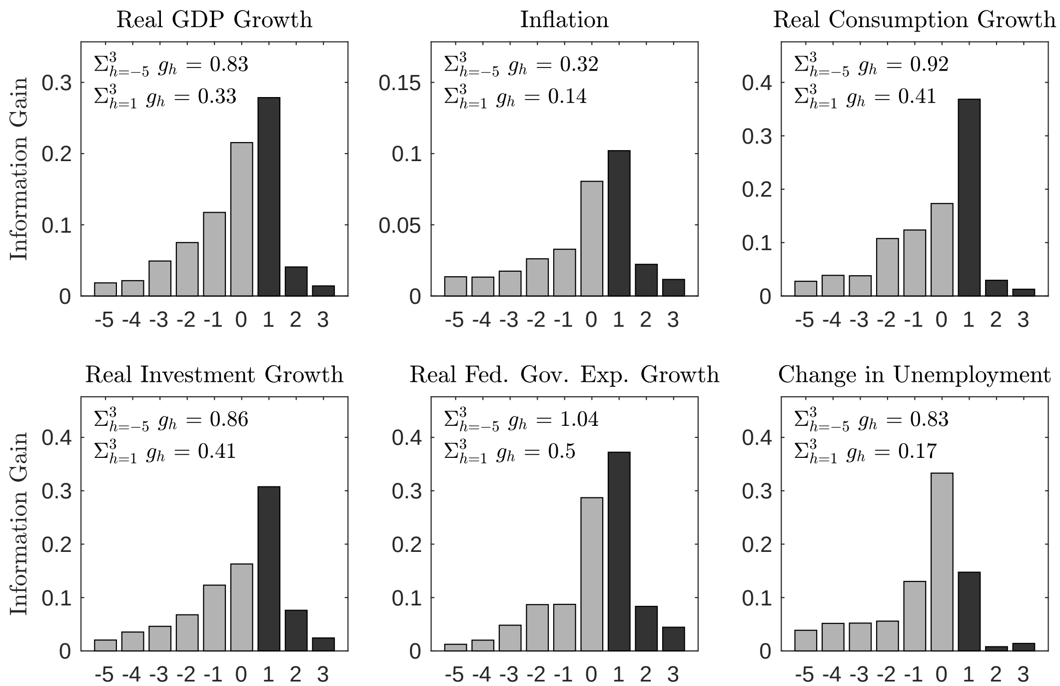
### 1.6.1 Forecaster Data

For applications, I use data on expectations from the Greenbooks, which are provided by the Federal Reserve Bank of Philadelphia. These back-, now, and forecasts are prepared for meetings of the Federal Open Market Committee (FOMC), and are available to the public with a delay of five years. I aggregate the forecasts to a quarterly series by taking average, if there is more than one meeting per quarter.

The Federal Reserve is likely the most informed economic agent in the economy, which is desirable for identification in this paper. This is because Fed's backcast revisions are likely caused by fluctuations that are also not observed by other agents so that there should be no feedback coming from expectations to this unobserved variation. An additional advantage of using Greenbooks is the wealth of forecasts across variables and horizons, which is useful for implementation.

The key identification condition for my proposed methodology, namely that agents revise their views not just about the future but also about the past, can be quantified directly through the

Figure 1.2: Information Gain




*Notes:* This figure displays information gain $g_h$ for six macro indicators. The metric is defined in equation (1.28) and it measures the share of variance that is learned $h$ quarters after realization, respectively, -$h$ quarters before realization, where $h$ is plotted on the x-axis. The share of the variance that is learned between 5 periods before and 3 periods after realization is equal to the sum of the gains, $\Sigma_{h=-5}^{3} g_h$, which is displayed at the top left in each plot. Right below that is the share of the variance that is learned after realization, $\Sigma_{h=1}^{3} g_h$. Information gain is measured using quarterly data on expectations from the Greenbook, 1978Q3 - 2011Q4. Inflation is defined as GDP price deflator, consumption as personal consumption expenditures, investment as business fixed investment, government expenditures as federal government consumption and gross investment.

following metric of information gain $g_h$:

$$g_h \equiv \frac{Var\left(E[y_t|\mathcal{F}_{t+h}] - E[y_t|\mathcal{F}_{t+h-1}]\right)}{Var(y_t)} \in [0,1]. \tag{1.28}$$

This metric measures the share of the variance of a variable $y_t$ that is learned $h$ periods after its realization, or $-h$ periods before its realization.

Under FI, information gain $g_h$ would be zero for all $h \geq 0$, as everything is observed when it is realized. For $h \leq -1$, the only reason why $g_h$ would not be equal to zero under FI is that the system is persistent so that information about today propagates into the future and thus changes expectations about the future. Under FI, these changes would reflect one to one the changes in expectations about today, which is why Manski's (1993) reflection problem occurs. When relaxing FI.ii, there is information gain after realization, $g_h \neq 0$ for $h \geq 0$, so that people learn about a period even after they experience that period. Relaxing FI.i means that people learn about the future from news shocks directly, and not just indirectly from propagating what they learn about today into the future. Hence, this metric provides a simple quantitative metric for assessing when information about different macroeconomic variables arrives and is processed by agents.

Full revelation, $\lim_{s\to\infty} E_{t+s}[y_t] = y_t$, together with RE imply $\sum_{h=-\infty}^{\infty} g_h = 1$. The share of the variance that is only observed after realization is equal to $\sum_{h=1}^{\infty} g_h$. Figure 1.2 plots information gain observed in Greenbook forecasts. Strikingly, around one third of real GDP growth is only observed after realization. This contradicts the full information assumption. Since this is a lower bound for missing information, as the Fed is one of the best forecaster in the U.S. economy, this finding provides particularly strong support for relaxing FI.

### 1.6.2 Estimation

For estimation, consider a more general version of model (1.1):

$$\mathbf{y}_t = \mathbf{q}_t + \sum_{p=1}^{P} \mathbf{\Gamma}_p \mathbf{y}_{t-p} + \sum_{r=1}^{R} \mathbf{\Phi}_r E[\mathbf{y}_{t+r}|\mathcal{F}_t] + \mathbf{\Pi}\mathbf{u}_t, \tag{1.29}$$

where $\mathbf{q}_t$ are $N$ trend components that are observed in advance, $E[\mathbf{q}_t|\mathcal{F}_{t-k}] = \mathbf{q}_t$, for all $k$, the same way as model parameters are observed in advance. These trend components take care of changes in the first moments of the data over time. This is necessary as for some variables, e.g. inflation, far horizon forecasts don't converge to a constant mean, but rather converge to a mean that changes over time. Model 1.29 also extends the number of lead and lag terms from one to $P$ lag variables, $\{\mathbf{y}_{t-1}, ..., \mathbf{y}_{t-P}\}$, and $R$ lead terms, $\{E[\mathbf{y}_{t+1}|\mathcal{F}_t], ..., E[\mathbf{y}_{t+R}|\mathcal{F}_t]\}$.

I will impose the same assumptions as discussed in section 1.4. Namely, I assume rational expectations (RE), but relax full information (FI), and I restrict the information structure so that there is no uncertainty about timing after shocks are realized (RTR), and information about shocks diffuse $H$ periods before until $K$ periods after realization (IDI).

Estimation is based on the variance-covariance matrix of the revisions observed in data on expectations, $E[\mathbf{y}_{t+h}|\mathcal{F}_t] - E[\mathbf{y}_{t+h}|\mathcal{F}_{t-1}]$, for all $h \in \{-K, ..., H\}$, where $K$ and $H$ are the same limits as assumed under IDI. Hence, there is no information gathered about shocks beyond horizon $H = 4$, which is reasonable according to figure 1.2, as information gains are close to zero for revisions further than three quarters before impact. I consider the one quarter backcasts as truth by setting $K = 1$ so that $E[\mathbf{y}_{t-1}|\mathcal{F}_t] = \mathbf{y}_{t-1}$. The reason is that backcast revisions about the further past are infrequent and sometimes capture changes in definitions rather than unobserved fluctuations. Revisions are uncorrelated across $t$ under RE, by law of iterated expectations. The generalized model with $P = H + 1$, and $R = H$ is exactly identified so that the model can fully match the variance-covariance matrix.

Similar to the more restricted model in section 1.4, appendix 1.9.3 shows that the model can be solved in terms of $\{\mathbf{A}_p\}_{p=1}^P$, $\{\mathbf{B}_r\}_{r=1}^R$, and $\mathbf{B}_0$, which are invertible functions of $\{\mathbf{\Gamma}_p\}_{p=1}^P$ and $\{\mathbf{\Phi}_r\}_{r=1}^R$. For illustrative purpose, let $K = 1$ and $H = 3$. The covariance matrix of revisions in expectations can be decomposed into the following elements:

$$\mathbf{R} \equiv E[\tilde{\mathbf{Y}}_t \tilde{\mathbf{Y}}_t'] = (\mathcal{A}\mathcal{B}_1\mathcal{B}_0\mathcal{C})\mathfrak{D}(\mathcal{A}\mathcal{B}_1\mathcal{B}_0\mathcal{C})', \quad \tilde{\mathbf{Y}}_t \equiv \begin{pmatrix} E[\mathbf{y}_{t-1}|\mathcal{F}_t] - E[\mathbf{y}_{t-1}|\mathcal{F}_{t-1}] \\ E[\mathbf{y}_t|\mathcal{F}_t] - E[\mathbf{y}_t|\mathcal{F}_{t-1}] \\ E[\mathbf{y}_{t+1}|\mathcal{F}_t] - E[\mathbf{y}_{t+1}|\mathcal{F}_{t-1}] \\ E[\mathbf{y}_{t+2}|\mathcal{F}_t] - E[\mathbf{y}_{t+2}|\mathcal{F}_{t-1}] \\ E[\mathbf{y}_{t+3}|\mathcal{F}_t] - E[\mathbf{y}_{t+3}|\mathcal{F}_{t-1}] \end{pmatrix}, \tag{1.30}$$

where appendix 1.9.4 shows that $\mathcal{A}$ is function of $\{\mathbf{A}_p\}_{p=1}^P$, $\mathcal{B}_1$ of $\{\mathbf{B}_r\}_{r=1}^R$, $\mathcal{B}_0$ of $\mathbf{B}_0$, $\mathcal{C}$ of $\mathbf{\Pi}$, and $\mathfrak{D}$ is defined in (1.13). Properties of the shocks, $E[\mathbf{u}_t\mathbf{u}_{t+h}] = 0, \forall h \neq 0$, $E[\mathbf{u}_t\mathbf{u}_t'] = I$, together with RTR and IDI imply the following covariance matrix of shock components:

$$\mathfrak{D} \equiv \begin{pmatrix} \mathfrak{D}_{-1,-1} & 0 & 0 & 0 & 0 \\ 0 & \mathfrak{D}_{00} & \mathfrak{D}_{01} & \mathfrak{D}_{02} & 0 \\ 0 & \mathfrak{D}_{01} & \mathfrak{D}_{11} & \mathfrak{D}_{12} & -\mathfrak{D}_{02} \\ 0 & \mathfrak{D}_{02} & \mathfrak{D}_{12} & \mathfrak{D}_{22} & -\sum_{i=0}^{1}\mathfrak{D}_{i,i+1} \\ 0 & 0 & -\mathfrak{D}_{02} & -\sum_{i=0}^{1}\mathfrak{D}_{i,i+1} & I - \sum_{i=0}^{2}\mathfrak{D}_{ii} \end{pmatrix}. \tag{1.31}$$

There are enough restrictions in the variance-covariance matrix of the shock components to exactly identify all the model parameters from matrix $\mathbf{R}$, which can be done with a minimization algorithm, or more efficiently with matrix calculations (see appendix 1.9.4). Inference is performed by bootstrapping the covariance matrix of revisions, $\mathbf{R}$, and re-estimating the model for each draw.

### 1.6.3 Counterfactual Moments

This section compares the effect of shocks on business cycle fluctuations when they are observed ahead of time to shocks that are only observed ex-post. If both types of shocks generate the same characteristics, the null that information and thus expectations matter can be rejected. Information enters the model exogenously, hence, according to Lucas's (1976) critique, this exercise does not reveal whether a policy that aims to provide more or less information is effective, because there might be a change in behaviour in response to such policy, which would then change the parameters of the model. There is however still a lot to learn about effects of expectations from quantifying the difference between early and late arrival of information.

Arrival of information is measured using the same metric of information gain (1.28) on shocks $u_{it}$ so that $g_h$ is the $i^{th}$ diagonal element in $\mathfrak{D}_{hh}$. Counterfactuals are generated by manipulating the information structure, $\mathfrak{D}$, while still maintaining the properties of the shocks discussed in section 2.5. Note that the counterfactual exercise does not require identification of the impact matrix $\mathbf{\Pi}$ of section 1.5, because the moments do not depend on how the residuals are decomposed. The only specification needed is the set of variables included in $\mathbf{y}_t$.

I estimate counterfactuals in a model with real GDP growth, real consumption growth, real investment growth, and inflation. I chose these four variables as they should capture the state of the economy in both, real business cycle (RBC) models and New Keynesian (NK) models at business cycle frequencies. NK models generally include interest rates, but they are fully observed in real time so that the identification strategy does not apply. If the interest rate is a state variable, estimation without it is fine as long as the four variables and expectations thereof capture the state of the interest rate. RBC models generally include total factor productivity (TFP), but the Greenbook does not provide expectations thereof so that the same idea applies for TFP as for interest rates. Counterfactuals are estimated with model (1.29) using five lag and four lead terms, based on quarterly Greenbook forecasts from 1978Q3 to 2011Q4. The moments are calculated for detrended data, hence, observed changes in means are subtracted from the variables, $\mathbf{y}_t - \mathbf{q}_t$, before calculating the moments, where $\mathbf{q}_t$ is defined in model (1.29).

Table 1.1 shows the moments of the data, and six counterfactuals. The first column lists the moments without changing the arrival of information. The first six rows describe when information arrives by calculating how much of the variance of an average shock is learned at $t-4$, $t-3$, up until $t+1$, where $t$ is the time when shocks are realized. About half the information is available before the shocks are realized ($0.07 + 0.04 + 0.07 + 0.12 + 0.25 = 0.54$), whereas the other half is available after realization (0.46). The second column simulates a *delayed info* counterfactual, where information is only gathered after the shocks are realized so that all the variance of the shocks is learned ex-post, $\mathfrak{D}_{-1-1} = I$. The *full info* counterfactual of column three simulates an economy with the standard full information (FI) assumption so that surprise components are the only non-zero components, and have a variance of one, $\mathfrak{D}_{00} = I$. The *news* counterfactuals of columns four to seven list the moments when all the information is gathered one to four periods ahead of time so that all the weights are put on the news components, $\mathfrak{D}_{hh} = I$, for $h = 1$ up to $h = 4$.

Table 1.1 reveals that whether information is available has economically significant implications. For example, the ability to respond to economic shocks ahead of time seems to increase persistence, measured as autocorrelation. In particular, the autocorrelations of output, consumption, and investment growth increase when moving from the *delayed info* to the *news* counterfactuals. Moreover, available information about the future seems to increase the overall variance, which suggests that information amplifies the effect of economic shocks.

Table 1.1 shows that the relationships between output, consumption, and investment growth become stronger the more information is available; the variables are more pro-cyclical in the *news* counterfactuals compared to the *delayed info* counterfactual, i.e. $\rho(\cdot, y_t)$ increases with informa-

Table 1.1: Moments of Data when Changing Arrival of Information

| Variance Observed | Data | Delayed Info | Full Info | News | | | |
|---|---|---|---|---|---|---|---|
| $t-4$ | 0.07 | 0 | 0 | 0 | 0 | 0 | 1.00 |
| $t-3$ | 0.04 | 0 | 0 | 0 | 0 | 1.00 | 0 |
| $t-2$ | 0.07 | 0 | 0 | 0 | 1.00 | 0 | 0 |
| $t-1$ | 0.12 | 0 | 0 | 1.00 | 0 | 0 | 0 |
| $t$ | 0.25 | 0 | 1.00 | 0 | 0 | 0 | 0 |
| $t+1$ | 0.46 | 1.00 | 0 | 0 | 0 | 0 | 0 |
| Real GDP Growth ($y_t$), detrended | | | | | | | |
| $\sigma(y_t)$ | 2.55 | 2.46 | 2.33 | 2.41 | 2.68 | 3.04 | 3.47 |
| Auto(1) | 0.43 | 0.32 | 0.31 | 0.46 | 0.51 | 0.55 | 0.61 |
| Real Consumption Growth, detrended | | | | | | | |
| $\sigma(\cdot)$ | 2.73 | 2.90 | 2.29 | 2.43 | 2.69 | 2.76 | 3.34 |
| Auto(1) | 0.15 | 0.01 | 0.05 | 0.20 | 0.26 | 0.34 | 0.42 |
| $\rho(\cdot, y_{t-1})$ | 0.34 | 0.30 | 0.31 | 0.21 | 0.28 | 0.40 | 0.46 |
| $\rho(\cdot, y_t)$ | 0.61 | 0.51 | 0.64 | 0.65 | 0.67 | 0.80 | 0.84 |
| $\rho(\cdot, y_{t+1})$ | 0.25 | 0.09 | 0.16 | 0.29 | 0.39 | 0.50 | 0.53 |
| Real Investment Growth, detrended | | | | | | | |
| $\sigma(\cdot)$ | 8.68 | 8.06 | 7.79 | 9.10 | 8.80 | 10.20 | 11.17 |
| Auto(1) | 0.35 | 0.17 | 0.23 | 0.59 | 0.63 | 0.65 | 0.65 |
| $\rho(\cdot, y_{t-1})$ | 0.08 | 0.04 | 0.05 | 0.09 | 0.11 | 0.14 | 0.17 |
| $\rho(\cdot, y_t)$ | 0.16 | 0.13 | 0.16 | 0.16 | 0.18 | 0.19 | 0.20 |
| $\rho(\cdot, y_{t+1})$ | 0.13 | 0.11 | 0.11 | 0.12 | 0.13 | 0.15 | 0.18 |
| Inflation (GDP Deflator), detrended | | | | | | | |
| $\sigma(\cdot)$ | 1.44 | 1.24 | 1.06 | 1.12 | 1.40 | 1.42 | 2.67 |
| Auto(1) | 0.31 | 0.17 | 0.16 | 0.33 | 0.42 | 0.44 | 0.29 |
| $\rho(\cdot, y_{t-1})$ | -0.33 | -0.13 | -0.15 | -0.04 | -0.29 | -0.14 | -0.61 |
| $\rho(\cdot, y_t)$ | -0.30 | -0.23 | -0.32 | -0.28 | -0.18 | -0.08 | -0.35 |
| $\rho(\cdot, y_{t+1})$ | -0.19 | -0.06 | -0.05 | -0.15 | -0.06 | -0.09 | -0.17 |

*Notes:* This table shows the moments of the data for different assumptions on the arrival of information. The first column lists the moments of detrended data, $\mathbf{y}_t - \mathbf{q}_t$, where $\mathbf{q}_t$ is the trend of system (1.29). The remaining columns show counterfactual moments by shifting the arrival of information from one quarter after realization (column 2) up to four quarters before realization (column 7). The moments of interest are standard deviation, $\sigma(\cdot)$, autocorrelation, $Auto(1)$, and correlations to detrended output growth with lag $\rho(\cdot, y_{t-1})$, without lag, $\rho(\cdot, y_t)$, and with a lead, $\rho(\cdot, y_{t+1})$. Trend and counterfactual moments are estimated based on model (1.29) with five lag and four lead terms, using quarterly data on expectations from the Greenbook, 1978Q3-2011Q4. The first six rows of column 1 list the average share of variances that is learned at time $t-k$, where $t$ is the period when shocks are realized: $\overline{Var}(E[\mathbf{u}_t|\mathcal{F}_{t-k}] - E[\mathbf{u}_t|\mathcal{F}_{t-k-1}])$. In the *delayed info* economy of column 2, the arrival of information is shifted to one period after realization so that information is only available ex-post. Column 3 displays the *full info* economy where all the information about shocks is learned at time of realization. The remaining four columns list the moments of *news* economies, where all the information arrives one quarter ahead of time (column 4), up to four quarters ahead of time (column 7). The counterfactual moments are generated by manipulating the variance-covariance matrix of the shock components, $\mathfrak{D}$, so that for example in column 7: $Var(E[\mathbf{u}_t|\mathcal{F}_{t-4}] - E[\mathbf{u}_t|\mathcal{F}_{t-5}]) = I$. Price level is measured as GDP price deflator, consumption as personal consumption expenditures, and investment as business fixed investment. Units are annualized percentage points.

tion. Expectations thus seem to contribute to the observed co-movement of macro series. According to table 1.1, investment growth is a leading variable, $\rho(\cdot, y_{t-1}) = 0.08 < 0.13 = \rho(\cdot, y_{t+1})$, while consumption growth lags output growth, $\rho(\cdot, y_{t-1}) = 0.34 > 0.25 = \rho(\cdot, y_{t+1})$. This difference seems to be driven by lack of information, only, because once information is available ahead of time, this leading and lagging behaviour no longer occurs: in the *news* counterfactual where information is available four quarters ahead of time, investment barely leads output growth anymore, $\rho(\cdot, y_{t-1}) = 0.17 < 0.18 = \rho(\cdot, y_{t+1})$, while consumption growth slightly leads output growth, $\rho(\cdot, y_{t-1}) = 0.46 < 0.53 = \rho(\cdot, y_{t+1})$, instead of responding with a delay.

The changes in the moments are more noisy for inflation. Inflation is countercyclical, but less so when information is available one to three quarters ahead of time. In particular, table 1.1 shows that the correlation of inflation and output growth is more negative in the actual economy of column 1, $\rho(\cdot, y_t) = -0.30$, than in the counterfactual economy of column 6, when information is available three quarters ahead of time, $\rho(\cdot, y_t) = -0.08$. The response of expectations to information about the near future thus seems to move inflation in the same direction as output growth, suggesting that this information causes a response on the demand side of the economy. However, if information about the future is available beyond horizon three, inflation becomes more countercyclical again, $\rho(\cdot, y_t) = -0.35$, suggesting that the supply side responds, too, as long as information is available far enough ahead of time.

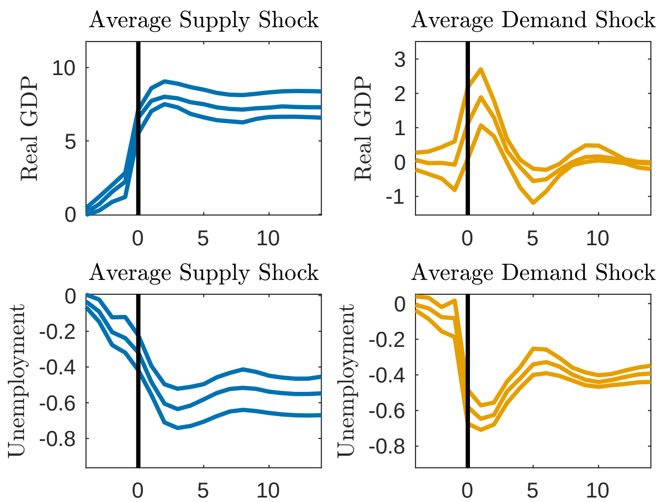
To summarize, three stylized facts show that information and therefore expectations have real effects on the economy. First, effects of information contribute to the variance of macro variables, second, they increase persistence of shocks, and third, they strengthen interdependences across variables. Hence, the exercise justifies the effort of including expectations in dynamic models, which is a common practice in the field of macroeconomics. Importantly, this conclusion is not drawn from a model where expectations need to matter by construction, in order to match persistence of the data. Instead, the model is general enough to replicate the persistence of the data without relying on effects of expectations. In other words, the model is not subject to Manski's reflection problem so that the importance of information is identified rather than assumed.

### 1.6.4 Impulse Responses

This section identifies economic shocks, how well they are observed, and their average impulse responses. In particular, the aim is to separate demand from supply shocks. Economic shocks are extracted from identified residuals, $\mathbf{\Pi u}_t$, but additional identifying restrictions are necessary to make the impact matrix unique, i.e. the variance-covariance matrix $\mathbf{\Pi\Pi}'$ together with identifying restrictions should provide a unique solution for $\mathbf{\Pi}$. Section 1.5 proposes different strategies to get restrictions from DSGE models or alternatively, from impulse responses directly. I will use impulse response functions to separate demand and supply shocks, based on the same idea as Blanchard and Quah (1989): supply shocks are the only shocks with long run impact on real GDP, while demand shocks have only a temporary effects on output. Similar to Blanchard and Quah (1989), I will separate demand from supply shocks in a bivariate system with real GDP growth and changes in unemployment rate.

Model (1.29) is estimated with five lag and four lead terms using quarterly Greenbook forecasts from 1969Q1-2011Q4. In contrast to Blanchard and Quah (1989), the model estimated in this paper differentiates between shock components observed before, at, or after the shock hits the economy. Depending on when information arrives, a shock has a stronger or a weaker effect on the economy. The zero restriction on the long run effect of a demand shock on real GDP is thus imposed on the *average* response according to (1.35).

Figure 1.3: Impulse Response Functions

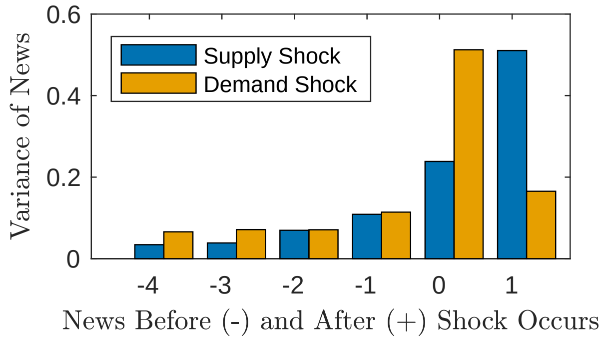



*Notes:* This figure displays responses of unemployment and real GDP to average supply and demand shocks, as obtained from system (1.29) together with a zero restriction on the long run impact of real GDP in response to an average demand shock. The model is estimated based on a backcast, a nowcast, and five forecasts, and it includes five lag and four lead terms. Data for estimation are quarterly Greenbook forecasts from 1969Q1-2011Q4. 68% confidence bands are estimated using bootstrap method. Units of real GDP are cumulated annualized percentage points, and unemployment is measured in percentage points..

Figure 1.3 plots the average impulse responses to demand and supply shocks. As imposed by the zero long run restriction, real GDP increases in response to an average demand shock, but then converges back to zero fairly quickly, while the average supply shock increases real GDP permanently. While Blanchard and Quah (1989) find that supply shocks lead to increased unemployment on impact, the impulse responses of figure 1.3 suggest that both demand and supply shocks reduce unemployment rates at all horizons, consistent with standard RBC models. Interestingly, the shocks cause significant responses even before they are realized. Information gathered ahead of time thus have real effects on the economy, by changing behaviour and decisions made in preparation of this future shock. This finding is consistent with the literature on news shocks started by Beaudry and Portier (2004, 2006) and extended by Barsky and Sims (2011), Beaudry, Dupaigne and Portier (2011), Beaudry and Portier (2014), and Barsky, Basu and Lee (2015) among others, who find that TFP shocks have real effects even before they are realized. These models use the assumption that identified TFP news shocks are indeed shocks that only realize in the future, rather than shocks that realize today. Hence, these models rely on assumptions to differentiate the direct effect of a shock, from an indirect effect of observing that shock (Manski's (1993) reflection problem). While relying on a different set of assumptions, the model estimated here can identify whether shocks are indeed realized or not so that this assumption is not necessary. Discovering significant responses before realization with a model where data alone identifies effects of expectations thus provides robust support in favour of the news shock literature.

The effect of expectations about the future seems to be stronger for supply shocks than for demand shocks, suggesting that either expectations matter more for supply shocks, or that there is more information available for supply shocks than for demand shocks. Figure 1.4 shows that the former is the case. The figure plots how much information on supply and demand shocks is

Figure 1.4: Information Gain of Shocks




*Notes:* This figure displays information gain $g_h$ for demand and supply shocks. The metric is defined in equation (1.28) and it measures the share of variance that is learned $h$ quarters after realization, respectively, -$h$ quarters before realization, where $h$ is plotted on the x-axis. The metric is equal to the variance of the shock components, $diag(\mathfrak{D}_{hh}) = Var(E[\mathbf{u}_t|\mathcal{F}_{t-h}] - E[\mathbf{u}_t|\mathcal{F}_{t-h-1}])$, as obtained from system (1.29) together with a zero restriction on the long run impact of real GDP in response to an average demand shock. The model includes five lag and four lead terms, and is estimated using quarterly Greenbook forecasts on real GDP growth and changes in unemployment from 1969Q1-2011Q4.

learned before, at, and after impact. Demand shocks are better observed, while a large share of the variance of supply shocks is only observed ex-post. This provides evidence that the demand side is well understood, but there is little consensus on the production side. The finding that supply shocks cause a significant response ahead of time suggests that supply shocks come along with large adjustments ahead of time, as for example investments in new infrastructure.

## 1.7 Conclusion

Forward-looking behavior is one of the reasons why macro variables are correlated over time according to rational expectations (RE) models. I show that in the standard estimation framework, the extent to which this persistence is driven by forward-looking behavior is a result of the underlying model rather than a feature of the data. I relate this issue to Manski's (1993) reflection problem of social interaction models, where the researcher cannot distinguish whether expectations about a group cause individual behavior or whether they just reflect individual behavior without causing it. FIRE models suffer from the same reflection problem, as the researcher cannot distinguish whether expectations about the future cause current behavior or whether they just reflect current behavior without causing it.

I use data on expectations at multiple horizons to separately identify the forward- and backward-looking components of the RE model. I first generalize the RE model by relaxing the full information (FI) assumption, which allows me to match data on expectations at all horizons. Specifically, information about a shock is not only revealed at realization as it is the case under FI, but it also diffuses before and after the shock is realized. This generalization is based on the literature on news shocks (see Beaudry and Portier, 2004, 2006), where information arrives ahead of time, and the literature on information rigidities (see Mankiw and Reis, 2002; Woodford, 2003; Sims, 2003), where information may arrive after realization. In this generalized framework, expectations about the future are no longer proportional to current and past realizations, because the agent receives separate information about the future. Hence, the generalized model no longer suffers from the reflection problem.

Identification makes use of the timing assumption that expectations about the future can only depend on information that is available today. Revisions in expectations about the past are therefore purely backward-looking and identify how expectations reflect actions. Once this direct effect is identified, the difference between projections and observed forecasts is used to identify the feedback of expectations to actions. Applying the new approach to U.S. Greenbook forecasts, I find that persistence and co-movement of macro series depend significantly on how information diffuses over time. Since information can only cause macro movements when expectations matter, this finding supports the standard assumption that expectations are relevant for business cycle fluctuations. Moreover, based on Blanchard and Quah's (1989) decomposition, I find that supply shocks are subject to more information frictions than demand shocks. Finally, consistent with RBC models, both demand and supply shocks increase output and decrease unemployment.

## 1.8 Proofs

### 1.8.1 Proof of Lemma 1

*Proof.* Guess $E[\mathbf{y}_{t+h+1}|\mathcal{F}_t] = \mathbf{A}E[\mathbf{y}_{t+h}|\mathcal{F}_t]$, and insert the guess in system (1.2), then by RE:

$$\begin{aligned}E[\mathbf{y}_{t+h}|\mathcal{F}_t] =&\mathbf{\Gamma}E[\mathbf{y}_{t+h-1}|\mathcal{F}_t] + \mathbf{\Phi}\mathbf{A}E[\mathbf{y}_{t+h}|\mathcal{F}_t]\\ =&(I-\mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}E[\mathbf{y}_{t+h-1}|\mathcal{F}_t].\end{aligned}$$

Hence, the guess can be verified if $\mathbf{A} = (I-\mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}$. □

### 1.8.2 Proof of Proposition 1.2.1

*Proof.* (a) Under RE, and FI.i, lemma 1 shows that system (1.1) can be expressed in terms of expectations about today:

$$\mathbf{y}_t =\mathbf{\Gamma}\mathbf{y}_{t-1} + \mathbf{\Phi}\mathbf{A}E[\mathbf{y}_t|\mathcal{F}_t] + \mathbf{\Pi}\mathbf{u}_t. \tag{1.32}$$

Under FI, $E[\mathbf{y}_t|\mathcal{F}_t] = \mathbf{y}_t$, so that $\mathbf{\Gamma}$, $\mathbf{\Phi}\mathbf{A}$, and $\mathbf{\Pi}$ cannot be separately identified. This is the reflection problem of Manski (1993), who denotes coefficient $\beta \equiv \mathbf{\Phi}\mathbf{A}$ as *endogenous effect* and coefficient $\eta \equiv \mathbf{\Gamma}$ as *direct effect*. (b) Take expectations of system (1.32) and solve:

$$E[\mathbf{y}_t|\mathcal{F}_t] =(I-\mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}E[\mathbf{y}_{t-1}|\mathcal{F}_t] + (I-\mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Pi}E[\mathbf{u}_t|\mathcal{F}_t], \tag{1.33}$$

then shift information sets to get for any $k \geq 0$:

$$\begin{aligned}E[\mathbf{y}_{t+k}|\mathcal{F}_t] =&E[\mathbf{y}_{t+k}|\mathcal{F}_{t-1}] + (E[\mathbf{y}_{t+k}|\mathcal{F}_t] - E[\mathbf{y}_{t+k}|\mathcal{F}_{t-1}])\\ =&(I-\mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}E[\mathbf{y}_{t+k-1}|\mathcal{F}_{t-1}] + (E[\mathbf{y}_{t+k}|\mathcal{F}_t] - E[\mathbf{y}_{t+k}|\mathcal{F}_{t-1}]).\end{aligned} \tag{1.34}$$

The regression of $E[\mathbf{y}_{t+k}|\mathcal{F}_t]$ on $E[\mathbf{y}_{t+k-1}|\mathcal{F}_{t-1}]$ is unbiased, as by law of iterated expectations:

$$E[E[\mathbf{y}_{t+k-1}|\mathcal{F}_{t-1}](E[\mathbf{y}_{t+k}|\mathcal{F}_t] - E[\mathbf{y}_{t+k}|\mathcal{F}_{t-1}])] = 0,$$

hence, composite parameter $\mathbf{A} = (I-\mathbf{\Phi}\mathbf{A})^{-1}\mathbf{\Gamma}$ is identified. □

### 1.8.3 Lemma 2 with Proof

Define average IRF, $IRF_h$, as the average difference between the variable at $t+h$ when the shock is set equal to its standard deviation, and the outcome when the shock is set equal to zero:

$$IRF_h \equiv E\left[E\left[\mathbf{y}_{t+h}|u_{it} = \sqrt{Var(u_{it})} = 1\right] - E[\mathbf{y}_{t+h}|u_{it} = 0]\right]. \tag{1.35}$$

Define average impulse or average shock as a shock which realization is learned with the same speed as the agent learns on average. The average shock $\bar{u}_{it}$ is learned as follows:

$$E[\bar{u}_{it}|\mathcal{F}_{t+k}] - E[\bar{u}_{it}|\mathcal{F}_{t+k-1}] = E\left[E[u_{it}|\mathcal{F}_{t+k}] - E[u_{it}|\mathcal{F}_{t+k-1}]\Big|u_{it} = 1\right] \equiv s_k,\ \forall k, t, \tag{1.35}$$

where $Var(E[u_{it}|\mathcal{F}_{t+k}] - E[u_{it}|\mathcal{F}_{t+k-1}]) \approx s_k$ is the linear approximation of the conditional expectation in (1.35), which is equal to $s_k$ under the assumption of normality. In the model presented in section 1.4, the following holds:

**Lemma 2.** *Average response to a shock is equal to the response to the average shock.*

*Proof.* The second expression in (1.35) can be expressed as expectations conditional on an information set where no information about the shock is available, while the first expression can be expressed as conditional expectations where all of the shock is observed so that

$$\begin{aligned} IRF_h &= E\left[E\left[E[\mathbf{y}_{t+h}|\mathcal{F}_{t+K}] - E[\mathbf{y}_{t+h}|\mathcal{F}_{t-H-1}]\Big|u_{it} = 1\right]\right], \\ &= E\left[E\left[\sum_{k=-H}^{K}\Big(E[\mathbf{y}_{t+h}|\mathcal{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathcal{F}_{t+k-1}]\Big)\Big|u_{it} = 1\right]\right], \end{aligned}$$

where the second equation follows from the fact that the agent learns about $u_{it}$ within the interval of $H$ periods before and $K$ periods after impact so that the effects of the shock are fully learned within that interval, as well. Let $x$ and $v$ be random variables, then $u_{it} = \sum_{m=-H}^{K}(E[u_{it}|\mathcal{F}_{t+m}] - E[u_{it}|\mathcal{F}_{t+m-1}])$, and $E[x|v=1] = E[E[x|v]|v=1]$ implies

$$IRF_h = E\left[E\left[\sum_{k=-H}^{K} E\left[E[\mathbf{y}_{t+h}|\mathcal{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathcal{F}_{t+k-1}]\Big|\sum_{m=-H}^{K}\Big(E[u_{it}|\mathcal{F}_{t+m}] - E[u_{it}|\mathcal{F}_{t+m-1}]\Big)\right]\Big|u_{it} = 1\right]\right]. \tag{1.36}$$

Note that by law of iterated expectations, the changes in expectations about shocks and variables are uncorrelated so that for any $z_1$ and $z_2$ (e.g. $z_1 = z_2 = u_{it}$ or $z_1 = \mathbf{y}_{t+h}$ and $z_2 = u_{it}$), and for all $m > n$:

$$\begin{aligned}
&E\left[\Big(E[z_1|\mathfrak{F}_{t+m}] - E[z_1|\mathfrak{F}_{t+m-1}]\Big)\Big(E[z_2|\mathfrak{F}_{t+n}] - E[z_2|\mathfrak{F}_{t+n-1}]\Big)\right] \\
=&E\left[E[z_1|\mathfrak{F}_{t+n}]E[z_2|\mathfrak{F}_{t+n}]\right] - E\left[E[z_1|\mathfrak{F}_{t+n-1}]E[z_2|\mathfrak{F}_{t+n-1}]\right] \\
&- E\left[E[z_1|\mathfrak{F}_{t+n}]E[z_2|\mathfrak{F}_{t+n}]\right] + E\left[E[z_1|\mathfrak{F}_{t+n-1}]E[z_2|\mathfrak{F}_{t+n-1}]\right] = 0.
\end{aligned} \tag{1.37}$$

The conditional expectation inside (1.36) can be illustrated as $E\left[E[x|v_1 + v_2]|v_1 + v_2 = 1\right]$, where $v_1, v_2$, respectively the revisions in expectations about the shocks are uncorrelated according to (1.37). Given linearity, $x = av_1 + bv_2$, the previous expression is equal to $E\left[E[x|v_1] + E[x|v_2]|v_1 + v_2|v_1 + v_2 = 1\right.$ $E\left[E[x|v_1] + E[x|v_2]|v_1 + v_2 = 1\right]$, hence,

$$IRF_h = E\left[E\left[\sum_{k=-H}^{K}\sum_{m=-H}^{K} E\left[E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k-1}]\Big|E[u_{it}|\mathfrak{F}_{t+m}] - E[u_{it}|\mathfrak{F}_{t+m-1}]\right]\Bigg|u_{it} = 1\right]\right].$$

By law of iterated expectations, $E\left[E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k-1}]|E[u_{it}|\mathfrak{F}_{t+m}] - E[u_{it}|\mathfrak{F}_{t+m-1}]\right] = 0$ for all $m \neq k$ so that

$$IRF_h = E\left[E\left[\sum_{k=-H}^{K} E\left[E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k-1}]\Big|E[u_{it}|\mathfrak{F}_{t+k}] - E[u_{it}|\mathfrak{F}_{t+k-1}]\right]\Bigg|u_{it} = 1\right]\right].$$

The above expression has the following form: $E\left[E\left[E[x_1|v_1] + E[x_2|v_2]|v_1 + v_2 = 1\right]\right]$, which is identical to $E\left[E\left[E[x_1|v_1]|v_1 + v_2 = 1\right]\right] + E\left[E\left[E[x_2|v_2]|v_1 + v_2 = 1\right]\right]$ so that

$$IRF_h = \sum_{k=-H}^{K} E\left[E\left[E\left[E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k-1}]\Big|E[u_{it}|\mathfrak{F}_{t+k}] - E[u_{it}|\mathfrak{F}_{t+k-1}]\right]\Bigg|u_{it} = 1\right]\right].$$

An expression inside the above equation has the following form: $E[E[x_1|v_1]|v_1 + v_2 = 1]$, which is equal to $E[x_1|E[v_1|v_1 + v_2 = 1]]$, since $E[x_1|v_1]$ is linear in $v_1$, hence,

$$IRF_h = \sum_{k=-H}^{K} E\left[E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k-1}]\Bigg|E\left[E[u_{it}|\mathfrak{F}_{t+k}] - E[u_{it}|\mathfrak{F}_{t+k-1}]\Big|u_{it} = 1\right]\right].$$

Note that $E\left[E[u_{it}|\mathfrak{F}_{t+k}] - E[u_{it}|\mathfrak{F}_{t+k-1}]|u_{it} = 1\right]$ is defined as $s_k$ in (1.35) and it is constant across $t$:

$$IRF_h = \sum_{k=-H}^{K} E\left[E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathfrak{F}_{t+k-1}]\Bigg|E[u_{it}|\mathfrak{F}_{t+k}] - E[u_{it}|\mathfrak{F}_{t+k-1}] = s_k\right]. \tag{1.38}$$

Taking the sum back inside expectations, and since $\sum_{k=-H}^{K}\left(E[\mathbf{y}_{t+h}|\mathcal{F}_{t+k}] - E[\mathbf{y}_{t+h}|\mathcal{F}_{t+k-1}]\right)$ is equal to $(\mathbf{y}_{t+h} - E[\mathbf{y}_{t+h}|\mathcal{F}_{t+H-1}])$:

$$IRF_h = \sum_{k=-H}^{K} E\left[\mathbf{y}_{t+h}\middle|E[u_{it}|\mathcal{F}_{t+k}] - E[u_{it}|\mathcal{F}_{t+k-1}] = s_k\right] - E[\mathbf{y}_{t+h}].$$

Let $w$ be a random variable that is linear in $v_1$ and $v_2$, then $E[w|v_1 = r_1] + E[w|v_2 = r_2] = E[w|v_1 = r_1, v_2 = r_2]$, respectively, in the big model:

$$\begin{aligned}IRF_h &= E\left[\mathbf{y}_{t+h}\middle|\begin{pmatrix}E[u_{it}|\mathcal{F}_{t-H}] - E[u_{it}|\mathcal{F}_{t-H-1}]\\ \vdots \\ E[u_{it}|\mathcal{F}_{t+K}] - E[u_{it}|\mathcal{F}_{t+K-1}]\end{pmatrix} = \begin{pmatrix}s_{-H}\\ \vdots \\ s_K\end{pmatrix}\right] - E\left[\mathbf{y}_{t+h}\middle|\begin{pmatrix}E[u_{it}|\mathcal{F}_{t-H}] - E[u_{it}|\mathcal{F}_{t-H-1}]\\ \vdots \\ E[u_{it}|\mathcal{F}_{t+K}] - E[u_{it}|\mathcal{F}_{t+K-1}]\end{pmatrix} = 0\right],\\ &= E\left[E\left[\mathbf{y}_{t+h}|u_{it} = \bar{u}_{it}\right] - E[\mathbf{y}_{t+h}|u_{it} = 0]\right],\end{aligned}$$

which is the response to the average shock, where average shock is defined in (1.35). $\square$

## 1.9 Model Solutions

### 1.9.1 Micro-Foundation and Solution of Simple RE Model (1.3)

Consider the following law of motion for inflation $\pi_t$:

$$\pi_t = \kappa\pi_{t-1} + e_t + \epsilon_t. \tag{1.39}$$

The central bank can intervene and exert effort $e_t$ to adjust inflation, while being subject to an $iid$ error which is only observed with a delay, i.e. $\epsilon_t \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$, $E[\epsilon_t|\mathcal{F}_t] = 0$ and $E[\epsilon_t|\mathcal{F}_{t+1}] = \epsilon_t$. The central bank chooses effort $e_t$ to solve the following problem:

$$\max \sum_{s=0}^{\infty} \beta^s \left[(a + \mu_{t+s})\pi_{t+s} - b\pi_{t+s}^2 - ce_{t+s}^2\right]. \tag{1.40}$$

Some level of inflation is beneficial to the economy, $a > 0$, as it keeps unemployment balanced, but too high inflation distorts prices and thus harms the economy, $b > 0$, and large interventions are costly, $c > 0$. Inflation is sometimes more beneficial than other times, $\mu_t \overset{iid}{\sim} N(0, \sigma_\mu^2)$, and the future is discounted by factor $\beta$.

Reformulate problem (1.40) in terms of expected inflation by replacing effort using (1.39), $e_{t+s} = E[\pi_{t+s}|\mathfrak{F}_t] - \kappa E[\pi_{t+s-1}|\mathfrak{F}_t]$, so that the central bank chooses $E[\pi_{t+s}|\mathfrak{F}_t]$ to maximize the following expression:

$$\max \sum_{s=0}^{\infty} \beta^s \Big[(a + \mu_{t+s|t})E[\pi_{t+s}|\mathfrak{F}_t] - bE[\pi_{t+s}|\mathfrak{F}_t]^2 - c(E[\pi_{t+s}|\mathfrak{F}_t] - \kappa E[\pi_{t+s-1}|\mathfrak{F}_t])^2\Big].$$

Solve for first order condition:

$$a + \mu_t - 2bE[\pi_t|\mathfrak{F}_t] - 2c(E[\pi_t|\mathfrak{F}_t] - \kappa E[\pi_{t-1}|\mathfrak{F}_t]) + \beta\kappa c(E[\pi_{t+1}|\mathfrak{F}_t] - \kappa E[\pi_t|\mathfrak{F}_t]) \stackrel{!}{=} 0,$$

and rearrange:

$$\begin{aligned}
[2(b+c) + \beta\kappa^2 c]E[\pi_t|\mathfrak{F}_t] &= a + 2\kappa c E[\pi_{t-1}|\mathfrak{F}_t] + \beta\kappa c E[\pi_{t+1}|\mathfrak{F}_t] + \mu_t, \\
E[\pi_t|\mathfrak{F}_t] &= \frac{a}{2(b+c)+\beta\kappa^2 c} + \frac{2\kappa c}{2(b+c)+\beta\kappa^2 c}E[\pi_{t-1}|\mathfrak{F}_t] \\
&\quad + \frac{\beta\kappa c}{2(b+c)+\beta\kappa^2 c}E[\pi_{t+1}|\mathfrak{F}_t] + \frac{1}{2(b+c)+\beta\kappa^2 c}\mu_t, \\
E[\pi_t|\mathfrak{F}_t] &= \alpha + \gamma E[\pi_{t-1}|\mathfrak{F}_t] + \phi E[\pi_{t+1}|\mathfrak{F}_t] + \mu_t.
\end{aligned}$$

From (1.39) we know that $\pi_t - E[\pi_t|\mathfrak{F}_t] = \epsilon_t$, hence,

$$\pi_t = \alpha + \gamma\pi_{t-1} + \phi E[\pi_{t+1}|\mathfrak{F}_t] + \sigma u_t,$$

where $\sigma u_t \equiv \mu_t + \epsilon_t$, and demeaning the system gets rid of $\alpha$.

### 1.9.2 Solution of Generalized RE Model

The rational expectations model (1.1) can be expressed similar to Binder and Pesaran (1997):

$$E[\mathbf{y}_t|\mathfrak{F}_t] = \mathbf{A}E[\mathbf{y}_{t-1}|\mathfrak{F}_t] + E[\mathbf{z}_t|\mathfrak{F}_t] \tag{1.15}$$

$$E[\mathbf{z}_t|\mathfrak{F}_t] = \mathbf{B}_1 E[\mathbf{z}_{t+1}|\mathfrak{F}_t] + \mathbf{B}_0\mathbf{\Pi}E[\mathbf{u}_t|\mathfrak{F}_t], \tag{1.16}$$

where $E[\mathbf{z}_t|\mathcal{F}_t]$ is a forward-looking place-holder. Now plug system (1.15) into current expectations of system (1.1):

$$\begin{aligned}\Big(\mathbf{A}E[\mathbf{y}_{t-1}|\mathcal{F}_t]+E[\mathbf{z}_t|\mathcal{F}_t]\Big) =&\boldsymbol{\Gamma}E[\mathbf{y}_{t-1}|\mathcal{F}_t]+\boldsymbol{\Phi}\Big(\mathbf{A}E[\mathbf{y}_t|\mathcal{F}_t]+E[\mathbf{z}_{t+1}|\mathcal{F}_t]\Big)+\boldsymbol{\Pi}E[\mathbf{u}_t|\mathcal{F}_t].\\ =&\boldsymbol{\Gamma}E[\mathbf{y}_{t-1}|\mathcal{F}_t]+\boldsymbol{\Phi}\Big(\mathbf{A}^2E[\mathbf{y}_{t-1}|\mathcal{F}_t]+\mathbf{A}E[\mathbf{z}_t|\mathcal{F}_t]+E[\mathbf{z}_{t+1}|\mathcal{F}_t]\Big)\\ &+\boldsymbol{\Pi}E[\mathbf{u}_t|\mathcal{F}_t]\end{aligned}$$

Rearrange:

$$\begin{aligned}(I-\boldsymbol{\Phi}\mathbf{A})E[\mathbf{z}_t|\mathcal{F}_t] =&(\boldsymbol{\Gamma}-\mathbf{A}+\boldsymbol{\Phi}\mathbf{A}^2)E[\mathbf{y}_{t-1}|\mathcal{F}_t]+\boldsymbol{\Phi}E[\mathbf{z}_{t+1}|\mathcal{F}_t]+\boldsymbol{\Pi}E[\mathbf{u}_t|\mathcal{F}_t]\\ E[\mathbf{z}_t|\mathcal{F}_t]=&\left[(I-\boldsymbol{\Phi}\mathbf{A})^{-1}\boldsymbol{\Gamma}-\mathbf{A}\right]E[\mathbf{y}_{t-1}|\mathcal{F}_t]+(I-\boldsymbol{\Phi}\mathbf{A})^{-1}\boldsymbol{\Phi}E[\mathbf{z}_{t+1}|\mathcal{F}_t]\\ &+(I-\boldsymbol{\Phi}\mathbf{A})^{-1}\boldsymbol{\Pi}E[\mathbf{u}_t|\mathcal{F}_t],\end{aligned}$$

so that together with system (1.16) one gets $[(I-\boldsymbol{\Phi}\mathbf{A})^{-1}\boldsymbol{\Gamma}-\mathbf{A}]=0,\ \mathbf{B}_1=(I-\boldsymbol{\Phi}\mathbf{A})^{-1}\boldsymbol{\Phi}$, and $\mathbf{B}_0=(I-\boldsymbol{\Phi}\mathbf{A})^{-1}$.

### 1.9.3 Solution with Multiple Lags, Leads, and Trend

The rational expectations model (1.29) can be expressed similar to Binder and Pesaran (1997):

$$E[\mathbf{y}_t|\mathcal{F}_t] =\mathbf{B}_0\mathbf{q}_t+\sum_{p=1}^{P}\mathbf{A}_pE[\mathbf{y}_{t-p}|\mathcal{F}_t]+E[\mathbf{z}_t|\mathcal{F}_t], \tag{1.41}$$

$$E[\mathbf{z}_t|\mathcal{F}_t] =\sum_{r=1}^{R}\mathbf{B}_rE[\mathbf{z}_{t+r}|\mathcal{F}_t]+\mathbf{B}_0\boldsymbol{\Pi}E[\mathbf{u}_t|\mathcal{F}_t]. \tag{1.42}$$

Now plug equation (1.41) into (1.29) repetitively until $E[\mathbf{z}_t|\mathcal{F}_t]$ is no longer function of current and future $E[\mathbf{y}_{t+r}|\mathcal{F}_t]$:

$$E[\mathbf{z}_t|\mathcal{F}_t] = \sum_{p=1}^{\max\{P,R\}} \left( \mathbb{1}_{p\leq P}(\mathbf{\Gamma}_p - \mathbf{A}_p) + \mathbb{1}_{p\leq Q} \sum_{r=1}^{R} \mathbf{\Phi}_r \mathbf{M}_p^{1+r} \right) E[\mathbf{y}_{t-p}|\mathcal{F}_t]$$
$$+ \sum_{r=1}^{R}\sum_{s=0}^{r} \mathbf{\Phi}_r \mathbf{M}_1^s E[\mathbf{z}_{t+r-s}|\mathcal{F}_t] + \mathbf{\Pi} E[\mathbf{u}_t|\mathcal{F}_t], \tag{1.43}$$

where $\mathbf{M}_p^{1+r}$ is function of $\{\mathbf{A}_1, ..., \mathbf{A}_P\}$. In particular, $\mathbf{M}_i^k$ is the $i^{th}$ top $N \times N$ submatrix from the left of $(\bar{\mathbf{M}})^k$, where $\bar{\mathbf{M}}$ is the companion matrix:

$$\bar{\mathbf{M}} \equiv \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{P-1} & \mathbf{A}_P \\ 1_N & 0_N & \cdots & 0_N & 0_N \\ \vdots & & \ddots & & \\ 0_N & 0_N & \cdots & 1_N & 0_N \end{bmatrix}.$$

The coefficients in front of past $E[\mathbf{y}_{t-p}|\mathcal{F}_t]$ need to be zero, as $E[\mathbf{z}_t|\mathcal{F}_t]$ is only forward-looking so that the following conditions need to hold:

$$\begin{aligned} \mathbf{\Gamma}_p - \mathbf{A}_p + \sum_{r=1}^{R} \mathbf{\Phi}_r \mathbf{M}_p^{1+r} &= 0, \quad \forall p \leq \min\{P, R\} \\ \mathbf{\Gamma}_p - \mathbf{A}_p &= 0, \quad \forall p > R \\ \sum_{r=1}^{R} \mathbf{\Phi}_r \mathbf{M}_p^{1+r} &= 0, \quad \forall p > P. \end{aligned}$$

Under these conditions, system (1.43) can be written as follows:

$$\begin{aligned} E[\mathbf{z}_t|\mathcal{F}_t] &= \sum_{r=1}^{R}\sum_{s=0}^{r} \mathbf{\Phi}_r \mathbf{M}_1^s E[\mathbf{z}_{t+r-s}|\mathcal{F}_t] + \mathbf{\Pi} E[\mathbf{u}_t|\mathcal{F}_t], \\ &= \sum_{r=1}^{R} \left(1 - \sum_{s=1}^{R} \mathbf{\Phi}_s \mathbf{M}_1^s\right)^{-1} \sum_{s=l}^{R} \mathbf{\Phi}_s \mathbf{M}_1^{s-r} E[\mathbf{z}_{t+r}|\mathcal{F}_t] + \left(1 - \sum_{s=1}^{R} \mathbf{\Phi}_s \mathbf{M}_1^s\right)^{-1} \mathbf{\Pi} E[\mathbf{u}_t|\mathcal{F}_t] \\ &\equiv \sum_{r=1}^{R} \mathbf{B}_r E[\mathbf{z}_{t+r}|\mathcal{F}_t] + \mathbf{B}_0 \mathbf{\Pi} E[\mathbf{u}_t|\mathcal{F}_t], \end{aligned}$$

and therefore,

$$\mathbf{B}_r = \sum_{s=l}^{R} \mathbf{\Phi}_s \mathbf{M}_1^{s-r}, \quad \mathbf{B}_0 = \left(1 - \sum_{s=1}^{R} \mathbf{\Phi}_s \mathbf{M}_1^{s}\right)^{-1}.$$

### 1.9.4 Decomposition of Revisions Variance-Covariance Matrix

Define $\mathbf{R}_{ij}$ as the $\{ij\}^{th}$ $N \times N$ matrix in $\mathbf{R}$ of system 1.30. Inspection of 1.18 reveals:

$$\begin{pmatrix} \mathbf{M}_1^1 \\ \mathbf{M}_1^2 \\ \mathbf{M}_1^3 \\ \mathbf{M}_1^4 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{21} \\ \mathbf{R}_{31} \\ \mathbf{R}_{41} \\ \mathbf{R}_{51} \end{pmatrix} \mathbf{R}_{11}^{-1}, \quad \mathcal{A} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\mathbf{A}_1 & 1 & 0 & 0 & 0 \\ -\mathbf{A}_2 & -\mathbf{A}_1 & 1 & 0 & 0 \\ -\mathbf{A}_3 & -\mathbf{A}_2 & -\mathbf{A}_1 & 1 & 0 \\ -\mathbf{A}_4 & -\mathbf{A}_3 & -\mathbf{A}_2 & -\mathbf{A}_1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \mathbf{M}_1^1 & 1 & 0 & 0 & 0 \\ \mathbf{M}_1^2 & \mathbf{M}_1^1 & 1 & 0 & 0 \\ \mathbf{M}_1^3 & \mathbf{M}_1^2 & \mathbf{M}_1^1 & 1 & 0 \\ \mathbf{M}_1^4 & \mathbf{M}_1^3 & \mathbf{M}_1^2 & \mathbf{M}_1^1 & 1 \end{pmatrix}$$

Define $\bar{\mathbf{R}} \equiv \mathcal{A}^{-1}\mathbf{R}\mathcal{A}\prime^{-1}$. Inspection of (1.21) reveals

$$\begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \\ \mathbf{B}_3' \end{pmatrix} = \begin{pmatrix} \sum_{i=3}^{5} \bar{\mathbf{R}}_{ii} & \sum_{i=3}^{4} \bar{\mathbf{R}}_{i,i+1} & \bar{\mathbf{R}}_{35} \\ \sum_{i=3}^{4} \bar{\mathbf{R}}_{i,i+1} & \sum_{i=4}^{5} \bar{\mathbf{R}}_{ii} & \bar{\mathbf{R}}_{45} \\ \bar{\mathbf{R}}_{35} & \bar{\mathbf{R}}_{45} & \bar{\mathbf{R}}_{55} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=2}^{4} \bar{\mathbf{R}}_{i,i+1} \\ \sum_{i=2}^{3} \bar{\mathbf{R}}_{i,i+2} \\ \bar{\mathbf{R}}_{25} \end{pmatrix}.$$

Matrices $\mathcal{B}_1$, $\mathcal{B}_0$, and $\mathcal{C}$ are defined as follows:

$$\mathcal{B}_1 \equiv \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\mathbf{B}_1 & -\mathbf{B}_2 & -\mathbf{B}_3 \\ 0 & 0 & 1 & -\mathbf{B}_1 & -\mathbf{B}_2 \\ 0 & 0 & 0 & 1 & -\mathbf{B}_1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}^{-1}, \ \mathcal{B}_0 \equiv \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{B}_0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{B}_0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{B}_0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{B}_0 \end{pmatrix}, \ \mathcal{C} \equiv \begin{pmatrix} \mathbf{\Pi} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{\Pi} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{\Pi} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{\Pi} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{\Pi} \end{pmatrix}.$$

Matrix $\mathbf{B}_0$ requires mapping of $\{\mathbf{A}_p\}_{p=1}^{P}$ and $\{\mathbf{B}_r\}_{r=1}^{R}$ into $\{\mathbf{\Gamma}_p\}_{p=1}^{P}$ and $\{\mathbf{\Phi}_r\}_{r=1}^{R}$, which is derived in appendix 1.9.3. Define $\bar{\bar{\mathbf{R}}} \equiv (\mathcal{B}_{\mathbf{1}}\mathcal{B}_{\mathbf{0}})^{-1}\bar{\mathbf{R}}(\mathcal{B}_{\mathbf{0}}{}'\mathcal{B}_{\mathbf{1}}')^{-1}$ so that $\mathbf{\Pi}\mathbf{\Pi}' = \sum_{i=1}^{5} \bar{\bar{\mathbf{R}}}_{ii}$, where additional identifying restrictions are imposed to get $\mathbf{\Pi}$, as discussed in section 1.5. Once $\mathbf{\Pi}$ is identified, the covariance matrix of shock components is identified, as well: $\mathfrak{D} = \mathcal{C}^{-1}\bar{\bar{\mathbf{R}}}\mathcal{C}'^{-1}$.

## 1.10 Model Extensions

### 1.10.1 RE Model where All Beliefs Matter

Consider the following extension of RE model (1.1):

$$\begin{aligned}\mathbf{y}_t =&\boldsymbol{\Gamma} E[\mathbf{y}_{t-1}|\mathcal{F}_t] + \boldsymbol{\Phi} E[\mathbf{y}_{t+1}|\mathcal{F}_t] + \boldsymbol{\Pi} E[\mathbf{v}_t|\mathcal{F}_t]\\ &+ \tilde{\boldsymbol{\Gamma}}(\mathbf{y}_{t-1} - E[\mathbf{y}_{t-1}|\mathcal{F}_t]) + \tilde{\boldsymbol{\Pi}}(\mathbf{v}_t - E[\mathbf{v}_t|\mathcal{F}_t]).\end{aligned} \tag{1.44}$$

Define $\mathbf{u}_t = E[\mathbf{v}_t|\mathcal{F}_t] + \boldsymbol{\Pi}^{-1}\tilde{\boldsymbol{\Pi}}(\mathbf{v}_t - E[\mathbf{v}_t|\mathcal{F}_t])$:

$$\begin{aligned}\mathbf{y}_t =&\boldsymbol{\Gamma} E[\mathbf{y}_{t-1}|\mathcal{F}_t] + \boldsymbol{\Phi} E[\mathbf{y}_{t+1}|\mathcal{F}_t] + \boldsymbol{\Pi}\mathbf{u}_t\\ &+ \tilde{\boldsymbol{\Gamma}}(\mathbf{y}_{t-1} - E[\mathbf{y}_{t-1}|\mathcal{F}_t]).\end{aligned} \tag{1.45}$$

By RTR assumption, $E[\mathbf{u}_t\mathbf{u}_{t+h}'] = 0$ for all $h \neq 0$. Redefine $\boldsymbol{\Pi}$ so that $E[\mathbf{u}_t\mathbf{u}_t'] = I$.

Taking expectations cancels out prediction errors so that the system resembles standard first order conditions:

$$E[\mathbf{y}_t|\mathcal{F}_t] =\boldsymbol{\Gamma} E[\mathbf{y}_{t-1}|\mathcal{F}_t] + \boldsymbol{\Phi} E[\mathbf{y}_{t+1}|\mathcal{F}_t] + \boldsymbol{\Pi} E[\mathbf{u}_t|\mathcal{F}_t].$$

Backcast revisions on the other hand only depend on revisions in fundamentals, rather than beliefs:

$$\begin{aligned}E[\mathbf{y}_{t-1}|\mathcal{F}_t] - E[\mathbf{y}_{t-1}|\mathcal{F}_{t-1}] =&\tilde{\boldsymbol{\Gamma}}\left(E[\mathbf{y}_{t-2}|\mathcal{F}_t] - E[\mathbf{y}_{t-2}|\mathcal{F}_{t-1}]\right)\\ &+ \boldsymbol{\Pi}\left(E[\mathbf{u}_{t-1}|\mathcal{F}_t] - E[\mathbf{u}_{t-1}|\mathcal{F}_{t-1}]\right).\end{aligned} \tag{1.17}$$

### 1.10.2 Additional Shock Terms

Relaxing FI might change the structure of the model, depending on how the shocks are micro-founded. A larger set of models is nested in a structure that depends on expected future shocks separately. Consider the following extension of RE model (1.29):

$$\mathbf{y}_t = \mathbf{q}_t + \sum_{p=1}^{P} \mathbf{\Gamma}_p \mathbf{y}_{t-p} + \sum_{r=1}^{R} \mathbf{\Phi}_r E[\mathbf{y}_{t+r}|\mathcal{F}_t] + \sum_{r=1}^{R} \mathbf{\Pi}_r E[\mathbf{u}_{t+r}|\mathcal{F}_t] + \mathbf{\Pi}_0 \mathbf{u}_t. \tag{1.46}$$

Repeated substitution of expected future shocks lead to an infinite forward-looking model that no longer depends on future shocks separately:

$$\mathbf{y}_t = \check{\mathbf{q}}_t + \sum_{p=1}^{P} \check{\mathbf{\Gamma}}_p \mathbf{y}_{t-p} + \sum_{r=1}^{\tilde{R}} \check{\mathbf{\Phi}}_r E[\mathbf{y}_{t+r}|\mathcal{F}_t] + \check{\mathbf{\Pi}} \mathbf{u}_t, \tag{1.47}$$

where $\tilde{R}$ is infinity, but it can be truncated if system is stable. The structure of (1.47) is the same as in system (1.29) so that parameters $\check{\mathbf{\Gamma}}_p$ and $\check{\mathbf{\Phi}}_r$, as well as residuals $\check{\mathbf{\Pi}}\mathbf{u}_t$ with all the components are identified. The identified parameters cannot be mapped to the original model parameters without additional assumptions. The identified parameters together with impulse response restrictions on $\check{\mathbf{\Pi}}$ are enough to calculate variance decompositions and impulse response functions to the shocks of the model.

# Chapter 2

# News Shocks and Uncertainty

## 2.1 Introduction

Macroeconomic fluctuations are often unrelated to changes in contemporaneous fundamentals. This phenomena is the source of A. Pigou's (1927) theory that not only variations in fundamentals, but also variations in *expectations* cause industrial fluctuations[1]. A revival of the "Pigouvian" hypothesis in modern macroeconomics is initiated by the pioneering work of Beaudry and Portier (2004, 2006). They argue that agents receive information about future developments in productivity, which changes their behaviour today. They show that these "news shocks" cause fluctuations in the economy before the anticipated event is happening.

The research question of this paper is whether and how these news shocks are related to uncertainty. Such relationship suggests itself because news shocks and uncertainty both deal with anticipation: news shocks are *anticipated* changes in the future, whereas uncertainty impairs this *anticipation*[2]. I discover a new stylized fact that manifests this relationship: uncertainty jumps whenever good and bad news shocks hit the economy.

Rich dynamics following news shocks are already observed by Beaudry and Portier (2006). They argue that during this period, agents try best to position themselves to take advantage of future technological change. I reason that some agents reposition themselves, because the initial news they receive turn out to be wrong. A news shock therefore increases uncertainty, until the news about future productivity is verified.

Research on the role of uncertainty in the macroeconomy is a topical field since the seminal paper of Nicholas Bloom (2009). Bloom (2009) and others show theoretically and empirically that

[1]Pigou's (1927) idea is that expectations are driven by 'real', 'psychological' or 'monetary' impulses, and that they cause immediate response in the industry (see David Collard, 1996).

[2]Frank H. Knight (1921) defines uncertainty as the inability to measure probabilities of future events.

uncertainty may depress the economy. The new stylized fact provides a new source of uncertainty, but it also challenges existing identification procedures. Bloom (2009) for example identifies an uncertainty shock as the shock that shifts all series on impact except stock prices.[3] I show that impact restrictions are not enough to separate news shocks from uncertainty shocks. This is because there is no clear lead-lag behaviour between news shocks and uncertainty. This means a news shock might either be initiated by a change in stock prices, or a change in uncertainty. The intuition is that both series capture expectations about the future, but which series moves first varies over time. Another issue is that the identified uncertainty shock produces similar dynamics as a bad news shock: a spike in uncertainty is followed by a drop in output. A structural VAR model is estimated where uncertainty shocks are separated from news shocks in section 2.3. The new model produces weaker effects of uncertainty shocks than previous models.

As far as I know, this is the first study on the relationship between macro uncertainty and news shocks.[4] The discussion is however related to the research on stock returns and volatility in the finance literature. On the one hand, stock prices and news shocks both deal with expectations: news shocks are innovations to expectations, whereas stock prices are highly sensitive to changes in expectations. On the other hand, volatility and uncertainty are closely related, because the more volatile the series, the harder it is to forecast. Stock return and volatility may therefore capture some of the relationship between news shocks and uncertainty. Indeed, section 2.4 shows that the asymmetric correlation of stock returns and volatility is mainly caused by news shocks in a quarterly time-series framework. News shocks therefore provide an alternative explanation for the

[3]Impact restrictions are also used among others in Rüdiger Bachmann, Steffen Elstner and Eric R. Sims (2013), Kyle Jurado, Sydney C. Ludvigson and Serena Ng (2015), Scott R. Baker, Nicholas Bloom and Steven J. Davis (2015), and Sylvain Leduc and Zheng Liu (2015) for the identification of uncertainty shocks. Section 2.3 discusses the literature.

[4]A news shock is defined as an anticipated TFP shock in the spirit of Beaudry and Portier (2006).

stock returns to volatility relationship discussed in the finance literature.[5]

Since both good and bad news shocks increase uncertainty, it is likely that they have asymmetric effects on other variables, as well. These asymmetries are accommodated for in section 2.5 with a non-linear VAR, using a new identification procedure: instead of maximizing the forecast error variance as in Neville Francis, Michael T. Owyang, Jennifer E. Roush and Riccardo DiCecio (2014), the integral of the forecast error variance is optimized over a normal distribution. The maximization of the average variance allows to identify news shocks with long-run restrictions, even though their effect is non-linear.

Impulse response functions (IRFs) of the non-linear model produce three new properties. First, good and bad news shocks both increase uncertainty, while the increase caused by the bad news shock is five times stronger. Second, the confidence bands of the IRFs to bad news shocks are large. The strong increase in uncertainty after bad news shocks explains this inaccuracy: uncertainty makes the signal about future TFP noisy, which then adds additional noise to the responses. Finally, both good and bad news shocks decrease inflation. Intuitively, a good news shock decreases inflation if the supply side anticipates the lower marginal costs of the future (see Barsky, Basu and Lee, 2015). The high uncertainty after bad news shocks seems to impair this anticipation so that the initial response of a bad news shock is purely demand driven.

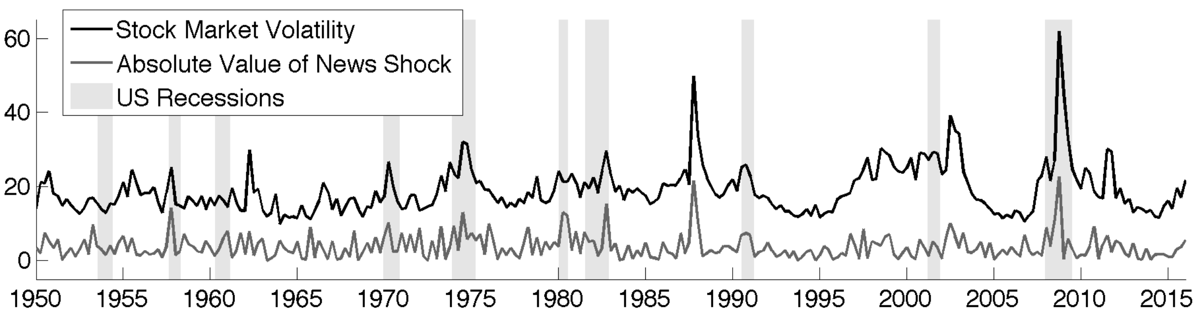
Section 2.2 presents the new stylized fact. Section 2.3 identifies an uncertainty shock that is orthogonal to good and bad news shocks. The relationship between stock returns and volatility is reassessed in section 2.4. Based on the findings of sections 2.2-2.4, a non-linear empirical

[5]Existing theories are among others the volatility feedback effect of Robert S Pindyck (1984), Kenneth R. French, G.William Schwert and Robert F. Stambaugh (1987), John Y. Campbell and Ludger Hentschel (1992), Geert Bekaert and Guojun Wu (2000), and Guojun Wu (2001), and the leverage effect of F. Black (1976), Andrew A. Christie (1982), G. William Schwert (1989). Section 2.4 discusses the literature.

model is estimated in section 2.5 that allows for asymmetric responses to news shocks. Section 3.6 concludes.

## 2.2 A New Stylized Fact

Figure 2.1: News Shocks and Stock Market Volatility




*Notes:* This figure plots U.S. stock market volatility and the absolute value of news shocks multiplied by five from 1950Q1 to 2015Q3. The shaded areas indicate U.S. recessions. **U.S. stock market volatility** is the quarterly average of the volatility index used in Bloom's (2009) figure 1: from 1986 onward the series is the Chicago Board of Options Exchange VXO index of percentage implied volatility (S&P100 option with 30 days to expiration). Pre 1986, the VXO index is unavailable, so actual returns volatilities are calculated as the standard deviation of the daily S&P500 index normalized to the same mean and variance as the VXO index when they overlap. **News shocks** are identified with Barsky and Sims's (2011) model (see section 2.2.1 and figure 2.3 for details about the model). The shaded areas are NBER based **U.S. recession** indicators for the United States, collected from the database of the Federal Reserve Bank of St.Louis (FRED).

Figure 2.1 reveals a new stylized fact: large news shocks coincide with spikes in uncertainty. Indeed, stock market volatility – a proxy for uncertainty – is positively correlated with good (0.22) and bad news shocks (0.71). Figure 2.1 shows the striking co-movement between uncertainty and the absolute value of a news shock (0.52). Before the new stylized fact is discussed in detail, I present literature, data and models of the two correlated series.

### 2.2.1 News Shock

A news shock in the spirit of Beaudry and Portier (2006) is an anticipated change in future productivity. A good news shock therefore increases future total factor productivity (TFP) with a delay, while boosting current activity through expectations. The empirical model of Beaudry and Portier (2006) consists of two series: TFP, reflecting a fundamental variable, and stock prices as proxy for expectations. The model makes use of the assumption that news about future productivity is reflected in current stock prices. Hence, a good news sock is assumed to increase expectations and therefore stock prices today. Barsky and Sims (2012) show that the consumer confidence index by the Michigan Survey of Consumers provides an alternative proxy for expectations, which also responds to news shocks.

Beaudry and Portier (2006) apply two different strategies to identify news shocks. The first model imposes that news shocks are the only shocks with a long-run impact on TFP. The second model restricts the effect of a news shock on TFP to be zero on impact. Their main finding is that both models produce almost the same residuals, suggesting that long-run fluctuations in TFP are driven by news shocks. Their expanded version of the model, and similar models by Paul Beaudry and Bernd Lucke (2010) and Beaudry, Dupaigne and Portier (2011) generate news shocks that produce co-movement in macro variables, and they find that news shocks are responsible for significant fluctuations in output at business cycle frequencies. Alternative models by Barsky and Sims (2011, 2012) and Barsky, Basu and Lee (2015) suggest that news shocks fail to account for most recessions, and that inflation, investment, and hours do not co-move with news shocks. The reason for the different findings is that results depend on how long-run co-variation is modelled, and on the identifying assumptions.

The underlying model of a news shock is a structural vector autoregression (SVAR) model that is either estimated in levels, or includes an error correction term (ECT)[6]. The series are not detrended, because of the very definition of a news shock: a news shock is an anticipated *long-run* increase in total factor productivity (TFP). The identification of a news shock therefore needs the information about the far future, which is stored in the low frequencies of the series. The drawback of working with non-stationary series is that results depend on how the long-run relationships are modelled, respectively, which and how many cointegrating relationships are accounted for. The original model by Beaudry and Portier (2006) is a vector error correction model (VECM) with one cointegrating relationship.

Barsky and Sims (2011) introduce an alternative identification procedure for news shocks. They identify the news shock as the shock that is orthogonal to current TFP, and that accounts for the largest share of future TFP movements.[7] The advantage of this procedure is that no other restrictions are required to identify TFP shocks, even if the model consists of more than two variables. In contrast to Beaudry and Portier (2006), their model is estimated in levels. Barsky, Basu and Lee (2015) estimate larger models with and without error correction terms, using a similar identification strategy as Barsky and Sims (2011).[8] Barsky and Sims (2012) apply an agnostic VAR identification that considers all shocks as news shocks that are orthognal to the innovation in current TFP and that explain most of future levels of productivity.[9]

Figure 2.2 plots news shocks produced with five different models using alternative mod-

[6]The error correction term captures the long-run co-variation between the series, if the variables are cointegrated, thus if they have a common long-run stochastic trend.

[7]This finite-horizon identification method of maximizing the expected forecast error variance is developed by Francis et al. (2014).

[8]In contrast to Barsky and Sims (2011) who maximize the share of movements in technology over all of the 40 quarters after impact, Barsky, Basu and Lee (2015) only maximize the share at a specific quarter.

[9]The agnostic identification is a method developed by Harald Uhlig (2004).

elling of long-run variation, and alternative identification strategies. The correlation among the shocks are between 0.54 and 0.88, and the new stylized fact holds for all models: correlations of the absolute values to stock market volatility are between 0.26 and 0.55.

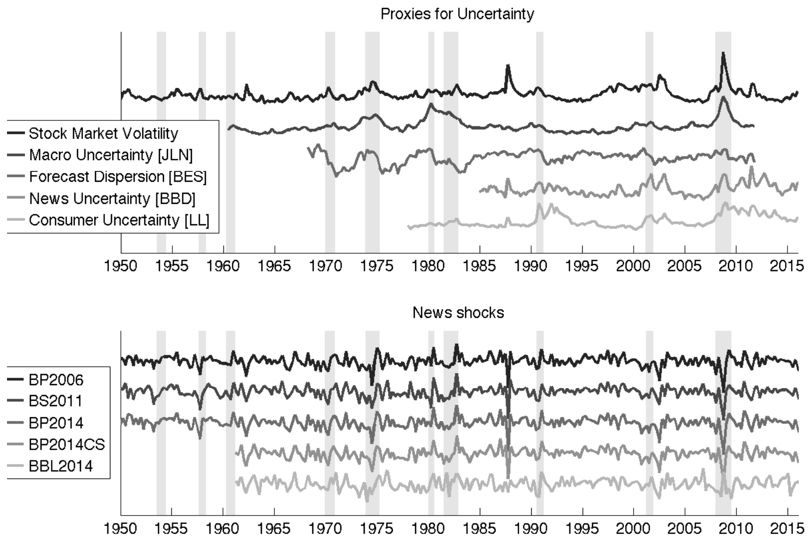
The discussed identification procedures require the model to be invertible. A non-invertability problem may occur if shocks are anticipated today, but only shift future observations. Intuitively, an anticipated shock that is only observed in the future cannot be captured with present and past observations of a backward-looking VAR. In the news shock literature this issue is solved by adding forward looking variables as stock prices or consumer confidence to the model. The necessary assumption is that these forward looking variables respond to an anticipated change in TFP so that the shocks are observed today and thus can be identified.[10]

Blanchard, L'Huillier and Lorenzoni (2013) argue that a VAR representation cannot distinguish a news shock from a noise shock. A noise shock is a shock that contaminates the news about future productivity. A noise shock therefore increases expectations today, even though future productivity remains the same. The VAR is still able to capture the correct response to the noisy signal the agents receive. Hence, the VAR captures the response to a mix of news and noise shocks.

In the news shock literature, technology is commonly measured as total factor productivity (TFP) by calculating the Solow residual. The Solow residual is then adjusted for fluctuations beyond TFP, as factor utilization. Susanto Basu, John G. Fernald and Miles S. Kimball (2006) provide adjusted annual TFP data constructed from disaggregated series with sectoral production functions. Fernald (2012) expands this series to quarterly TFP. Consistent with the most recent

[10]For non-invertability problem see Lucia Alessi, Matteo Barigozzi and Marco Capasso (2008) for a review, Eric M. Leeper, Todd B. Walker and Shu-Chun Susan Yang (2008) for problems with anticipated shocks, and Fernández-Villaverde et al. (2007) for adding forward looking variables.

Figure 2.2: Proxies for Uncertainty and News Shocks Produced with Different Models



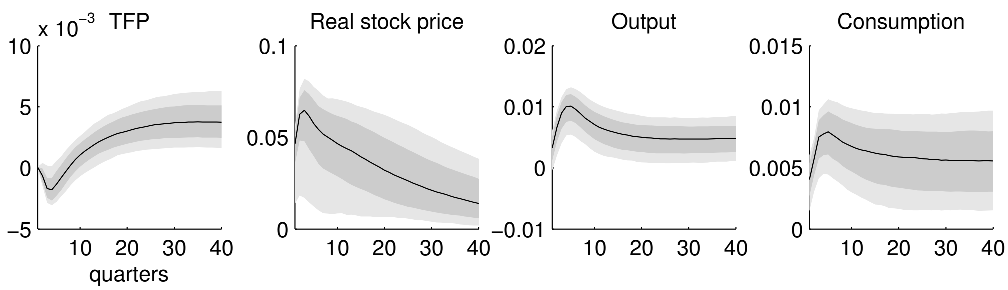

Correlations across news shocks to BS2011 (benchmark): 0.86, 1.00, 0.88, 0.88, 0.54.
Correlations across proxies to Stock Market Volatility (benchmark): 1.00, 0.50, 0.19, 0.52, 0.23.
Correlations different proxies for uncertainty to absolute value of BS2011: 0.52, 0.38, 0.08, 0.36, 0.13.
Correlations absolute value of different news shock to Stock M. Volatility: 0.55, 0.52, 0.54, 0.49, 0.26.

*Notes:* This figure plots five alternative proxies for uncertainty of the U.S. economy (upper panel), and five news shocks identified with different models based on U.S. data (lower panel). The standard deviations of the uncertainty proxies are normalized to one. Proxies for uncertainty: 1) **Stock market volatility** is the measure declared in figure 2.1, 2) **Macro uncertainty [JLN]** is Jurado, Ludvigson and Ng's (2015) approximate of uncertainty based on hundreds of macroeconomic and financial indicators. 3) **Forecast dispersion [BES]** is Bachmann, Elstner and Sims's (2013) measure that captures the level of disagreement amongst respondents of the Business Outlook Survey. 4) **News uncertainty [BBD]** is Baker, Bloom and Davis's (2015) measure of policy uncertainty based on the frequency of certain words written in news reports. 5) **Consumer uncertainty [LL]** is Leduc and Liu's (2015) measure based on Michigan's Consumer Survey: uncertainty is the share of consumers that state uncertainty is the reason why they don't buy a car. News shocks are estimated with John G. Fernald's (2012) adjusted TFP measure, all variables are in logs, and quantity variables are seasonally adjusted. The models that include consumer sentiment of the Michigan Consumer Survey cover the period 1960Q1-2015Q3, and the other models start from 1950Q1. 1) **BP2006**: Beaudry and Portier's (2006) VECM(3) model with one cointegrating relationship, and with TFP and R. J. Shiller's (2015) nominal stock prices, where news shock is the only shock with a long-run impact on TFP. 2) **BS2011**: Barsky and Sims' (2011) VAR(3) model in levels, with TFP, Shiller's (2015) real stock prices, per capita real GDP and per capital real consumption expenditures, where news shock is the only shock orthogonal to current TFP that maximizes the share of TFP movements of the following 40 quarters. 3) **BP2014**:Beaudry and Portier's (2014) VECM(3) model with two cointegrating relationships, and with TFP and Shiller's (2015) nominal stock prices, where news and surprise TFP shocks are the only shocks with a long-run impact on TFP, and where news shock is orthogonal to current TFP. 4) **BP2014CS**: Beaudry and Portier's (2014) VAR(4) model in levels with TFP, Shiller's (2015) nominal stock prices, and consumer sentiment (CS) from the Michigan Consumer Survey, where news and surprise TFP shocks are the only shocks with a long-run impact on TFP, and where news shock is orthogonal to current TFP. 5) **BBL2014**: Barsky, Basu and Lee's (2015) VAR(4) model in levels with TFP, Shiller's (2015) real stock prices, consumer sentiment (CS), three months treasury bill rate, CPI inflation, per capita real consumption of non-durables, per capita real consumption of durables, per capita real GDP investment, and per capita hours in non-farm sector, where news shock is the only shock orthogonal to current TFP that maximizes the share of TFP movements at the $40^{th}$ quarter after impact.

studies on news shocks, I use Fernald's (2012) series as measure for productivity.

The news shock of figure 2.1 is produced with a VAR model in levels, using Barsky and Sims's (2011) identification strategy: a news shock is orthogonal to current TFP and maximizes the share of future TFP movements.[11] The model consists of four variables: Fernald's (2012) TFP measure, real stock prices S&P500 from Robert J. Shiller's (2015) *Irrational Exuberance*, real personal consumption expenditures and real GDP gathered from the Federal Reserve Bank of St. Louis (FRED). The two quantity variables are seasonally adjusted and per capita, divided by the civilian, non-institutional population. All variables are in logs.

Figure 2.3: IRFs to a News Shock of Barsky and Sims's (2011)




*Notes:* This figure plots the impulse response functions of a news shock produced with Barsky and Sims' (2011) VAR(3) model in levels, with Fernald's (2012) adjusted total factor productivity (**TFP**), Shiller's (2015) **Real stock prices**, per capita real GDP (**Output**, source: FRED) and per capital real consumption expenditures (**Consumption**, source: FRED). News shocks are the only shocks orthogonal to current TFP that maximize the share of future TFP movements of the following 40 quarters. Variables are in logs and the quantity series are seasonally adjusted. The bootstrapped confidence bands show the $68^{th}$ and $90^{th}$percentiles.

The impulse responses are plotted in figure 2.3. The responses are as expected: a good

[11]The identifying assumptions are imposed by rotating the impact matrix $A_0$ of the reduced form VAR: $X_t = X_{t-1}B + A_0\epsilon_t$ (see James D. Hamilton (1994)). The rotation matrices are constructed following the approach of Luca Benati (2016). In contrast to the analytical approach of previous studies, this approach does not need to impose zero impact restrictions on TFP of the remaining shocks. The angles are optimized using the medium-scale quasi-Newton line search of Matlab's *fminunc* function.

news shock increases expectations and thus stock prices, consumption and output today, whereas, TFP only increases with a delay. Note that uncertainty is not part of the model that identifies the news shock. Hence, the strong correlation to uncertainty is not produced endogenously. Moreover, the other three shocks of the model are not correlated to uncertainty, as discussed in section 2.2.3.

### 2.2.2 Uncertainty

In economics, uncertainty is often defined based on Frank Knight's (1921) work on profits. He argues that uncertainty causes profits, whereas risk does not. Risk occurs if the future consists of more than one state, while the probabilities of the states are measurable. These probabilities are incalculable under uncertainty. Events under uncertainty can therefore not be anticipated. Hence, firms are not able to insure future states under uncertainty so that profits arise.

The definition of uncertainty in Macroeconomics is often a mix between Knight's (1921) understanding of risk and uncertainty. For example in the literature on "real options", uncertainty refers to an increase in possible future states.[12] It is assumed that in each period firms have a series of real options to choose from. For example, a real option is investing in a specific technology. This investment only pays out in a subset of future states. The investment is profitable, if a large enough share of the future states are within this subset. However, if the number of future states increases due to uncertainty, the share of future states within the subset gets smaller. The expected return on a specific investment therefore decreases with uncertainty. Hence, if uncertainty is high, the real option value of investing decreases, whereas the real option value of waiting increases. This relationship holds when investment cannot be undone without cost, thus, if there are adjustment

[12] Literature on real options include Ben S. Bernanke (1983), Michael J. Brennan and Eduardo S. Schwartz (1985), Robert McDonald and Daniel Siegel (1986), and Bloom (2009).

costs.

Bloom (2009) models this uncertainty effect and shows that with depreciation of capital and exogenous attrition of workers, high uncertainty leads to a drop in employment, investment, and output. He also shows that the real option effect dampens the reallocation of resources toward more productive units, and therefore temporarily decreases aggregate productivity. There are many other channels through which uncertainty might dampen the economy. Nicholas Bloom (2014) provides an overview of the literature.

Bloom (2014) argues that there is no perfect measure of uncertainty, but only a range of proxies. The most common proxy in the uncertainty literature is stock market volatility. Stock market volatility is an uncertainty proxy, because the more volatile the series, the harder it is to forecast. Bloom (2009) shows with regression analysis that a number of cross-sectional proxies such as the variance of firm-level shocks, or dispersion across macro forecasters are highly correlated with stock-market volatility. There is a large body of literature discussing alternative proxies for uncertainty.

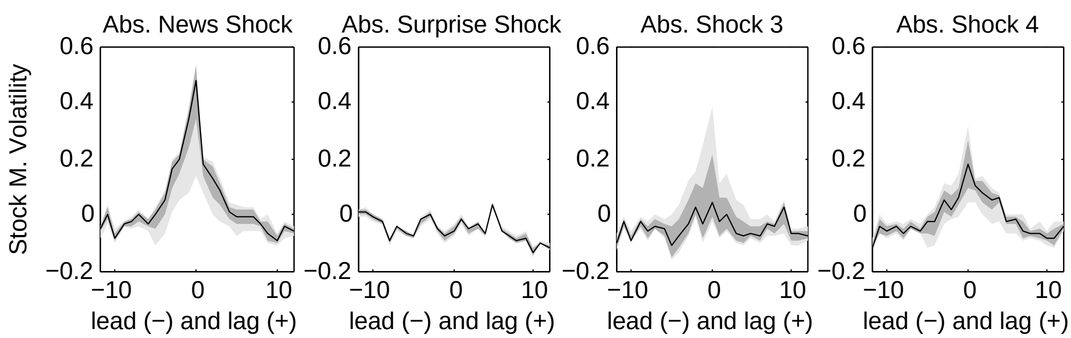
The closest proxy to Knight's (1921) definition of uncertainty is Jurado, Ludvigson and Ng's (2015) [JLN] measure. They approximate uncertainty based on hundreds of macroeconomic and financial indicators. Uncertainty is measured as the conditional volatility of the purely unforecastable component of future movements. Bachmann, Elstner and Sims (2013) [BES] use survey dispersion as proxy for uncertainty, which captures the level of disagreement amongst respondents of the Business Outlook Survey. Baker, Bloom and Davis (2015) [BBD] measure policy uncertainty based on the frequency of certain words written in news reports. Leduc and Liu (2015) [LL] measure perceived uncertainty of consumers based on the Michigan Consumer Survey: uncertainty is the share of consumers that state uncertainty is the reason why they don't buy a car. An overview of the different proxies are provided by Bloom (2014).

Figure 2.2 plots stock market volatility and the four proxies for uncertainty: JLN, BES, BBD, and LL. The correlation among the proxies are between 0.19 and 0.52, and the new stylized fact holds for all proxies: correlations of the the different proxies to the magnitude of a news shock are between 0.08 and 0.52.

### 2.2.3 News Shock and Uncertainty

So far, a relationship between uncertainty and the absolute value of news shocks is observed in figures 2.1 and 2.2. Figure 2.4 plots the cross-correlations of stock market volatility to the absolute value of a news shock, and to the other three structural shocks in Barsky and Sims' (2011) model: surprise technology shock, and the remaining shocks 3 and 4.

Figure 2.4: Cross-Correlations of News Shocks and Stock Market Volatility




*Notes:* This figure plots cross-correlations of stock market volatility (**Stock M. Volatility**) with the absolute value (**Abs.**) of structural shocks identified with Barsky and Sims' (2011) model. **News shock** is the only shock orthogonal to current TFP that maximizes the share of future TFP movements of the following 40 quarters. **Surprise shocks** is the shock that maximizes the share of future TFP movements, but is not orthogonal to current TFP. **Shocks 3 and 4** are the remaining shocks that contribute little to TFP movements. Barsky and Sims' (2011) VAR(3) model is estimated in levels, with Fernald's (2012) adjusted TFP, Shiller's (2015) real stock prices, per capita real GDP (source: FRED) and per capital real consumption expenditures (source: FRED) from 1950Q1-2015Q3. Variables are in logs and the quantity series are seasonally adjusted. The bootstrapped confidence bands show the $68^{th}$ and $90^{th}$percentiles.

The first panel of figure 2.4 shows cross-correlations of stock market volatility to the absolute value of a news shock. The correlations are positive for current stock market volatility to past and future magnitudes of news shocks. The correlations are significant for the previous and following two quarters. The graph neither points towards a lead nor a lag behaviour of stock market volatility. Hence, a news shock changes today's expectations, which either affect uncertainty or stock prices first. If they affect stock prices first, then news shocks, which are identified with stock prices, move first. If expectations affect stock market volatility first, the uncertainty proxy moves first. This is an important property for the identification of uncertainty shocks, which is discussed in section 2.3.

The remaining panels of figure 2.4 plot the correlations of stock market volatility to the magnitude of the other three shocks. There is either no correlation or significantly smaller correlation compared to the correlation to news shocks. This is evidence that the correlation to stock market volatility is a specific feature of a news shock.

The following sections examine implications of the new stylized fact for the literature on macro uncertainty (section 2.3), the finance literature (section 2.4), and the literature on news shocks (section 2.5).

## 2.3 News Shock and the Effect of Uncertainty

As shown in section 2.2, good and bad news shocks coincides with higher uncertainty. This is intuitive, considering the news shock as a shock on expectations about the future. If the expected future suddenly changes due to a news shock, it is reasonable that the ability to anticipate the "new" future changes, as well. The close relationship between news shocks and uncertainty challenges the identification of uncertainty shocks in VAR models: both, a bad news shock and an uncertainty shock increase uncertainty and slow down the economy. This section shows that the drop in output after spikes in uncertainty is mainly driven by bad news shocks, whereas there is little role for uncertainty shocks.

Bloom (2009) estimates a structural vector autoregression (SVAR) model with macro variables and a dummy variable capturing 17 uncertainty events from 1962 to 2008. An uncertainty event occurs if stock-market volatility exceeds 1.65 standard deviations of the Robert J Hodrick and Edward C Prescott (1997) detrended mean of stock-market volatility. He finds that an uncertainty shock is followed by substantial drop in output and employment over the next 6 months. The uncertainty shock is identified as the shock that leaves stock prices unchanged on impact, but it is allowed to shift all other macro variables.

Bloom (2009) and other empirical studies on macro uncertainty apply a Cholesky decomposition of the impact matrix. Bloom's (2009) identification is executed by placing the proxy for uncertainty second after stock prices. Other studies place the proxy either first before stock prices, second after financial variables, or last.[13]

In contrast to Bloom's (2009) uncertainty shock, a bad news shock shifts both, uncertainty and stock prices on impact. Indeed, impulse response functions of the previous section (see figure 2.3) show that a bad news shock decreases stock prices on impact. However, stock prices drop

[13] Jurado, Ludvigson and Ng (2015) place the proxy last, Simon Gilchrist, Jae W. Sim and Egon Zakrajšek (2014) place the proxy after both, credit spreads and short-term interest rates in their first model and after short-term interest rates in their second model. Bachmann, Elstner and Sims (2013), Baker, Bloom and Davis (2015), and Leduc and Liu (2015) place the proxy first before stock prices, which makes the separation of uncertainty and news shocks even more difficult. Dario Caldara, Cristina Fuentes-Albero, Simon Gilchrist and Egon Zakrajšek (2016) do not apply Cholesky decomposition, instead, they use a criterion that the uncertainty shock maximizes the impulse response of the uncertainty proxy. They implement this criterion with a penalty function approach developed by Jon Faust (1998) and Harald Uhlig (2005). Since financial conditions can react immediately to uncertainty shocks, the procedure does not seem to be able to separate news shocks from uncertainty shocks, either.

again in the following quarters, before it increases in the long-run. The second drop might be explained by increased uncertainty that initially reduces the negative effect of a bad news shock. When uncertainty decreases in the following quarters, the anticipation of the bad future is improved so that stock prices decrease again. A spike in uncertainty is therefore followed by a second drop in stock prices, output and consumption. The model of Bloom (2009) is not able to separate this second drop caused by news shocks from uncertainty shocks.

Another issue is that a bad news shock might initially increase uncertainty, before it shifts stock prices, output and consumption. Since news shocks are identified based on the latter three series, the identified news shock might occur after the spike in uncertainty. This theory is consistent with the findings of the previous section 2.2.3 that there is no clear lead-lag behaviour between news shocks and uncertainty. Hence, impact restrictions are not enough to separate news shocks from uncertainty shocks.

Since impact restrictions are not enough, other restrictions are required to separate the two shocks. Fortunately, the standard identification procedure for news shocks allows to separate news shocks from uncertainty shocks. A news shock is identified as a shock that moves TFP in the long-run. This separates news shocks from uncertainty shocks, because the effect of uncertainty on TFP is only temporary. The real option effect of uncertainty discussed in section 2.2.2 states that increased uncertainty dampens the reallocation of resources in the economy. This includes the reallocation of less productive toward more productive firms. When uncertainty is resolved, reallocation is increased which causes a rebound and an increase in productivity. News shocks, on the other hand, shift technology in the long run by definition. Moreover, the common restriction that a news shock is orthogonal to current TFP further supports the separation. Uncertainty shifts TFP only while uncertainty is high, and therefore only in the very short run.

In order to control for news shocks, good and bad news shocks are added to Bloom's (2009) structural VAR as exogenous variables. The two series are generated by setting the negative values, respectively, the positive values equal to zero. The remaining variables are the same as in Bloom (2009), except that the series are quarterly instead of monthly, because there is no monthly adjusted TFP measure.[14] Moreover, the series cover a longer time period until 2015. Hence, the model's uncertainty dummy includes another spike: the shock of the European Debt Crisis in 2011.

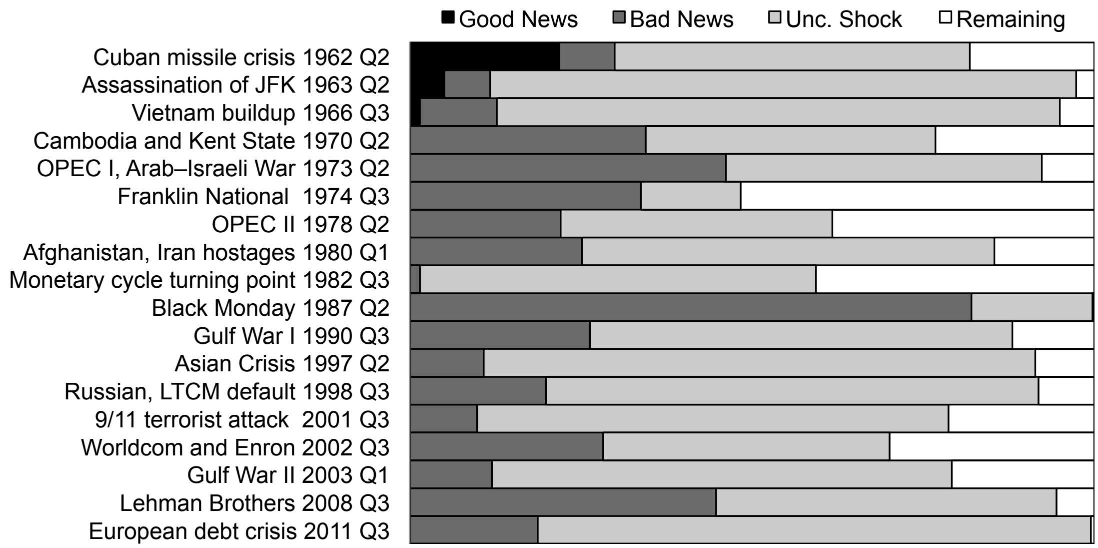
Figure 2.5 decomposes the 18 uncertainty events into good news shocks, bad news shocks, uncertainty shocks, and the sum of the remaining shocks. Bad news shocks make up a significant fraction of many of the uncertainty events. The largest contribution of news shocks are observed in the events: OPEC I, Black Monday, and the bankruptcy of Lehman Brothers. Good news shocks only contribute to the first three uncertainty events.

The only good event according to Bloom (2009) is the uncertainty spike at the monetary cycle turning point in 1982. And indeed, the contribution of bad news shocks is the lowest at this event. Good news shocks do not explain this spike, either. This might be because the good news about monetary policy only boosts demand, while leaving TFP unchanged.

Figure 2.6 takes a look at the cyclical component of industrial production before and after the 18 uncertainty events. The dashed line plots average (HP-filtered) industrial production before and after news-driven uncertainty events. News-driven uncertainty events are events, where bad news shocks are responsible for more than 30% of the dummy. The continuous line shows the average evolution before and after the remaining events. Both lines are normalized to zero at the event.

[14]Data is listed in the footnote of figure 2.5.

Figure 2.5: Contribution to Uncertainty Events

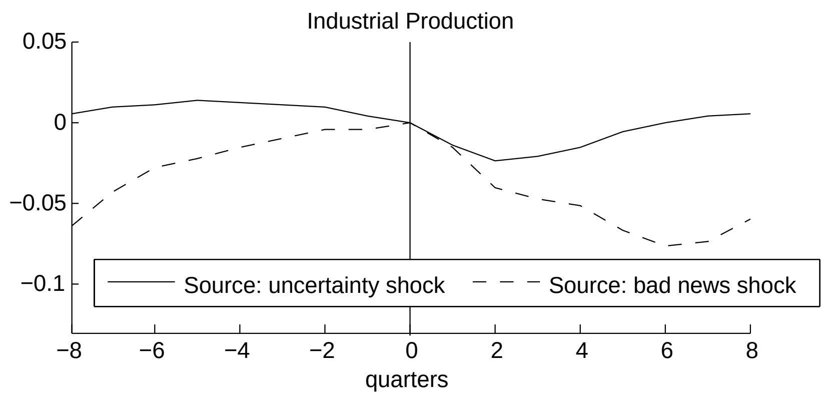



*Notes:*

This figure decomposes 18 uncertainty events form 1960Q1-2015Q1 into **good news shocks** (black), **bad news shocks** (dark grey), **uncertainty shocks** (light grey), and **remaining shocks** (white). A Bloom (2009) uncertainty event occurs if stock-market volatility exceeds 1.65 standard deviations of the Hodrick and Prescott (1997) detrended mean of the stock-market volatility series. The variables of the underlying VAR in the estimation order are log(S&P500 stock market index), a dummy variable indicating the 18 uncertainty events, Federal Funds Rate, log(average hourly earnings), log(consumer price index), hours, log(employment), and log(industrial production). All variables are Hodrick and Prescott (1997) detrended as in Bloom (2009). The variables are aggregated to quarterly series in order to include good and bad news shocks, which enter the system exogenously. The news shocks are estimated with Barsky and Sims' (2011) VAR(3) model in levels, with the logs of Fernald's (2012) adjusted TFP, Shiller's (2015) real stock prices, per capita real GDP and per capital real consumption expenditures, where news shock is the only shock orthogonal to current TFP that maximizes the share of TFP movements of the following 40 quarters.

The drop in industrial production is more significant when uncertainty is caused by bad news shocks. When uncertainty is exogenous, industrial production only decreases by little. The impulse response functions (IRFs) to news and uncertainty shocks confirm this finding (see figure 2.8 in appendix): the dynamics of uncertainty and bad news shocks are similar, but the magnitudes

Figure 2.6: Path Before and After Uncertainty Events




*Notes:* This figure plots the average path of industrial production of the quarters before (-4, -2, ...) and after (2, 4, ...) the 18 uncertainty events listed in figure 2.5. The dashed line plots the average path of events, where bad news shocks contribute more than 30% to the dummy: "Cambodia and Kent State", "OPEC I Arab-Israeli War", "Franklin National", "Black Monday", and "Lehman Brothers". The continuous line shows the average evolution before and after the remaining events. Both lines are normalized to zero at the event. The industrial production series is in logs and HP-filtered.

of the responses to uncertainty shocks are smaller and insignificant. A more detailed analysis of IRFs to good and bad news shocks is conducted in section 2.5.

Moreover, industrial production seems to increase before news-driven uncertainty spikes. Hence, the economy experiences a boom before a bad news shock causes a spike in uncertainty. Not so when uncertainty is caused exogenously. This may serve as a property that distinguishes uncertainty spikes driven by bad news shocks from exogenous uncertainty events.

To conclude, spikes in uncertainty are followed by a significant drop in output and other real variables. These dynamics are however rather driven by bad news shocks than by uncertainty itself. In the next section, further implications of the new stylized fact are discussed in the literature on finance.

## 2.4 News Shock and the Stock Returns to Volatility Relationship

A large body of literature in Finance deals with the relationship of stock returns to its volatility. The analysis are on different levels of aggregation: research has shown that the relationship is not only strong in a micro setting, but also in a macro time series framework. This section analyses the role of news shocks in the stock return to volatility relationship.

The stock returns to volatility relationship has a long history in Finance. Harry Markowitz (1952) remarks that investors consider not only expected returns, but also the variance of returns. This idea is incorporated in risk-averse utility functions by John W. Pratt (1964) and K.J. Arrow (1965). Due to risk averse investors, the market requires expected returns to be higher when volatility is high as shown by William F. Sharpe (1964) and John Lintner (1965), which is the cornerstone of the capital asset pricing model (CAPM). Asset pricing models therefore predict a positive relation between excess returns and ex ante volatility.

Pindyck (1984) observes a strong negative relationship between volatility and expected returns during the 1970s. He argues that this is due to the increase in risk premia caused by higher volatility. This does not contradict the findings of asset pricing models, when the increase in volatility is unexpected: an unexpected increase in volatility raises the required expected returns on stocks due to risk averse investors. Holding everything else constant, expected returns can only increase if current stock price drop. Hence, there is a negative relationship between current returns and unexpected volatility. This *volatility feedback effect* is first modelled in French, Schwert and Stambaugh (1987). Another theory explaining the negative relationship is the *leverage effect*: when stock prices drop, firms' leverage ratio increases, which increases risks and therefore stock market volatility (see Black, 1976; Christie, 1982; Schwert, 1989). Bekaert and Wu (2000) and Wu (2001) address the two effects simultaneously, and they find that the *volatility feedback effect* dominates the *leverage effect*.

Schwert (1989) supports the empirical evidence by Pindyck (1984) and French, Schwert and Stambaugh (1987): market volatility increases in recessions. Hany A. Shawky and Achla Marathe (1995) find that there is no negative relationship between excess returns and market volatility in periods where market prices are rising. Chia-Shang James Chu, Gary J. Santoni and Tung Liu (1996) find that volatility is higher when returns are either above or below normal. They argue that the inverse relationship that others have found is because the increase in volatility is larger when returns drop than when they increase. Hence, when a straight line is forced through the data the relationship is inverse, but in fact the relationship is V-shaped.

Estimates of table 2.1 confirm Chu, Santoni and Liu's (1996) findings with a much less sophisticated model: stock market volatility is regressed on stock returns, a constant, and a trend

Table 2.1: Regression Analysis Stock Returns and Stock Market Volatility

| | Stock Market Volatility | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Stock Returns | -0.460** | | | |
| | [0.114] | | | |
| Positive Stock Returns (S.R.) | | 0.351** | | |
| | | [0.095] | | |
| Negative Stock Returns (S.R.) | | -1.149** | | |
| | | [0.122] | | |
| News Shock Driven S.R. | | | -0.470** | |
| | | | [0.123] | |
| S.R. w/o News Shocks | | | -0.415* | |
| | | | [0.136] | |
| Positive News Shock Driven S.R. | | | | **0.533** ** |
| | | | | [0.125] |
| Negative News Shock Driven S.R. | | | | **-1.133** ** |
| | | | | [0.139] |
| Positive S.R. w/o News Shocks | | | | **-0.160** |
| | | | | [0.195] |
| Negative S.R. w/o News Shocks | | | | **-0.667** |
| | | | | [0.283] |
| Constant | 0.168** | 0.133** | 0.168** | 0.120** |
| | [0.005] | [0.006] | [0.005] | [0.008] |
| Trend | 0.019** | 0.019** | 0.019** | 0.025** |
| | [0.004] | [0.004] | [0.005] | [0.004] |
| Observations | 265 | 265 | 265 | 265 |
| Period | 1948-2016 | 1948-2016 | 1948-2016 | 1948-2016 |
| Adjusted R Square | 0.249 | 0.476 | 0.246 | 0.453 |

*Notes:* Robust standard errors in brackets, small P-values marked as *p<0.01, and **p<0.001.

over the period of 1950Q1-2016Q1. The significant parameter of negative 0.460 of the first column suggests that an increase in stock returns is associated with a significant decrease in stock market volatility. Then, stock returns are divided into two series: positive stock returns, where negative values are set equal to zero, and similarly, negative stock returns. The regression analysis of the second column shows that an increase as well as a decrease in stock returns is associated with higher volatility, and the latter relationship is three times stronger.

In order to examine the role of news shocks in this relationship, I construct two counterfactuals, based on the estimated model of section 2.2.1.[15] The first counterfactual re-runs history by setting all shocks except news shocks equal to zero. The second counterfactual sets news shocks equal to zero, and leaves the other shocks unchanged. Hypothetical stock returns are then calculated based on these counterfactuals so that the two series add up to stock returns.

The estimates of the last column of table 2.1 imply that the V-shaped relationship between stock returns and stock market volatility is driven by news shocks. On the one hand, the coefficients on news shock-driven positive and negative stock returns are significant and suggest a V-shaped connection. On the other hand, the coefficients on the positive and negative stock returns without news shocks are insignificant. Hence, the regression analyses suggest that the volatility feedback and leverage effects only occur when stock returns are driven by news shocks, or that there are alternative channels from news shocks to volatility.

A possible channel comes from the arrival of vague information associated with news shocks. Traditional technology shocks shift TFP first, which then causes a response of other macro variables. News shocks, on the other hand, shift the fundamental variable only in the future so that the initial response is only due to altered expectations, and not based on an observed change. The accuracy of information is therefore much lower with news shocks than with traditional shocks. Consequently, a jump in stock returns is likely to increase uncertainty more if it is based on news, than if it is caused by an observed change in TFP.

[15] A structural VAR model with Barsky and Sims's (2011) identification procedure, see section 2.2.1 and figure 2.3 for details.

To conclude, news shocks seem to be the main cause for the link between stock returns and stock market volatility in a quarterly time series framework. Moreover, the section confirms the findings of the previous sections that there is a non-linear relationship between news shocks and uncertainty. In the next section, an alternative news shock is estimated in a model that allows for the discovered asymmetries.

## 2.5 Identification of News Shocks with a Non-Linear Model

The analysis so far is based on news shocks identified with linear models. In this section, a non-linear structural VAR is estimated. The model allows for asymmetric responses to good and bad news shocks. The impulse response functions (IRFs) of the two shocks are then compared.

The model is estimated in two steps. First, a core model is estimated with total factor productivit (TFP), and three forward looking variables: stock prices, consumer sentiment, and the one quarter forecast of output by the Federal Reserve. The aim of the core model is to correctly identify the news shock without imposing symmetry. The model is a regime-switching VAR[16]. The parameters depend on whether previous changes in consumer sentiment are positive or negative. The regime is function of consumer sentiment, because it is the forward-looking variable that responds the most to news shocks, a feature observed by Barsky and Sims (2012) and Barsky, Basu and Lee (2015). A good news shock is therefore allowed to differ from a bad news shock through its impact on consumer sentiment. Since long-run fluctuations matter in order to correctly identify news shocks, it is assumed that stock prices are cointegrated with TFP as in Beaudry and Portier (2006). The cointegration relationship does not depend on regimes.

[16]see Hamilton (1994) for regime-switching VAR models.

A drawback of regime-switching models is that asymmetries are only allowed to occur with a lag. Hence, regimes cannot depend on current changes, only on past changes in consumer sentiment. The impact response to news shocks is however non-linear, as shown with the new stylized fact: both good and bad news shocks coincide with higher uncertainty. This is the reason why the model is estimated in two steps. In the second step, the identified news shocks of the first model enter a second VAR exogenously, together with a range of macro variables. The news shocks are divided into good news and bad news shocks by setting the opposite entries equal to zero. The macro variables are therefore allowed to respond differently to good and bad news shocks on impact.

The identification of news shocks non-trivial, because the path of regimes induced by a shock depends on the sign and size of a shock. A new identification procedure is introduced that identifies structural shocks with finite horizon restrictions for non-linear models. This allows to identify an average news shock with the same identifying assumptions as previous studies: a news shock is orthogonal to current TFP and maximizes the forecast error variance of future TFP.

Identification is based on Barsky, Basu and Lee (2015): a news shock is the only shock orthogonal to current TFP that maximizes the share of TFP movements at the $20^{th}$ quarter after impact[17]. However, instead of maximizing the forecast error variance of a one unit increase, the integral of the forecast error variances over a normal distribution is maximized. The identifying restrictions are therefore based on the average response to news shocks.

Since the integral is too complex to solve analytically, it is calculated numerically. A possibility is to apply Monte Carlo integration. This method calculates the integral as the sum of

[17]The identification procedure of maximizing the forecast error variance is introduced by Francis et al.'s (2014).

the functions produced with simulated shocks. Computationally, this method is inefficient. Instead, I integrate over a sparse grid proposed by Florian Heiss and Viktor Winschel (2006). The sparse grid approximates a Gaussian distribution with an optimized range of values and weights for each value. The finite horizon long-run restrictions on the integral are then imposed by rotating the impact matrix using numerical maximization following Benati (2016): the angles of six rotation matrices are optimized using the medium-scale quasi-Newton line search of Matlab's *fminunc* function.

In the second step, a VAR model is estimated with exogenous good and bad news shocks from the core model. The VAR also includes eight additional macro variables, thus, the model consists of 12 variables (4+8). The model is estimated in levels. The VAR is estimated 1,000 times for each of the bootstrapped news shocks of the core model. Moreover, for each of the VAR models, 10,000 additional bootstrapped models are estimated so that the percentiles of the impulse response functions are picked from ten million models ($1,000 \times 10,000$). The VARs of the first and second step both account for four lags.

The VAR is estimated over the 1967Q1-2009Q1 period, which is the period where the forecast data of the Federal Reserve is available. The following U.S. data is used: log of Fernald's (2012) adjusted TFP, log of Shiller's (2015) real stock prices, changes in consumer sentiment of the Michigan Consumer Survey, the Federal Reserve's one quarter ahead projection of real GDP growth (Greenbook), and stock market volatility (the measure is discussed in figure 2.1). The remaining series are collected from the database of the Federal Reserve Bank of St. Louis (FRED): three months treasury bill rate, CPI inflation, real GDP, real consumption of non-durables, real consumption of durables, real investment, and total hours in the non-farm sector. The quantity variables are seasonally adjusted and per capita, divided by the civilian, non-institutional population and in logs.

Interestingly, the news shock series produced here does not differ significantly from traditional news shocks: the correlation to the news shock with Barsky and Sims's (2011) identification is 0.81.

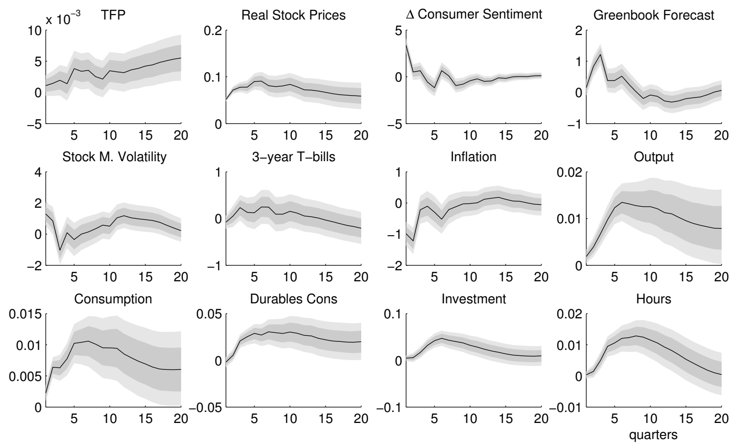
Figure 2.7 plots the impulse response functions to an exogenous shift in good and bad news shocks. Four characteristics attract my attention. First, stock market volatility spikes after both, good and bad news shocks. However, the response of stock market volatility to a bad news shock is five times stronger. This is consistent with the new stylized fact and confirms the correlation analysis of the previous sections.

Second, the new stylized fact does not seem to significantly affect the responses of the quantity variables: the responses of output, consumption, investment, and hours are not significantly different between a good and a negative bad news shock. If uncertainty causes the economy to slow down as suggested by Bloom (2009) and others, one would expect the responses to differ: uncertainty should dampen the effect of a good news shock and enhances the effect of a bad news shock. If anything, then the opposite is true: the effect of a good news shock on output is more significant than the effect of a bad news shock. However, the weaker response of bad news shocks could also be explained with the noisiness of the signal, as discussed next.

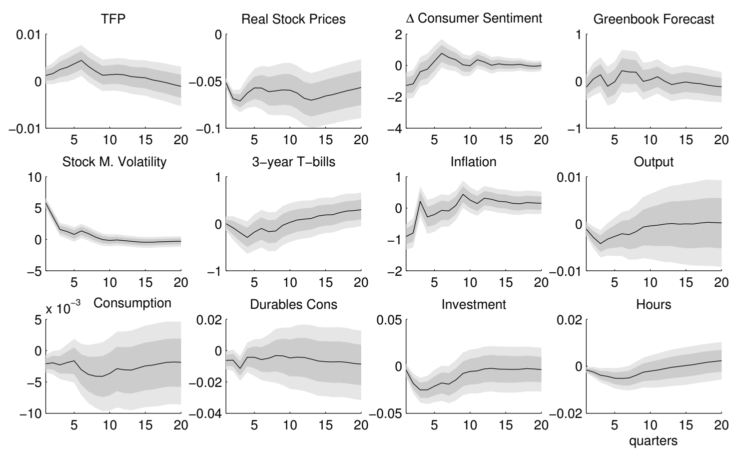
Third, the responses to bad news shocks have larger confidence bands. This can be explained by the increase in stock market volatility on impact, which is five times stronger than the increase after good news shocks. A bad news shock is a signal of a future decrease in TFP, and a good news shock is a signal of a future increase. The uncertainty responses suggest that the signal of the bad news shock is noisier than the signal of a good news shock. Hence, the anticipation of

Figure 2.7: IRFs Non-Linear Model

IRFs to Good News Shock




IRFs to Bad News Shock




*Notes:* This figure plots impulse response functions (IRFs) to a good news shock (upper panel) and a bad news shock (lower panel). The model is described in section 2.5. Responses are plotted for log of Fernald's (2012) adjusted TFP, log of Shiller's (2015) real stock prices, changes in consumer sentiment of the Michigan Consumer Survey, the Federal Reserve's one quarter ahead projection of real GDP growth (Greenbook), stock market volatility (the measure is discussed in figure 2.1), three months treasury bill rate, CPI inflation = 400× changes in quarterly log(CPI), log of per capita real GDP, log of per capita real consumption of non-durables, log of per capita real consumption of durables, log of per capita real investment, and log of total hours in the non-farm sector. The bootstrapped confidence bands show the $68^{th}$ and $90^{th}$percentiles.

the future is worse when a bad news shock hits the economy. The responses to bad news shocks are therefore contaminated with more noise than good news shocks, and confidence bands are thus larger. A lack of anticipation is also apparent in the response of the Greenbook forecast: the good news shock changes the forecast significantly, while the bad news shock doesn't.

Fourth, the response of inflation differs significantly. A good news shock causes output and inflation to move in the opposite directions, whereas a bad news shock causes co-movement. A bad news shock has the characteristics of a demand shock. This is consistent with the original idea of the news shock literature: today's anticipation increases demand, even though productivity only increases in the future. The response of inflation to a good news shock is consistent with Barsky, Basu and Lee's (2015) work. They find that inflation declines significantly after a new shock, and they rationalize this result with a New Keynesian model with wage and price rigidities. The decline of future marginal costs causes firms to change wages and prices today, because they might not get the chance to change prices in the future, given price and wage rigidities.

The large uncertainty spike caused by bad news shocks is a possible explanation for the different responses of inflation: the anticipation of a bad news shock is not good enough for firms to change price and wages today so that the price response of a bad news shock is purely demand driven.

## 2.6 Conclusion

Good news shocks and especially bad news shocks increase uncertainty. This new stylized fact combines three fields dealing with anticipation: research on news shocks, research on macro uncertainty, and finance.

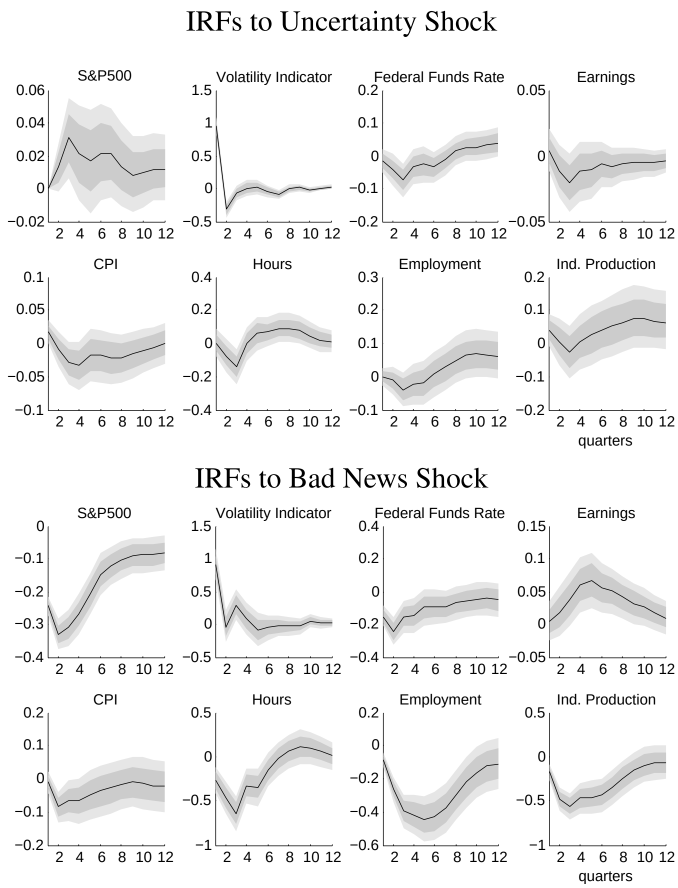
Structural VAR analyses show that bad news shocks cause spikes in macro uncertainty. These spikes are followed by significant drops in real macro variables. Exogenous spikes in uncertainty have little effects.

Times series and regression analyses of quarterly macro series suggest that the close relationship between stock returns and volatility is driven by news shocks.

A non-linear VAR model shows that bad news shocks increase stock market volatility - a proxy for uncertainty - five times more than good news shocks. This could be the source for two additional asymmetries. First, bad news shocks produce larger confidence bands. Second, while good news shocks move inflation and output in the opposite direction, bad news shocks cause a co-movement, because the spike in uncertainty impairs the anticipation of future marginal costs.

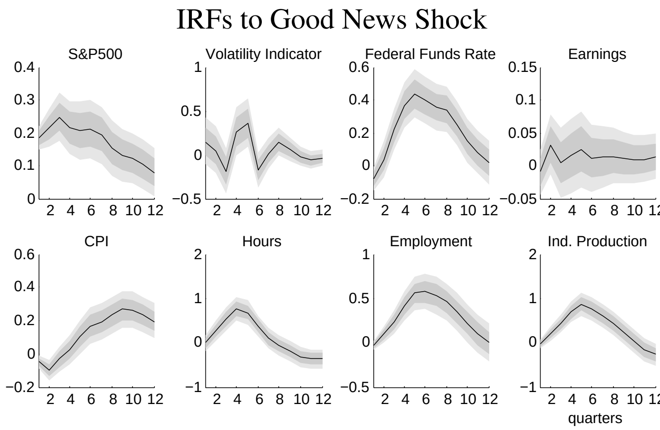
## 2.7 Additional Graphs

Figure 2.8: IRFs Augmented Version of Bloom's (2009) VAR Model (1/2)




(see figure notes on the next page)

Figure 2.8: IRFs Augmented Version of Bloom's (2009) VAR Model (2/2)




*Notes:* This figure shows the impulse response functions (IRFs) to an uncertainty shock (top), a bad news shock (middle), and a good news shock (bottom). The underlying model is Bloom's (2009) VAR model augmented with exogenous good and bad news shocks. The **Volatility Indicator** is a dummy variable that indicates an uncertainty event, which occurs if stock-market volatility exceeds 1.65 standard deviations of the Hodrick and Prescott (1997) detrended mean of the stock-market volatility series. The variables of the underlying VAR in the estimation order are log(S&P500 stock market index) [**S&P500**], volatility indicator, **Federal Funds Rate**, log(average hourly earnings) [**Earnings**], log(consumer price index) [**CPI**], **Hours**, log(employment) [**Employment**], and log(industrial production) [**Ind. Production**]. All variables are Hodrick and Prescott (1997) detrended as in Bloom (2009). The variables are aggregated to quarterly series in order to include good and bad news shocks, which enter the system exogenously. The VAR is estimated over the 1960Q1-2015Q1 period, and includes four lags. The news shocks are estimated with Barsky and Sims' (2011) VAR(3) model in levels, with the logs of Fernald's (2012) adjusted TFP, Shiller's (2015) real stock prices, per capita real GDP and per capital real consumption expenditures, where news shock is the only shock orthogonal to current TFP that maximizes the share of TFP movements of the following 40 quarters. The bootstrapped confidence bands show the $68^{th}$ and $90^{th}$percentiles.

# Chapter 3

# Identification of Treatment Effects when Treatment is an Aggregate Variable

## 3.1 Introduction

Business cycle economists examine the booms and the busts in the economy, and elaborate on how the government can mitigate these business fluctuations. The ups and downs in economic growth are caused by economic shocks such as new tariffs that reduce trade, new ideas that boost productivity, or sudden spreads of fear that make people spend less money. Whether these shocks produce large or small fluctuations in economic growth depends on the economic system that is in place; for example, on how the central bank responds to new tariffs, on how the law protects new ideas, or on how fast firms adjust prices when people reduce spending. Economic shocks and their effects are thus informative about both, the drivers and the characteristics of an economy. The challenge is to identify these economic shocks and to estimate their effects; in particular, it is difficult to separate economic shocks from each other, because each period is exposed to a large number of shocks that hit the economy at the same time. I introduce a framework that identifies the *average* effect of a group of shocks, which is a compromise to measuring the effect of each shock separately. For example, if interested in productivity shocks, I show how the average effect of a variety of productivity shocks can be identified without the need of defining and separating the different types of productivity shocks.

Textbooks and research papers usually treat groups of shocks as homogeneous disturbances, as if there is only one type of productivity, preference, or government expenditure shock. For example, a rise in government spending due to a new military threat is assumed to have the exact same effect on the economy than the same rise in spending caused by a new infrastructure project; in other words, it is assumed that all government expenditure shocks have the same effect up to scale. One exception is Blanchard and Quah (1989), who not only discuss how to separate demand from supply shocks, but they also lay out the necessary assumptions for their method to work when demand and supply shocks are combinations of shocks with heterogeneous effects. I make this aggregation more explicit by writing a model that consists of a large number of disaggregate shocks, before I compile the model into an empirical model with aggregate shocks. This procedure brings to light that effects of aggregate shocks change over time so that the model parameters are time-varying. The effects change over time because the response to a government expenditure shock, for example, depends on how much of this shock is driven by a military spending shock that has a different effect than let's say an infrastructure shock. The effect of the aggregate shock thus depends on the realizations of the disaggregate shocks at each point in time. This has implications on the common estimation techniques, which usually rely on the assumption of constant parameters.

This paper shows what standard identification strategies with constant parameters identify, when the estimated model is in fact an aggregation of a larger model. It turns out that under certain conditions, the constant parameter models identify a weighted average of all the different effects of the disaggregate shocks, but the weights on the disaggregate effects depend on the identification strategy. This provides an explanation for why different empirical macro papers estimate different effects of the same shock, even when dealing with the same country and the same time period. Hence, instead of arguing that if two papers get different results then one of them must be misspecified, I provide a framework where both identification strategies may extract the shocks of interest correctly, but one identification strategy puts more weights on disaggregate shocks with stronger effects than the other.

Another feature the model brings to light is the bridge between identifying effects of policy interventions, and identifying effects of shocks to policy interventions, which may differ when model parameters are no longer constant. When evaluating a policy, the aim is not to measure the effects of policy shocks by themselves. Policy shocks are for example mistakes by policy makers that are exogenous to the rest of the economy, which are hopefully irrelevant for macro business cycle fluctuations. Instead, the aim is to identify effects of policy shocks to make inference on what the effect of a systematic policy response would be. This extrapolation is done frequently, because effects of policy shocks are equal to effects of policy responses in standard empirical models with constant parameters. This one-to-one extrapolation between effects of shocks and responses does not always make sense. For example, the effect of a random hospital visit is almost zero, while the effect of going to the hospital when having an emergency is large. Similarly, a manipulation of the interest rate by the monetary authority might have a different effect if the policy intervention is a response to an economic downturn, than when it is a random mistake. I show that specifying the treatment with multiple variables instead of just one variable helps to link effects of shocks to effects of responses. Hence, instead of a one-to-one extrapolation from effects of shocks to effects of responses, I discuss how to make the extrapolation more realistic using additional characteristics of the treatment of interest.

When defining objects of interest and identification strategies, I build on the literature dealing with heterogeneous treatment effects. The heterogeneous treatment effect literature allows individuals to experience different effects of the same treatment so that for example a hard-working person gains more from education than a person that enjoys leisure time. Donald B. Rubin (1974) defines the heterogeneous causal effect as the gap between an individual's outcome with and without treatment, and then he defines the average causal effect as the average gap across individuals. James J. Heckman and Richard Robb (1985) introduce the notion of average treatment effect on the treated (ATT), which only accounts for the heterogeneous treatment effects of individuals that experience treatment, and thus may differ from the average treatment effect (ATE), respectively, the average causal effect discussed in Rubin (1974). Guido W Imbens and Joshua D Angrist (1994) and Joshua D. Angrist, Guido W. Imbens and Donald B. Rubin (1996) show that under the assumption of heterogeneous treatment effects, instrumental variable (IV) estimates produce local average treatment effects (LATE), which again weighs individuals differently than the ATE. James J. Heckman and Edward J. Vytlacil (1999); James J. Heckman and Edward Vytlacil (2005); James J. Heckman and Edward J. Vytlacil (2007) define the marginal treatment effect (MTE) as the gain from treatment for individuals at the border between treatment and no treatment, and then show how ATE, ATT, and other summary treatment effects can be expressed in terms of MTEs. Quang Vuong and Haiqing Xu (2017) show how a counterfactual mapping between potential and actual outcome can identify individual treatment effects (ITE) directly. Note that instead of individuals, business cycle models deal with different time periods so that the individual treatment effect discussed here refers to the treatment effect of a specific time period, rather than a person.

Section 3.2 aggregates the data generating process into an empirical model with aggregate shocks. Section 3.3 uses the empirical model to identify objects of interest. Section 3.4 improves identification by combining different identification strategies. Section 3.5 discusses the identification of policy effects, and section 3.6 concludes.

## 3.2 Model

Consider the following model for output $y_i$ and reference variable $s_i$:

$$\begin{aligned} y_i &= \sum_{j=1}^{M} \phi_j \mu_{ji}, \\ s_i &= \sum_{j=1}^{M} \gamma_j \mu_{ji}, \end{aligned} \tag{3.1}$$

where the $M$ components are independent drivers with zero mean and unit variance, i.e. $E[\mu_{ji}| \mu_{1i}, ..., \mu_{j-1i},\ \mu_{j+1i}, ..., \mu_{Mi}] = E[\mu_{ji}] = 0$, and $E[(\mu_{ji})^2] = 1$, for all $j$. Note that this data generating process nests all linearised macro models that have a vector moving average (VMA) representation.[1]

The first $K < M$ components are the shocks of interest, which sign and size are normalized to the reference variable $s_i$; a positive shock thus increases the reference variable, and a negative shock decreases the reference variable so that $\gamma_j > 0$, for all $j \leq K$. An impulse response (IR) is defined as the response of outcome $y_i$ to an impulse of component $\mu_{ji}$ that corresponds to a unit increase in the reference variable:

$$IR_j \equiv \frac{\partial y_i/\partial \mu_{ji}}{\partial s_i/\partial \mu_{ji}} = \frac{\phi_j}{\gamma_j}, \quad \forall i,\ \forall j \leq K. \tag{3.2}$$

For example, the $IR_j$ of a military spending shock is the response in outcome ($\Delta y_i$) to a unit increase in total government expenditures ($\Delta s_i = 1$), caused by a new military threat ($\Delta \mu_{ji}$). An impulse response function (*IRF*) is then the set of impulse responses (*IR*s) where the reference variable stays the same, but the output variable changes from current output to future output; for instance, *IRF* is the combination of $IRs$ for all outputs $\{y_i^0, y_i^1, y_i^2, ...\}$, where $y_i^h \equiv GDP_{t+h}$, while the reference variable stays the same, e.g. $s_i \equiv G_t$.

Note that a component of interest $\mu_{ji}$ is defined as an independent event that has the same relative effect on output and reference variable across all observations $i$. It is thus very unlikely, for example, that there is only one government expenditure shock that has always the same relative effect on output and government expenditures. Even when reducing the focus on just defense

[1]In the VMA framework, $y_i$ would be one out of H variables of the vector, and the model's dependence on the past would imply $\mu_{ji} = \mu_{lk}$, for some $j \neq l$ and $i \neq k$, which does not violate the imposed assumptions.

spending shocks, there are defense expenditure shocks that are more labor heavy and others that are more capital intense so that the relative effect of output and government expenditures depends on the type of defense spending shock. In fact, it is likely that the number of components is infinitely large, i.e. $K \to \infty$. This means that the importance of a single component becomes infinitely small, i.e. $\gamma_j \to 0$. It thus makes sense to aggregate the components of interest into a treatment variable as follows:

$$d_i = \sum_{j=1}^{K} \gamma_j \mu_{ji}, \tag{3.3}$$

where the aggregation is based on the components' contribution to the reference variable $s_i$ in model (3.1). For example, let monetary policy shocks be the components of interest, and let the interest rate be the reference variable $s_i$. The treatment variable $d_i$ then captures the part of the interest rate that depends on current monetary policy components only, but not on other shocks such as supply shocks or past monetary policy shocks.

There are at least three reasons why the different components $\mu_{ji}$ need to be aggregated into a meaningful aggregate shock $d_i$. First, even if the components were observed, their variances would be too small to estimate significant effects; the variance of a sum of components, in contrast, is large enough. Second, even if components were observed and their effects identified, it is impossible to distinguish one component from another in an economically meaningful way. For example, there are infinitely many different government expenditure shocks so that it gets impossible to allocate each shock to a specific policy action. Third, the components themselves are unobserved and too small to be identifiable; it is however possible to identify a combination of components.

Consider the following empirical model:

$$y_i = f_i(d_i, u_i) = \beta_i d_i + u_i, \tag{3.4}$$

where $\beta_i d_i$ is the contribution of the components of interest, and residual $u_i$ aggregates the remaining components, i.e. $u_i = \Sigma_{j=K+1}^{M} \phi_j \mu_{ji}$. Parameter $\beta_i$ captures the effect of the treatment variable:

$$\beta_i = \frac{\sum_{j=1}^{K} \phi_j \mu_{ji}}{\sum_{j=1}^{K} \gamma_j \mu_{ji}}, \tag{3.5}$$

which varies across observations $i$. Hence, while the effect of a single component $\mu_i$ is constant across observations, i.e. $IR_j$ does not depend on $i$ in expression (3.2), the effect of the sum of components $d_i$ varies, i.e. parameter $\beta_i$ depends on $i$. This is because each observation is exposed to a different combination of disaggregate shocks so that some observations put more weight on components with strong effects, while other observations put more weight on components with weak effects. This is apparent when rewriting parameter $\beta_i$ in terms of the constant impulse responses ($IR_j$) defined in (3.2): (see derivation in appendix 3.7.1)

$$\beta_i = \sum_{j=1}^{K} IR_j \, \omega_{ji}^{ITE}, \quad \omega_{ji}^{ITE} = \frac{\gamma_j \mu_{ji}}{\sum_{j=1}^{K} \gamma_j \mu_{ji}}. \tag{3.6}$$

This expression shows how each observation puts different weights on the $K$ impulse responses so that each observation experiences a different treatment effect $\beta_i$. Note that if all impulse responses were the same, i.e. $IR_1 = ... = IR_K$, parameter $\beta_i$ would be constant across observations $i$, as the weights add up to one, i.e. $\Sigma_j^K \omega_{ji}^{ITE} = 1$, for all $i$. If the effects of the components differ significantly, however, then the treatment effects $\beta_i$ change significantly across observations $i$, as well. Note that this type of heterogeneity is not captured in time-varying parameter models of the macro literature, where parameters either have a linear representation such as an autoregressive form, or where parameters are regime-switching, changing form one value to another with a certain probability.

I build on the literature on heterogeneous treatment effects to define and identify different objectives of interest. One of the most informative object of interest is the individual treatment effect (*ITE*), $\beta_i$, which is the treatment effect that is experienced by a specific observation, i.e. by an individual, a time period, or a region. It is difficult to identify *ITE* for each observation so that econometricians often rely on summary treatment effects that aggregate over different parts of the population. Common summary treatment effects are the average treatment effect (*ATE*), and the average treatment effect on the treated (*ATT*). In addition to the standard summary statistics, I also make use of an object of interest that I call variance explained by treatment (*VET*). They are defined as follows:

$$\begin{aligned} ITE_i &\equiv \frac{\partial f_i(d_i + h, u_i)}{\partial h} = \beta_i \\ ATE &\equiv E[ITE_i] \\ ATT &\equiv E\left[ITE_i \frac{(d_i)^2}{E[(d_l)^2]}\right] \\ VET &\equiv Var\left(f_i(d_i, u_i) - f_i(0, u_i)\right). \end{aligned} \tag{3.7}$$

Appendix 3.7.2 shows that within model 3.4, $ITE_i$, $ATE$, and $ATT$ are consistent with the original definitions of the literature when treatment is a dummy variable. $ITE_i$ is the effect of a marginal change in treatment variable $d_i$ on an individual's outcome $y_i$, holding everything else constant. $ATE$ is the average of $ITE$s over the entire population. The average treatment effect on treated ($ATT$) is the $ATE$ of the individuals that are exposed to treatment. Since everyone is exposed to treatment when the treatment variable is continuous, $ATT$ is defined here as the weighted average of $ITE$s, where the weights reflect exposure to treatment $d_i$. The $VET$ is the variance of the dependent variable responsible by the treatment variable. In business cycle models it is common to calculate the forecast error variance (*FEV*), which is the $VET$ divided by the variance explained by both, the treatment variable and all other contemporaneous shocks.

Expression (3.6) rewrites the *ITE* in terms of the impulse responses defined in (3.2). Let's rewrite the remaining objects of interest of (3.7) in that way as well: (see derivation in appendix 3.7.3)

$$\begin{aligned} ATT &= \sum_{j=1}^{K} IR_j \, \omega_j^{ATT}, \quad \omega_j^{ATT} = \frac{(\gamma_j)^2}{\sum_{k=1}^{K} (\gamma_k)^2} \\ VET &= \sum_{j=1}^{K} (IR_j)^2 \; \omega_j^{ATT} \, Var(d_i) \; \geq \; (ATT)^2 \, Var(d_i). \end{aligned} \tag{3.8}$$

Expression (3.8) shows that $ATT$'s weights on the impulse responses $IR_j$ are all positive and reflect the contribution of the components to the variance of the treatment variable. $ATE$ does not collapse into a simple weighted average of impulse responses, but if the components are normally distributed then $ATT = ATE$ (see appendix 3.7.3). The $VET$ is the weighted average of the squared impulse responses multiplied by the treatment's variance. Appendix 3.7.3 shows that the squared $ATT$ multiplied by the treatment's variance provides a lower bound of the $VET$.

In the next section, I show how to identify $ATT$ and a lower bound of $VET$ with linear regressions.

## 3.3 Identification

The identification problem is that the treatment variable is unobserved. I discuss three methods that identify treatment effects with different set of assumptions. First, assumption A3.3 restricts the reference variable to be equal to the treatment variable so that identification of *ATT* is possible with observables $s_i$ and $y_i$ only. Second, assumptions A3.3, A3.3, A3.3, A3.3 and A3.3 restrict an instrumental variable (IV) $z_i$ to be correlated with the treatment variable but independent of the error term so that an IV-regression identifies *LATT*, which is the *ATT* but with different weights. Third, assumptions A3.3, A3.3, A3.3, A3.3, and A3.3 limit control variables (CV) in a way so that they control for some of the components, but don't interfere with the components of interest so that *ATT* is identified.

Model (3.4) does not satisfy some of the standard assumptions imposed in the heterogeneous treatment effect literature. The literature usually relies on the assumption that there is variation in the treatment variable that is independent from the rest of the model, i.e. there exists $\eta_i$ so that $corr(\eta_i, d_i) \neq 0$ and $f_i(h, u_i) \perp\!\!\!\perp \eta_i$ for all $h$. Note that the independence assumption requires not only the error term to be independent, i.e. $u_i \perp\!\!\!\perp \eta_i$, but also the treatment effect, i.e. $\beta_i \perp\!\!\!\perp \eta_i$. In the model presented here, there is no variation in the treatment variable that is independent of the treatment effect so that this assumption cannot be imposed here. Specifically, the treatment effect is a function of the treatment variable's decomposition, which is the very reason why the effect is heterogeneous (see equation 3.5). In this section I show that independence with respect to the error term is enough to identify $ATT$. I use the notation *partially* random and *partially* exclusive to refer to the weaker assumptions imposed here.

**A3.3** (No Measurement Error)**.** *Treatment $d_i$ is observed without measurement error. In particular, $s_i = d_i$ so that $\gamma_j = 0$ for all $j > K$ in (3.1).*

Assumption A3.3 states that the treatment variable is observed. Note however that the selection of treatment $d_i$ is only partially random, i.e. $f_i(h, u_i) \perp\!\!\!\perp d_i$ for $h = 0$ only. Hence, identification of *ATE* requires additional assumptions, while *ATT* is identified:

**Proposition 3.3.1** (Identification of ATT)**.** *Under the assumption of no measurement error A3.3, the ordinary least square (OLS) regression of the dependent variable* $y_i$ *on the reference variable* $s_i$ *identifies the average treatment effect on treated (ATT).*

**Proof**: See appendix 3.7.4.

Proposition 3.3.1 shows that we can estimate ATT with a linear regression, which according to appendix 3.7.3, also identifies $ATE = ATT$ when components are normally distributed.

Next, suppose assumption A3.3 does not hold so that the treatment variable or some of its variation needs to be extracted first. Let's assume the following IV is observed:

$$z_i = \sum_{j=1}^{M} \lambda_j \mu_{ji}, \tag{3.9}$$

which satisfies the following assumptions:

**A3.3** (Partial Exclusion)**.** *Instrument* $z_i$ *is partially random, i.e.* $f_i(h, u_i) \perp\!\!\!\perp z_i$ *for* $h = 0$*. In particular, for all* $j > K$*,* $\phi_j \neq 0$ *implies* $\lambda_j = 0$ *in (3.1) and (3.9).*

**A3.3** (Independence to Measurement Error)**.** *Instrument* $z_i$ *does not depend on measurement error, i.e.* $(s_i - d_i) \perp\!\!\!\perp z_i$*. In particular, for all* $j > K$*,* $\gamma_j \neq 0$ *implies* $\lambda_j = 0$ *in (3.1) and (3.9).*

**A3.3** (Relevance)**.** *Instrument* $z_i$ *is correlated with treatment variable* $d_i$*. In particular,* $\lambda_j \neq 0$ *and* $\gamma_j \neq 0$ *for at least one* $j$ *in equations (3.3) and (3.9).*

**A3.3** (Monotonicity)**.** *The treatment variable* $d_i$ *is monotone w.r.t. the instrument* $z_i$*. In particular,* $sign(\gamma_j \lambda_j) = sign(\gamma_k \lambda_k)$ *for all* $j \neq k \leq K$ *in equations (3.3) and (3.9).*

The following testable assumption turns out to be useful when there are more than one instrument, i.e. $\{z_i^1, ..., z_i^L\}$:

**A3.3** (Conditional Monotonicity). *The sign of the OLS regression coefficient of the reference variable on the instrument does not change when adding additional instruments, i.e. the coefficient* $\rho_k$ *of the small regression,* $s_i = \rho_k z_i^k + r_i^k$*, has the same sign as the coefficient* $\tilde{\rho}_k$ *of the big regression,* $s_i = \sum_{r=1}^{L} \tilde{\rho}_k z_i^k + \tilde{r}_i$ *for all* $k$*.*

**Proposition 3.3.2** (IV-Identification of LATT). *Under the assumptions of partial exclusion A3.3, independence to measurement error A3.3, and relevance A3.3, OLS-IV-regression of* $y_i$ *on* $s_i$ *with instrument* $z_i$ *identifies a local average treatment effect on treated (LATT), which is the ATT of expression (3.8) but with different weights:*

$$\beta_{IV} = LATT \equiv \sum_{j=1}^{K} IR_j \, \omega_j^{IV}, \quad \omega_j^{IV} = \frac{\gamma_j \lambda_j}{\sum_{m=1}^{K} \gamma_m \lambda_m}. \tag{3.10}$$

*Under the assumption of monotonicity A3.3, all weights are positive, i.e.* $\omega_j^{IV} \geq 0, \forall j$*. OLS-IV regression with multiple instruments identifies* $LATT$*, as well, and the weights are positive if both assumptions hold A3.3 and A3.3.*

**Proof**: See appendix 3.7.5.

Note that different instruments put different weights on the components so that the estimated effects depends on the identification strategy. In particular, the weights $\omega_j^{IV}$ of the *LATT* depends on the instrument used; the identified *LATT* therefore varies across instruments. *LATT* is only meaningful if the weights are positive so that the different *IRs* don't cancel each other out. The instrument should thus satisfy the monotonicity assumption A3.3. Interestingly, for multiple instruments, monotonicity alone is not enough for the weights to be positive so that the additional assumption A3.3 needs to be verified.

Next, let there be $L$ control variables (CVs) that depend on the components as follows:

$$x_i^l = \sum_{j=1}^{M} \tau_j^l \mu_{ji}, \quad l = 1, ..., L. \tag{3.11}$$

Remember that the IV approach identifies treatment effects using linear regressions, even when the relationship between instrument and treatment is non-linear. The CV approach, however, needs to remove some of the variation that might be non-linear so that a more complicated estimation technique is necessary. The goal here is to gain intuition about the CV approach, rather than to stay as general as possible. I therefore assume that linear regressions are enough to control for any unwanted variation by imposing normality:

**A3.3** (Normality)**.** *Components are normally distributed, i.e.* $\mu_{ji} = N(0,1)$*, for all* $j$*.*

Normality A3.3 implies that the best linear predictor is also the best predictor so that the controlled reference variable is linear, as well: (see derivation in appendix 3.7.6)

$$s_i - E[s_i|x_i^1, ..., x_i^L] = \sum_{j=1}^{M} \kappa_j \mu_{ji}. \tag{3.12}$$

The following assumptions A3.3, A3.3, and A3.3 make sure that the controlled reference variable only depends on the components of the treatment variable, i.e. $\kappa_j = 0$ for all $j > K$. The decomposition of the components may however differ so that $\kappa_j \neq \gamma_j$ for some $j < K$. The last assumption A3.3 assures that the decomposition remains the same so that $\kappa_j = \gamma_j$ for all $j \leq K$.

**A3.3** (Partial Ignorability)**.** *Selection of reference variable* $s_i$ *is partially random conditional on control variables, i.e.* $f_i(0, u_i) \perp\!\!\!\perp s_i | x_i^1, ..., x_i^L$*. In particular, for all* $j > K$ *where* $\phi_j \neq 0$*, the assumption implies* $\tau_j^l = 0$*, for all* $l$ *in (3.1) and (3.11).*

**A3.3** (Ignorability of Measurement Error)**.** *Reference variable $s_i$ is independent of the measurement error conditional on control variables, i.e. $(s_i - d_i) \perp\!\!\!\perp s_i | x_i^1, ..., x_i^L$. In particular, for all $j > K$ where $\gamma_j \neq 0$, the assumption implies $\tau_j^l = 0$, for all $l$ in (3.1) and (3.11).*

**A3.3** (No Measurement Error in CVs)**.** *There is no variation in CVs independent of reference variable $s_i$. In particular, for any $j$ where $\gamma_j = 0$, the assumption implies $\tau_j^l = 0$, for all $l$ in (3.1) and (3.11)..*

**A3.3** (Disjointedness)**.** *CVs are independent of treatment $d_i$, i.e. $d_i \perp\!\!\!\perp x_i^1, ..., x_i^L$. In particular, the assumption implies $\tau_j^l = 0$ for $j \leq K$ and for all $l$ in (3.11).*

**Proposition 3.3.3** (CV-Identification of ATT)**.** *Under the assumptions of A3.3, A3.3, A3.3, and A3.3, the linear regression of the dependent variable $y_i$ on the reference variable $s_i$ and controls $\{x_i^1, ..., x_i^L\}$ identifies the following combination of IRs:*

$$\beta_{CV} = \sum_{j=1}^{K} IR_j \, \omega_j^{CV}, \quad \omega_j^{CV} = \frac{\gamma_j \kappa_j}{\sum_{m=1}^{K} (\kappa_m)^2}. \tag{3.13}$$

*This combination is not informative, as $\omega_j^{CV}$ may take negative values and the weights may not add up to one. Under the assumption of disjointedness A3.3, however, we have $\kappa_j = \gamma_j$ for all $j$ so that $\beta_{CV} = ATT$.*

**Proof**: See appendix 3.7.6.

Proposition (3.3.3) provides new insights into regressions with control variables. It is not enough to find observables that control for the measurement error of the reference variable. The control variables also need to be independent of the treatment variable, otherwise the estimated parameter $\beta_{CV}$ is not informative. The more heterogeneous the fluctuations that need to be controlled for, the more difficult it gets to find enough control variables that take care of this variation. But none of these control variables should be correlated with the treatment.

Finally, the variance explained by the treatment variable *VET* is usually estimated as follows, which I denote as predicted variance explained by treatment *PVET*:

$$\begin{aligned} y_i = b\tilde{d}_i + u_i, \quad E[u_i|\tilde{d}_i] = E[u_i] = 0, \\ PVET = b^2 Var(\tilde{d}_i), \end{aligned} \tag{3.14}$$

where $\tilde{d}_i = d_i$ when treatment is observed, $\tilde{d}_i \propto z_i$ when using an instrument, and $\tilde{d}_i = s_i - E[s_i|x_i^1, ..., x_i^L]$ when using control variables. The following proposition shows how the identified $PVET$ is only a lower bound of the $VET$:

**Proposition 3.3.4** (Identification of *PVET*)**.** *The $VET$ defined in (3.7) and the $PVET$ defined in (3.14) have the following relationship:*

$$VET = PVET + Var(\phi_j - \varphi m_j),$$

*where $\varphi m_j$ is the linear predictor of $\phi_j$ given $m_j$. In particular, $m_j = \gamma_j$ in the OLS-regression of proposition 3.3.1, $m_j \propto \lambda_j$ in the IV-regression of proposition 3.3.2, and $m_j = \kappa_j$ in the CV-regression of proposition 3.3.3. Note that the weights $\omega_j^{IV}$ and $\omega_j^{CV}$ in propositions 3.3.2 and 3.3.3 don't need to be positive or add up to one so that assumptions A3.3, A3.3, and A3.3 are not required.*

**Proof**: See appendix 3.7.7.

Proposition 3.3.4 shows that we only identify a lower bound of the variance explained by treatment *VET*. Moreover, the difference between *PVET* and *VET* depends on how well the effects of the components can be approximated by the identified treatment variable.

To summarize, the above identification strategies of propositions 3.3.1, 3.3.2, and 3.3.3 extract the components of interest from the reference variable in different ways. By doing so, the resulting weights on the components may no longer mirror the contribution of the components to the reference variable; the weights thus depend on the identification strategy. This could explain why different research papers find different effects for the same shock in the same country and time period. The effects differ, because each research paper weighs the disaggregate shocks differently. This is an alternative explanation to arguing that if two papers get different results, one of them must be misspecified.

In the next section, I show how combining different identification strategies can be beneficial.

## 3.4 Combining Identification Strategies

This section shows that we can do better when combining different identification strategies in three ways. First, we can get closer to the true variance explained by treatment *VET*. Second, we can approximate the individual treatment effect *ITE*. Third, we can define the treatment more precisely by using a reference vector instead of a single reference variable.

Let there be $Q$ identification strategies that produce $Q$ different weighted averages of the components of interest:

$$d_i^q = \Sigma_{j=1}^K m_j^q \mu_{ji}, \tag{3.15}$$

where, for example, $m_j^q = \gamma_j$ in the OLS-regression of proposition 3.3.1, $m_j^q \propto \lambda_j$ in the IV-regression of proposition 3.3.2, and $m_j^q = \kappa_j$ in the CV-regression of proposition 3.3.3. Note that different weighted averages can also be extracted by using different reference variables $s_i$, in addition to using different identification strategies.

Combine the set of identified treatment variables to get closer to the $VET$ using the following linear regression and definition of $PVET^*$:

$$\begin{aligned} y_i &= b_1^* d_i^1 + ... + b_Q^* d_i^Q + u_i^*, \\ PVET^* &\equiv Var(b_1^* d_i^1 + ... + b_Q^* d_i^Q), \end{aligned} \tag{3.16}$$

where $E[u_i^*|d_i^1, ..., d_i^Q] = E[u_i^*] = 0$. The following expression extends proposition 3.3.4 and shows that $PVET^*$ gets closer to the $VET$ than $PVET$ with only one identified treatment variable: (see derivation in appendix 3.7.8)

$$VET = PVET^* + Var(\phi_j - \varphi_1 m_j^1 - ... - \varphi_Q m_j^Q), \tag{3.17}$$

where $(\varphi_1 m_j^1 + ... + \varphi_Q m_j^Q)$ is the linear predictor of $\phi_j$ given $\{m_j^1, ..., m_j^Q\}$, and the weights $m_j^q$ don't need to be positive or add up to one. Hence, combining identification strategies helps to get closer to the true *VET*.

Next, combine the set of identified treatment variables to get an approximation of the $ITE_i$. Let $d_i^1$ be the treatment variable of interest defined in (3.3). The following linear regression can then be used to approximate the individual treatment effect *ITE*: (see derivation in appendix 3.7.9)

$$\begin{aligned} y_i &= b_1^* d_i^1 + ... + b_Q^* d_i^Q + u_i^*, \\ ITE_i^* &\equiv \frac{b_1^* d_i^1 + ... + b_Q^* d_i^Q}{d_i^1} \equiv \frac{\sum_{j=1}^K \phi_j^* \mu_{ji}}{\sum_{j=1}^K \gamma_j \mu_{ji}}, \\ &= ITE_i + \sum_{j=1}^K \frac{(\phi_j^* - \phi_j)}{\gamma_j} \omega_{ji}^{ITE}, \end{aligned} \tag{3.18}$$

where $E[u_i^*|d_i^1, ..., d_i^Q] = E[u_i^*] = 0$. $ITE$'s approximation improves with $Q$, as additional variables increase the match and thus decrease the distance between $\phi_j^*$ and $\phi_j$ on average.

Finally, let's define the treatment in $H$ dimensions, instead of just one dimension. This is related to Atsushi Inoue and Barbara Rossi's (2019), who define shocks as functions rather than scalars. Here I extend the reference variable $s_i$ in (3.3) to a reference vector $\mathbf{s}_i$. The treatment of interest $\delta_i$ is a scalar variable and affects the reference vector as follows

$$\mathbf{s}_i = \mathbf{c}\delta_i + \boldsymbol{\epsilon}_i, \tag{3.19}$$

where $\mathbf{c}$ is a vector of known constants and $\boldsymbol{\epsilon}_i$ is an unknown error vector, where $E[\boldsymbol{\epsilon}_i|\delta_i] = E[\boldsymbol{\epsilon}_i] = 0$. Treatment $\delta_i$ is function of the components of interest only, i.e. $\delta_i = \Sigma_{j=1}^{K}\lambda_j\mu_{ji}$. Hence, given a known vector $\mathbf{c}$, the aim is to find a combination of identified treatment variables that maximizes the explained variance, i.e. $Var(\mathbf{c}\delta_i)$, while maintaining the properties of the error term $\boldsymbol{\epsilon}_i$ so that

$$\delta_i = a_1 d_i^1 + ... + a_Q d_i^Q. \tag{3.20}$$

Vector $\mathbf{c}$ characterizes the treatment of interest. For example, let the treatment of interest be transitory government expenditure shocks. The vector of reference variables then consists of current and future government expenditures, and the values in vector $\mathbf{c}$ are large for current and small for future government expenditures so that the treatment is transitory.

The next section extends the model by including policy effects.

## 3.5 Extension with Policy

Consider the following refinement of model (3.1):

$$\begin{aligned} y_i &= \sum_{j=1}^{M} \phi_j \mu_{ji} = \sum_{j=1}^{M} (h_j(\pi_r) + \psi_j)\mu_{ji} \\ s_i &= \sum_{j=1}^{M} \gamma_j \mu_{ji} = \sum_{j=1}^{M} \left(g_j(\pi_1) + \theta_j\right)\mu_{ji} \end{aligned} \tag{3.21}$$

where the dependent variable $y_i$ and the reference variable $s_i$ are observed for a policy $\pi_1$. As in model (3.1), there are $M$ unobserved components which are independent with zero mean and unit variance, i.e. $E[\mu_{ji}|\ \mu_{1i},...,\mu_{j-1i},\ \mu_{j+1i},...,\mu_{Mi}] = E[\mu_{ji}] = 0$, and $E[(\mu_{ji})^2] = 1$, for all $j$. Model (3.21) refines model (3.1) by dividing the parameters into effects caused by policy, i.e. $h_j(\pi_r)$ and $g_j(\pi_r)$, and effects that occur even when there is no policy, i.e. $\psi_j$ and $\theta_j$; as a result, no policy implies $h_j(\pi_0) = g_j(\pi_0) = 0$, for all $j$. The policy in place $\pi_1$ is constant across all observations $i$ so that data with alternative policies are not available. For instance, counterfactual data with optimal policy $\pi_2$ or no policy $\pi_0$ is unobserved.

Let the first $K$ components be the ones that only occur when the policy is in place so that $\psi_j = \theta_j = 0$ for all $j \leq K$. These components are denoted as *non-systematic* policy interventions, which are for example random mistakes by policy makers, generating additional business cycle fluctuations. The remaining components are *systematic* policy interventions that consistently amplify or dampen economic shocks. A shock $\mu_{ji}$ is amplified when $(h_j(\pi_r) + \psi_j)^2 > (\psi_j)^2$, and dampened when $(h_j(\pi_r) + \psi_j)^2 < (\psi_j)^2$. Some components are not affected by the policy at all so that $h_j(\pi_r) = g_j(\pi_r) = 0$ for all $j > L > K$.

While systematic policy interventions are intentional and may optimize business cycles, non-systematic interventions are random; these random shocks, consequently, manifest a side effect of the policy. In fact, policy makers would do a bad job in balancing business cycles, if these non-systematic policy interventions were an important driver. Nonetheless, policy shocks are widely used in empirical macroeconomics, as they help to identify policy effects under some conditions.

The reference variable is the policy variable that normalizes the policy responses and policy shocks. For example, an expansionary monetary policy response to an oil price shock is normalized to decrease the interest rate, while a contractionary response increases the interest rate. The policy impulse response ($PIR$) is then defined as the part of the impulse response that is caused by the policy:

$$\begin{aligned} PIR_j &\equiv IR_j(\pi_1) - IR_j(\pi_0) \\ &= \frac{\partial y_i(\pi_1)/\partial \mu_{ji}}{\partial s_i(\pi_1)/\partial \mu_{ji}} - \frac{\partial y_i(\pi_0)/\partial \mu_{ji}}{\partial s_i(\pi_0)/\partial \mu_{ji}} = \frac{h_j(\pi_1)}{g_j(\pi_1)}, \quad \forall i,\ \forall j. \end{aligned} \tag{3.22}$$

As discussed in section 3.2, there are infinitely many components, i.e. $M \to \infty$, so that identification of every $PIR_j$ seems impossible. Hence, consider the following aggregation into treatment variable $d_i$:

$$d_i = \sum_{j=1}^{M} g_j(\pi_1)\mu_{ji}. \tag{3.23}$$

For example, monetary policy responses are defined in how the Federal Reserve changes the interest rate at $t$, in response to economic shocks at $t-k$. The reference variable is thus the interest rate, and the treatment variable is the interest rate caused by policy responses to $t-k$ shocks.

Consider the following empirical model:

$$y_i = f_i(d_i, u_i) = \beta_i d_i + u_i, \tag{3.24}$$

where residual $u_i$ aggregates the effects that are not caused by policy, i.e. $u_i = \Sigma_{j=1}^{M}\psi_j\mu_{ji}$. Parameter $\beta_i$ is the effect of the policy treatment variable:

$$\beta_i = \frac{\sum_{j=1}^{M} h_j(\pi_1)\mu_{ji}}{\sum_{j=1}^{M} g_j(\pi_1)\mu_{ji}}, \tag{3.25}$$

which varies across observations $i$.

In addition to the standard summary statistics introduced in (3.7), I define the average treatment effect of systematic treatment (ATST), and average treatment effect of non-systematic treatment (ATNT) as follows:

$$\begin{aligned} ATST &\equiv E\left[ITE_i \frac{(d_i)^2}{E[(d_l)^2|\mu_1, ..., \mu_K]}\middle|\mu_1 = ... = \mu_K = 0\right] \\ ATNT &\equiv E\left[ITE_i \frac{(d_i)^2}{E[(d_l)^2|\mu_{K+1}, ..., \mu_M]}\middle|\mu_{K+1} = ... = \mu_M = 0\right]. \end{aligned} \tag{3.26}$$

ATNT is the average effect of policy shocks, and ATST is the average effect of any policy that responds to economic shocks other than policy shocks.

Similar to section 3.2, I rewrite the objects of interest (3.7) and (3.26) in terms of weighted averages of policy impulse responses ($PIR$): (see derivation in appendix 3.7.10)

$$\begin{aligned}
ITE_i &= \sum_{j=1}^{M} PIR_j\, \omega_{ji}^{ITE}, \quad \omega_{ji}^{ITE} = \frac{g_j(\pi_1)\mu_{ji}}{\sum_{j=1}^{M} g_j(\pi_1)\mu_{ji}} \\
ATT &= \sum_{j=1}^{M} PIR_j\, \omega_{j}^{ATT}, \quad \omega_{j}^{ATT} = \frac{g_j(\pi_1)^2}{\sum_{k=1}^{M} g_k(\pi_1)^2} \\
ATNT &= \sum_{j=1}^{K} PIR_j\, \omega_{j}^{ATNT}, \quad \omega_{j}^{ATNT} = \frac{\omega_j^{ATT}}{\sum_{k=1}^{K} \omega_k^{ATT}} \\
ATST &= \sum_{j=K+1}^{M} PIR_j\, \omega_{j}^{ATST}, \quad \omega_{j}^{ATST} = \frac{\omega_j^{ATT}}{\sum_{k=K+1}^{M} \omega_k^{ATT}} \\
VET &= \sum_{j=1}^{M} (PIR_j)^2\ \omega_j^{ATT}\, Var(d_i) \\
&\geq (ATT)^2\, Var(d_i).
\end{aligned} \tag{3.27}$$

ATT is the weighted average of ATNT and ATST; the former only accounts for PIRs of non-systematic components, and the latter depends on the remaining PIRs. Identification of ATNT is the same as identification of aggregate shocks as described in section 3.3, because $PIR_j = IR_J$ for all $j \leq K$. The following proposition shows that the remaining objects of interest are not identified, because data cannot separate $PIR_j$ from $IR_j$ for $j > K$:

**Proposition 3.5.1** (Simultaneity Problem of Policy Effects)**.** *Even when all components were observed, policy impulse responses $PIR_j$ for $j > K$ are not identified; as a result, when $PIR_j \neq 0$ for at least one $j > K$, $ATST$, $ATT$, $ATE$, and $VET$ are not identified, either.*

**Proof**: See appendix 3.7.11.

Remember that non-systematic policy interventions are just side effects of a policy that aims to reduce business cycle fluctuations. Nevertheless, identification of non-systematic policy interventions is key in empirical macroeconomics. This is because of the following assumption that is often imposed:

**A3.5** (Extrapolation of Non-Systematic Policy)**.** *Systematic policy interventions have the same effect as non-systematic policy interventions, i.e.* $ATST = ATNT$ *in (3.26).*

Once $ATNT$ is identified according to section 3.3, A3.5 implies that $ATST = ATNT$ and therefore $ATT = ATNT$ is identified, as well. This extrapolation from effects of policy shocks to effects of systematic policy is not always reasonable; for instance, the effect of taking pain relievers by accident is almost zero, while the effect of taking pain relievers when suffering pain is large. Hence, the missing effect of non-systematic consumption of pain relievers is not a good proxy for the effect of pain relievers consumed under pain. Similarly, a random policy mistake may not have the same persistent and strong effect as an intentional policy intervention that fights an economic downturn. Assumption A3.5 may thus be violated.

Note also that even when assumption A3.5 holds, the variance contribution of policy shocks provide no information on neither the magnitude, nor on the frequency of systematic policy interventions; accordingly, measuring the contribution of policy shocks to business cycle fluctuations does not reveal anything about whether a policy is important or not. In particular, nothing can be said about the variance explained by treatment $VET$ of the policy.

## 3.6 Conclusion

Business cycle economists combine economic shocks in just a handful of aggregate shocks. While this aggregation is necessary and meaningful, this paper relaxes the assumption that all shocks within an aggregate shock have the same effect on the economy. Relaxing this assumption brings to light that different identification strategies produce different results, as each strategy weighs the disaggregate components of the model differently.

## 3.7 Proofs

### 3.7.1 Derivation of Model Parameter (3.6)

This appendix derives expressions (3.6), which expresses parameter $\beta_i$ defined in (3.5) in terms of impulse responses (IR) defined in (3.2).

*Proof.* Rearrange equation (3.5) and use definition (3.2) to get

$$\begin{aligned}\beta_i &= \frac{\sum_{j=1}^{K}\phi_j\mu_{ji}}{\sum_{l=1}^{K}\gamma_l\mu_{li}} = \sum_{j=1}^{K}\frac{\phi_j}{\gamma_j}\frac{\gamma_j\mu_{ji}}{\sum_{l=1}^{K}\gamma_l\mu_{li}} \\ &= \sum_{j=1}^{K} IR_j\,\omega_{ji}^{ITE}, \quad \omega_{ji}^{ITE} = \frac{\gamma_j\mu_{ji}}{\sum_{j=1}^{K}\gamma_j\mu_{ji}}.\end{aligned}$$

□

### 3.7.2 Consistency of Summary Treatment Effects with Literature

This appendix shows that within model (3.4), the summary treatment effects defined in (3.7) are consistent with the definitions of the literature where the treatment variable is a dummy.

The traditional definitions are denoted with a bar on the top:

$$\begin{aligned}
\overline{ITE}_{i,d_i\in\{0,1\}} &\equiv f_i(1,u_i) - f_i(0,u_i) = \beta_i = ITE_i \\
\overline{ATE}_{d_i\in\{0,1\}} &\equiv E[ITE_i] = ATE \\
\overline{ATT}_{d_i\in\{0,1\}} &\equiv E\left[ITE_i | d_i = 1\right] = E\left[ITE_i \frac{(d_i)^2}{E[(d_l)^2]}\right] = ATT.
\end{aligned}$$

### 3.7.3 Derivation of Summary Statistics (3.8)

This appendix derives expressions (3.8): ATE, ATT, and VET in terms of IRs.

*Proof.* Use definitions (3.7) and (3.2) to derive ATT:

$$\begin{aligned}
ATT &= E\left[\beta_i \frac{(d_i)^2}{E[(d_l)^2]}\right] \\
&= E\left[\frac{\sum_{j=1}^K \phi_j \mu_{ji}}{\sum_{m=1}^K \gamma_m \mu_{mi}}\left(\frac{\left(\sum_{j=1}^K \gamma_j \mu_{ji}\right)^2}{E\left[\left(\sum_{j=1}^K \gamma_j \mu_{ji}\right)^2\right]}\right)\right] \\
&= E\left[\frac{\sum_{j=1}^K \phi_j \mu_{ji} \sum_{j=1}^K \gamma_j \mu_{ji}}{E\left[\left(\sum_{j=1}^K \gamma_j \mu_{ji}\right)^2\right]}\right] = \frac{\sum_{j=1}^K \phi_j \gamma_j}{\sum_{m=1}^K (\gamma_m)^2} \\
&= \sum_{j=1}^K \frac{\phi_j}{\gamma_j} \frac{(\gamma_j)^2}{\sum_{m=1}^K (\gamma_m)^2} = \sum_{j=1}^K IR_j\, \omega_j^{ATT}, \quad \omega_j^{ATT} = \frac{(\gamma_j)^2}{\sum_{m=1}^K (\gamma_m)^2}.
\end{aligned}$$

Next, assume a normal distribution, i.e. $\mu_{ji} \sim N(0,1)$ for all $j$, and combine expression

(3.5) with ATE's definition (3.7) to show that ATE = ATT under normality:

$$
\begin{aligned}
ATE \equiv E\left[ITE_i\right] = E\left[\frac{\sum_{j=1}^{K}\phi_j\mu_{ji}}{\sum_{m=1}^{K}\gamma_m\mu_{mi}}\right] &= E\left[\sum_{j=1}^{K}\frac{\phi_j}{\gamma_j}\frac{\gamma_j\mu_{ji}}{\sum_{m=1}^{K}\gamma_m\mu_{mi}}\right] \\
&= \sum_{j=1}^{K}\frac{\phi_j}{\gamma_j}E\left[\frac{1}{\sum_{m=1}^{K}\gamma_m\mu_{mi}}E\left[\gamma_j\mu_{ji}\middle|\sum_{m=1}^{K}\gamma_m\mu_{mi}\right]\right] \\
&= \sum_{j=1}^{K}\frac{\phi_j}{\gamma_j}E\left[\frac{1}{\sum_{m=1}^{K}\gamma_m\mu_{mi}}\left(\frac{(\gamma_j)^2}{\sum_{m=1}^{K}(\gamma_m)^2}\sum_{m=1}^{K}\gamma_m\mu_{mi}\right)\right] \\
&= \sum_{j=1}^{K}\frac{\phi_j}{\gamma_j}\frac{(\gamma_j)^2}{\sum_{m=1}^{K}(\gamma_m)^2} = ATT.
\end{aligned}
$$

Use definitions (3.2) and (3.7) to solve for $VET$:

$$
\begin{aligned}
VET = Var\left(f_i(d_i,u_i) - f_i(0,u_i)\right) = Var\left(\sum_{j=1}^{K}\phi_j\mu_{ji}\right) &= \sum_{j=1}^{K}\left(\phi_j\right)^2 \\
&= \sum_{j=1}^{K}\frac{(\phi_j)^2}{(\gamma_j)^2}\frac{(\gamma_j)^2}{\sum_{m=1}^{K}(\gamma_m)^2}\sum_{m=1}^{K}(\gamma_m)^2 \\
&= \sum_{j=1}^{K}\left(IR_j\right)^2\ \omega_j^{ATT}\ Var(d_i), \quad \omega_j^{ATT} = \frac{(\gamma_j)^2}{\sum_{m=1}^{K}(\gamma_m)^2}.
\end{aligned}
$$

Finally, proof that $VET \geq (ATT)^2\, Var(d_i)$ using the above expressions:

$$
VET = \sum_{m=1}^{K}(\phi_m)^2 \geq \left(\frac{\sum_{j=1}^{K}\phi_j\gamma_j}{\sum_{m=1}^{K}(\gamma_m)^2}\right)^2\sum_{m=1}^{K}(\gamma_m)^2 = (ATT)^2\, Var(d_i)
$$

$$
\sum_{m=1}^{K}(\phi_m)^2 \geq \frac{\left(\sum_{j=1}^{K}\phi_j\gamma_j\right)^2}{\sum_{m=1}^{K}(\gamma_m)^2}
$$

$$
\left(\sum_{m=1}^{K}(\phi_m)^2\right)\left(\sum_{m=1}^{K}(\gamma_m)^2\right) \geq \left(\sum_{j=1}^{K}\phi_j\gamma_j\right)^2,
$$

by Cauchy-Schwarz inequality. Equality holds if and only if $\phi_j = k\gamma_j$ for a non-zero constant $k$. □

### 3.7.4 Proposition 3.3.1 (Identification of ATT)

*Proof.* OLS-regression refers to the following model:

$$y_i = bs_i + e_i, \quad E[e_i s_i] = E[e_i] = 0.$$

Hence,

$$0 = E[e_i s_i] = E[(y_i - bs_i)s_i] \quad \Leftrightarrow \quad b = \frac{E[y_i s_i]}{E[(s_i)^2]}.$$

Plug in for $s_i$ and $y_i$ according to model (3.1):

$$b = \frac{E\left[\left(\sum_{j=1}^{M} \gamma_j \mu_{ji}\right)\left(\sum_{j=1}^{M} \phi_j \mu_{ji}\right)\right]}{E\left[\left(\sum_{j=1}^{M} \gamma_j \mu_{ji}\right)^2\right]} = \frac{\sum_{j=1}^{M} \gamma_j \phi_j}{\sum_{j=1}^{M} (\gamma_j)^2}.$$

A3.3 implies $\gamma_j = 0$ for all $j > K$ so that

$$b = \frac{\sum_{j=1}^{K} \gamma_j \phi_j}{\sum_{j=1}^{K} (\gamma_j)^2} = ATT = \sum_{j=1}^{K} IR_j \, \omega_j^{ATT}, \quad \omega_j^{ATT} = \frac{(\gamma_j)^2}{\sum_{k=1}^{K} (\gamma_k)^2}.$$

Appendix 3.7.3 shows that $ATE = ATT$ when the components are normally distributed. Hence, under A3.3 and normality, the regression of $y_i$ on $d_i$ identifies ATE. □

### 3.7.5 Proposition 3.3.2 (IV-Identification of LATT)

*Proof.* The OLS-IV-regression refers to the following two-stage model:

1. $s_i = az_i + \eta_i, \quad E[\eta_i z_i] = E[\eta_i] = 0$
2. $y_i = b(az_i) + \nu_i, \quad E[\nu_i z_i] = E[\nu_i] = 0$

Hence,

$$0 = E[\eta_i z_i] = E[(s_i - a z_i) z_i] \quad \Leftrightarrow \quad a = \frac{E[s_i z_i]}{E[(z_i)^2]}.$$
$$0 = E[\nu_i z_i] = E[(y_i - b a z_i) z_i] \quad \Leftrightarrow \quad b = \frac{E[z_i y_i]}{aE[(z_i)^2]} = \frac{E[z_i y_i]}{E[z_i s_i]}.$$

Plug in for $s_i$, $y_i$, and $z_i$ according to equations (3.1) and (3.9):

$$\begin{aligned} b &= \frac{E\left[\left(\sum_{j=1}^{M} \lambda_j \mu_{ji}\right)\left(\sum_{j=1}^{M} \phi_j \mu_{ji}\right)\right]}{E\left[\left(\sum_{j=1}^{M} \lambda_j \mu_{ji}\right)\left(\sum_{j=1}^{M} \gamma_j \mu_{ji}\right)\right]} \\ &= \frac{\sum_{j=1}^{M} \phi_j \lambda_j}{\sum_{m=1}^{M} \gamma_m \lambda_m} \end{aligned}$$

Note that the partial exclusion assumption A3.3 restricts $\lambda_j = 0$ whenever $\phi_j \neq 0$ for $j > K$, and independence to measurement error assumption A3.3 restricts $\lambda_j = 0$ whenever $\gamma_j \neq 0$ for $j > K$ so that

$$b = \frac{\sum_{j=1}^{K} \phi_j \lambda_j}{\sum_{m=1}^{K} \gamma_m \lambda_m} = \sum_{j=1}^{K} IR_j \, \omega_j^{IV}, \; \omega_j^{IV} = \frac{\gamma_j \lambda_j}{\sum_{m=1}^{K} \gamma_m \lambda_m}.$$

□

### 3.7.6 Proposition 3.3.3 (CV-Identification of ATT)

*Proof.* First, let's derive equation (3.12). Remember that under normality A3.3, the best linear predictor is also the best predictor so that there exists $\{c_l\}_{l=1}^{L}$ that satisfy

$$E[s_i | x_i^1, ..., x_i^L] = \sum_{l=1}^{L} c_l x_i^l.$$

Next, use the left hand side of equation (3.12) and replace the reference variable and the CVs according to (3.3) and (3.11):

$$
\begin{aligned}
s_i - E[s_i|x_i^1, ..., x_i^L] &= \sum_{j=1}^{M} \gamma_j \mu_{ji} - \sum_{l=1}^{L} c_l \sum_{j=1}^{M} \tau_j^l \mu_{ji} \\
&= \sum_{j=1}^{M} \left( \gamma_j - \sum_{l=1}^{L} c_l \tau_j^l \right) \mu_{ji} \\
&= \sum_{j=1}^{M} \kappa_j \mu_{ji} \equiv \eta_i,
\end{aligned}
$$

where $\kappa_j = \gamma_j - \Sigma_{l=1}^{L} c_l \tau_j^l$.

The OLS-CV-regression refers to the following two-stage model:

1. $s_i = a_1 x_i^1 + ... + a_L x_i^L + \eta_i, \quad E[\eta_i x_i^l] = E[\eta_i] = 0,\ \forall l$
2. $y_i = b\eta_i + \nu_i, \quad E[\nu_i \eta_i] = E[\nu_i] = 0.$

Hence,

$$
\begin{aligned}
0 = E[\eta_i x_i] = E[(s_i - a x_i) x_i] \quad &\Leftrightarrow \quad a = \frac{E[x_i s_i]}{E[(x_i)^2]} \\
0 = E[\nu_i \eta_i] = E[(y_i - b\eta_i)\eta_i] \quad &\Leftrightarrow \quad b = \frac{E[\eta_i y_i]}{E[(\eta_i)^2]}.
\end{aligned}
$$

Next, plug in equations (3.1) and (3.11):

$$
b = \frac{E\left[\left(\sum_{j=1}^{M} \kappa_j \mu_{ji}\right)\left(\sum_{j=1}^{M} \phi_j \mu_{ji}\right)\right]}{E\left[\left(\sum_{j=1}^{M} \kappa_j \mu_{ji}\right)^2\right]} = \frac{\sum_{j=1}^{M} \phi_j \kappa_j}{\sum_{m=1}^{M} (\kappa_m)^2}.
$$

All terms drop out for $j > K$ under the assumptions of A3.3 and A3.3, and A3.3 so that

$$
b = \frac{\sum_{j=1}^{K} \phi_j \kappa_j}{\sum_{m=1}^{K} (\kappa_m)^2} = \sum_{j=1}^{K} IR_j\, \omega_j^{CV}, \quad \omega_j^{CV} = \frac{\gamma_j \kappa_j}{\Sigma_{m=1}^{K} (\kappa_m)^2}.
$$

$\square$

### 3.7.7 Proposition 3.3.4 (Identification of *PVET*)

*Proof.* Consider the following hypothetical linear regression for $j = 1, ..., K$:

$$\phi_j = \varphi m_j + \epsilon_j, \quad E[\epsilon_j | m_j] = E[\epsilon_j] = 0,$$
$$\Rightarrow \varphi = \frac{E[\phi_j m_j]}{E[(m_j)^2]}.$$

Next, square both sides and take expectations across $j$:

$$\begin{aligned} E[(\phi_j)^2] &= E[(\varphi m_j + \epsilon_j)^2], \\ &= \varphi^2 E[(m_j)^2] + E[(\epsilon_j)^2], \\ &= \frac{E[\phi_j m_j]^2}{E[(m_j)^2]} + Var(\epsilon_j), \end{aligned}$$

and therefore

$$\begin{aligned} VET \equiv \sum_{j=1}^{K} (\phi_j)^2 &= \frac{\left(\sum_{j=1}^{K} \phi_j m_j\right)^2}{\sum_{j=1}^{K} (m_j)^2} + Var(\epsilon_j), \\ &= \left(\frac{\sum_{j=1}^{K} \phi_j m_j}{\sum_{j=1}^{K} (m_j)^2}\right)^2 \sum_{j=1}^{K} (m_j)^2 + Var(\epsilon_j), \\ &= (b)^2 Var(\tilde{d}_i) + Var(\phi_j - \varphi m_j), \\ &= PVET + Var(\phi_j - \varphi m_j), \end{aligned}$$

where $\tilde{d}_i \equiv \sum_{j=1}^{K} m_j \mu_{ji}$, and $b\tilde{d}_i$ is the linear predictor of $y_i$ given $\tilde{d}_i$. □

### 3.7.8 Identification of *PVET*$^*$ *(3.17)*

*Proof.* Consider the following two linear regression:

$$y_i = \beta_1 d_i^1 + ... + \beta_Q d_i^Q + u_j, \quad E[u_i | d_i^1, ..., d_i^Q] = E[u_i] = 0,$$

$$\phi_j = \varphi_1 m_j^1 + ... + \varphi_Q m_j^Q + \epsilon_j, \quad E[\epsilon_j | m_j^1, ..., m_j^Q] = E[\epsilon_j] = 0,$$

where $d_i^q = \Sigma_{j=1}^K m_j^q \mu_{ji}$. Using matrix notation,

$$\mathbf{y} = \mathbf{D}\boldsymbol{\beta} + \mathbf{u}, \quad \boldsymbol{\beta} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y},$$

$$\boldsymbol{\phi} = \mathbf{M}\boldsymbol{\varphi} + \boldsymbol{\epsilon}, \quad \boldsymbol{\varphi} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\boldsymbol{\phi}.$$

The predicted variance explained by multiple treatments ($PVET^*$) is therefore

$$\begin{aligned} PVET^* &= \boldsymbol{\beta}'\mathbf{D}'\mathbf{D}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y}, \end{aligned}$$

and the $VET$:

$$\begin{aligned} VET = \boldsymbol{\phi}'\boldsymbol{\phi} &= \boldsymbol{\varphi}'\mathbf{M}'\mathbf{M}\boldsymbol{\varphi} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} \\ &= \boldsymbol{\phi}'\mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\boldsymbol{\phi} + \boldsymbol{\epsilon}'\boldsymbol{\epsilon}. \end{aligned}$$

Hence, if (a) $\mathbf{M}'\mathbf{M} = \mathbf{D}'\mathbf{D}$ and (b) $\mathbf{M}'\boldsymbol{\phi} = \mathbf{D}'\mathbf{y}$ , then

$$\begin{aligned} VET &= PVET^* + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} \\ &= PVET^* + Var(\phi_j - \varphi_1 m_j^1 + ... + \varphi_Q m_j^Q). \end{aligned}$$

(a) To show that $\mathbf{M}'\mathbf{M} = \mathbf{D}'\mathbf{D}$, calculate row-column entry $(q, k)$ of $\mathbf{D}'\mathbf{D}$:

$$\begin{aligned} (\mathbf{D}'\mathbf{D})(q,k) \equiv \begin{pmatrix} d_1^q & \cdots & d_N^q \end{pmatrix} \begin{pmatrix} d_1^k \\ \vdots \\ d_N^k \end{pmatrix} &= \sum_{i=1}^N d_i^q d_i^k \\ &= \sum_{i=1}^N \left( \sum_{j=1}^K m_j^q \mu_{ji} \right) \left( \sum_{j=1}^K m_j^k \mu_{ji} \right) \\ &= \sum_{j=1}^K \sum_{l=1}^K m_j^q m_l^k E[\mu_{ji}\mu_{li}] = \sum_{j=1}^K m_j^q m_j^k, \end{aligned}$$

which is equal to the row-column entry $(q, k)$ of $\mathbf{M}'\mathbf{M}$:

$$(\mathbf{M}'\mathbf{M})(q,k) \equiv \begin{pmatrix} m_1^q & \cdots & m_K^q \end{pmatrix} \begin{pmatrix} m_1^k \\ \vdots \\ m_K^k \end{pmatrix} = \sum_{j=1}^K m_j^q m_j^k.$$

(b) To show that $\mathbf{M}'\boldsymbol{\phi} = \mathbf{D}'\mathbf{y}$, calculate the $q^{th}$ entry of $\mathbf{D}'\mathbf{y}$:

$$
\begin{aligned}
(\mathbf{D}'\mathbf{y})(q) \equiv \begin{pmatrix} d_1^q & \cdots & d_N^q \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} &= \sum_{i=1}^{N} d_i^q y_i \\
&= \sum_{i=1}^{N} \left( \sum_{j=1}^{K} m_j^q \mu_{ji} \right) \left( \sum_{j=1}^{M} \phi_j \mu_{ji} \right) \\
&= \sum_{j=1}^{K} \sum_{l=1}^{M} m_j^q \phi_l E[\mu_{ji}\mu_{li}] = \sum_{j=1}^{K} m_j^q \phi_j,
\end{aligned}
$$

which is equal to the $q^{th}$ entry of $\mathbf{M}'\boldsymbol{\phi}$:

$$
(\mathbf{M}'\boldsymbol{\phi})(q) \equiv \begin{pmatrix} m_1^q & \cdots & m_K^q \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_K \end{pmatrix} = \sum_{j=1}^{K} m_j^q \phi_j.
$$

□

### 3.7.9 Derivation of ITE's Approximation (3.18)

This appendix derives expression (3.18). Note that regression parameter $b$ of expression (3.18) is equal to $ATT$ according to proposition 3.3.1 so that

$$
\begin{aligned}
ITE_i^* = \frac{\sum_{j=1}^{K} \phi_j^* \mu_{ji}}{\sum_{j=1}^{K} \gamma_j \mu_{ji}} &= \sum_{j=1}^{K} \frac{\phi_j^*}{\gamma_j} \frac{\gamma_j \mu_{ji}}{\sum_{j=1}^{K} \gamma_j \mu_{ji}} = \sum_{j=1}^{K} \frac{(\phi_j + \phi_j^* - \phi_j)}{\gamma_j} \frac{\gamma_j \mu_{ji}}{\sum_{j=1}^{K} \gamma_j \mu_{ji}}, \\
&= ITE_i + \sum_{j=1}^{K} \frac{(\phi_j^* - \phi_j)}{\gamma_j} \omega_{ji}^{ITE}.
\end{aligned}
$$

### 3.7.10 Derivation of Summary Statistics (3.27)

Appendix 3.7.3 derives ITE, ATT, ATE, and VET of expression (3.8), which are the same as the summary statistics of expression (3.27) with the only difference that model (3.24) is used instead of model (3.8). This appendix derives the remaining statistics ATNT and ATST.

*Proof.* Use definitions (3.26) and (3.22) to derive ATNT:

$$
\begin{aligned}
ATNT &= E\left[ ITE_i \frac{(d_i)^2}{E[(d_l)^2 | \mu_{K+1}, ..., \mu_M]} \middle| \mu_{K+1} = ... = \mu_M = 0 \right] \\
&= E\left[ \frac{\sum_{j=1}^K h_j(\pi_1)\mu_{ji}}{\sum_{m=1}^K g_m(\pi_1)\mu_{mi}} \left( \frac{\left(\sum_{j=1}^K g_j(\pi_1)\mu_{ji}\right)^2}{E\left[\left(\sum_{j=1}^K g_j(\pi_1)\mu_{ji}\right)^2\right]} \right) \right] \\
&= E\left[ \frac{\sum_{j=1}^K h_j(\pi_1)\mu_{ji} \sum_{j=1}^K g_j(\pi_1)\mu_{ji}}{E\left[\left(\sum_{j=1}^K g_j(\pi_1)\mu_{ji}\right)^2\right]} \right] = \frac{\sum_{j=1}^K h_j(\pi_1) g_j(\pi_1)}{\sum_{m=1}^K g_m(\pi_1)^2} \\
&= \sum_{j=1}^K \frac{h_j(\pi_1)}{g_j(\pi_1)} \frac{g_j(\pi_1)^2}{\sum_{m=1}^K g_m(\pi_1)^2} = \sum_{j=1}^K IR_j\, \omega_j^{ATNT}, \quad \omega_j^{ATNT} = \frac{g_j(\pi_1)^2}{\sum_{m=1}^K g_m(\pi_1)^2},
\end{aligned}
$$

and similarly, ATST is the following function of TCEs:

$$
ATST = \sum_{j=K+1}^M IR_j\, \omega_j^{ATST}, \quad \omega_j^{ATST} = \frac{g_j(\pi_1)^2}{\sum_{m=K+1}^M g_m(\pi_1)^2}.
$$

ATT is the weighted average of the two:

$$
ATT = ATNT \left( \frac{\sum_{j=1}^K g_j(\pi_1)^2}{\sum_{m=1}^M g_m(\pi_1)^2} \right) + ATST \left( \frac{\sum_{j=K+1}^M g_j(\pi_1)^2}{\sum_{m=1}^M g_m(\pi_1)^2} \right).
$$

□

### 3.7.11 Proposition 3.5.1 (Simultaneity Problem of Policy Effects)

From model (3.21) we know that even if all components $\{\mu_{1i}, ..., \mu_{Mi}\}$ are observed, $h_j(\pi_1)$ and $\psi_j$ cannot be separately identified, because no variation separates these two effects. Hence,

$PIR_j$ for $j > K$ in equation (3.22) can only be identified if additional information about the relative importance of the policy is available.

# Bibliography


**Adam, Klaus, and Mario Padula.** 2011. “Inflation dynamics and subjective expectations in the United States.” *Economic Inquiry*, 49(1): 13–25.

**Alessi, Lucia, Matteo Barigozzi, and Marco Capasso.** 2008. “A review of nonfundamentalness and identification in structural VAR models.” European Central Bank Working Paper Series 0922.

**Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin.** 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association*, 91(434): 444–455.

**Arrow, K.J.** 1965. *Aspects of the theory of risk-bearing. Yrjö Jahnsson lectures*, Yrjö Jahnssonin Säätiö.

**Bachmann, Rüdiger, Steffen Elstner, and Eric R. Sims.** 2013. “Uncertainty and Economic Activity: Evidence from Business Survey Data.” *American Economic Journal: Macroeconomics*, 5(2): 217–49.

**Baker, Scott R., Nicholas Bloom, and Steven J. Davis.** 2015. “Measuring Economic Policy Uncertainty.” Centre for Economic Performance, LSE CEP Discussion Papers dp1379.

**Barsky, Robert B., and Eric R. Sims.** 2011. “News shocks and business cycles.” *Journal of Monetary Economics*, 58(3): 273–289.

**Barsky, Robert B., and Eric R. Sims.** 2012. “Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence.” *American Economic Review*, 102(4): 1343–77.

**Barsky, Robert B., Susanto Basu, and Keyoung Lee.** 2015. “Whither News Shocks?” *NBER Macroeconomics Annual*, 29(1): 225–264.

**Basu, Susanto, John G. Fernald, and Miles S. Kimball.** 2006. “Are Technology Improvements Contractionary?” *American Economic Review*, 96(5): 1418–1448.

**Beaudry, Paul, and Bernd Lucke.** 2010. “Letting Different Views about Business Cycles Compete.” In *NBER Macroeconomics Annual 2009, Volume 24*. *NBER Chapters*, 413–455. National Bureau of Economic Research, Inc.

**Beaudry, Paul, and Franck Portier.** 2004. “An exploration into Pigou’s theory of cycles.” *Journal of Monetary Economics*, 51(6): 1183 – 1216.

**Beaudry, Paul, and Franck Portier.** 2006. “Stock Prices, News, and Economic Fluctuations.” *American Economic Review*, 96(4): 1293–1307.

**Beaudry, Paul, and Franck Portier.** 2014. “News-Driven Business Cycles: Insights and Challenges.” *Journal of Economic Literature*, 52(4): 993–1074.

**Beaudry, Paul, Martial Dupaigne, and Franck Portier.** 2011. “Modeling news-driven international business cycles.” *Review of Economic Dynamics*, 14(1): 72 – 91. Special issue: Sources of Business Cycles.

**Bekaert, Geert, and Guojun Wu.** 2000. "Asymmetric Volatility and Risk in Equity Markets." *Review of Financial Studies*, 13(1): 1–42.

**Benati, Luca.** 2016. "Investment-Specific and Neutral News Shocks and Macroeconomic Fluctuations." University of Bern.

**Bernanke, Ben S.** 1983. "Irreversibility, Uncertainty, and Cyclical Investment." *The Quarterly Journal of Economics*, 98(1): 85–106.

**Binder, Michael, and M. Hashem Pesaran.** 1997. "Multivariate Linear Rational Expectations Models: Characterization of the Nature of the Solutions and Their Fully Recursive Computation." *Econometric Theory*, 13(6): 877–888.

**Black, F.** 1976. "Studies of stock price volatility changes." Proceedings of the Meetings of the American Statistical Association, Business and Economics Statistics Division, pp. 177-181.

**Blanchard, Olivier Jean, and Danny Quah.** 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review*, 79(4): 655–73.

**Blanchard, Olivier J., Jean-Paul L'Huillier, and Guido Lorenzoni.** 2013. "News, Noise, and Fluctuations: An Empirical Exploration." *American Economic Review*, 103(7): 3045–70.

**Bloom, Nicholas.** 2009. "The Impact of Uncertainty Shocks." *Econometrica*, 77(3): 623–685.

**Bloom, Nicholas.** 2014. "Fluctuations in Uncertainty." *The Journal of Economic Perspectives*, 28(2): 153–175.

**Blundell, Richard, and Stephen Bond.** 2000. "GMM Estimation with persistent panel data: an application to production functions." *Econometric Reviews*, 19(3): 321–340.

**Brennan, Michael J., and Eduardo S. Schwartz.** 1985. "Evaluating Natural Resource Investments." *The Journal of Business*, 58(2): 135–157.

**Bullard, James.** 2018. "R-Star Wars: The Phantom Menace." Presentation at the 34th Annual National Association for Business Economics (NABE) Economic Policy Conference. Washington, D.C.

**Caldara, Dario, Cristina Fuentes-Albero, Simon Gilchrist, and Egon Zakrajšek.** 2016. "The Macroeconomic Impact of Financial and Uncertainty Shocks." National Bureau of Economic Research, Inc NBER Working Papers 22058.

**Campbell, John Y., and Ludger Hentschel.** 1992. "No news is good news." *Journal of Financial Economics*, 31(3): 281 – 318.

**Christiano, Lawrence J., Cosmin Ilut, Roberto Motto, and Massimo Rostagno.** 2010. "Monetary policy and stock market booms." *Proceedings - Economic Policy Symposium - Jackson Hole*, 85–145.

**Christie, Andrew A.** 1982. "The stochastic behavior of common stock variances: Value, leverage and interest rate effects." *Journal of Financial Economics*, 10(4): 407–432.

**Chu, Chia-Shang James, Gary J. Santoni, and Tung Liu.** 1996. "Stock market volatility and regime shifts in returns." *Information Sciences*, 94(1): 179 – 190.

**Clarida, Richard, Jordi Galí, and Mark Gertler.** 2000. "Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory." *The Quarterly Journal of Economics*, 115(1): 147–180.

**Coibion, Olivier, and Yuriy Gorodnichenko.** 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120(1): 116 – 159.

**Coibion, Olivier, and Yuriy Gorodnichenko.** 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review*, 105(8): 2644–78.

**Collard, David.** 1996. "Pigou and Modern Business Cycle Theory." *The Economic Journal*, 106(437): 912–924.

**Davis, Joshua Mark.** 2007. "News and the Term Structure in General Equilibrium." Unpublished.

**Del Negro, Marco, and Frank Schorfheide.** 2013. "Chapter 2 - DSGE Model-Based Forecasting." In *Handbook of Economic Forecasting*. Vol. 2, , ed. Graham Elliott and Allan Timmermann, 57 – 140. Elsevier.

**Del Negro, Marco, and Stefano Eusepi.** 2011. "Fitting observed inflation expectations." *Journal of Economic Dynamics and Control*, 35(12): 2105 – 2131.

**Faust, Jon.** 1998. "The robustness of identified VAR conclusions about money." Board of Governors of the Federal Reserve System (U.S.) International Finance Discussion Papers 610.

**Fernald, John G.** 2012. "A quarterly, utilization-adjusted series on total factor productivity." Federal Reserve Bank of San Francisco Working Paper Series 2012-19.

**Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson.** 2007. "ABCs (and Ds) of Understanding VARs." *The American Economic Review*, 97(3): 1021–1026.

**Francis, Neville, Michael T. Owyang, Jennifer E. Roush, and Riccardo DiCecio.** 2014. "A Flexible Finite-Horizon Alternative to Long-Run Restrictions with an Application to Technology Shocks." *The Review of Economics and Statistics*, 96(3): 638–647.

**French, Kenneth R., G.William Schwert, and Robert F. Stambaugh.** 1987. "Expected stock returns and volatility." *Journal of Financial Economics*, 19(1): 3 – 29.

**Fuhrer, Jeff.** 2017. "Expectations as a Source of Macroeconomic Persistence: Evidence from Survey Expectations in a Dynamic Macro Model." *Journal of Monetary Economics*, 86: 22 – 35.

**Galí, Jordi, and Mark Gertler.** 1999. "Inflation dynamics: A structural econometric analysis." *Journal of Monetary Economics*, 44(2): 195 – 222.

**Gilchrist, Simon, Jae W. Sim, and Egon Zakrajšek.** 2014. "Uncertainty, Financial Frictions, and Investment Dynamics." National Bureau of Economic Research, Inc NBER Working Papers 20038.

**Hamilton, James D.** 1994. *Time series analysis.* Princeton:Princeton University Press.

**Heckman, James J., and Edward J. Vytlacil.** 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences of the United States of America*, 96(8): 4730–4734.

**Heckman, James J., and Edward J. Vytlacil.** 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New." In *Handbook of Econometrics*. Vol. 6 of *Handbook of Econometrics*, , ed. J.J. Heckman and E.E. Leamer, Chapter 71. Elsevier.

**Heckman, James J., and Edward Vytlacil.** 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation1." *Econometrica*, 73(3): 669–738.

**Heckman, James J., and Richard Robb.** 1985. "Alternative methods for evaluating the impact of interventions: An overview." *Journal of Econometrics*, 30(1): 239 – 267.

**Heiss, Florian, and Viktor Winschel.** 2006. "Estimation with Numerical Integration on Sparse Grids." University of Munich, Department of Economics Discussion Papers in Economics 916.

**Hirose, Yasuo, and Takushi Kurozumi.** 2012. "Identifying News Shocks with Forecast Data." Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University CAMA Working Papers 2012-01.

**Hodrick, Robert J, and Edward C Prescott.** 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking*, 29(1): 1–16.

**Imbens, Guido W, and Joshua D Angrist.** 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–75.

**Inoue, Atsushi, and Barbara Rossi.** 2019. "The Effects of Conventional and Unconventional Monetary Policy: A New Approach." Barcelona Graduate School of Economics Working Papers 1082.

**Jain, Monica.** 2017. "Perceived Inflation Persistence." *Journal of Business & Economic Statistics*, 1–11.

**Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng.** 2015. "Measuring Uncertainty." *American Economic Review*, 105(3): 1177–1216.

**Knight, Frank H.** 1921. *Risk, Uncertainty, and Profit.* Hart, Schaffner and Marx; Houghton Mifflin Company, Boston.

**Kydland, Finn E., and Edward C. Prescott.** 1982. "Time to Build and Aggregate Fluctuations." *Econometrica*, 50(6): 1345–1370.

**Laubach, Thomas, and John C. Williams.** 2003. "Measuring the Natural Rate of Interest." *The Review of Economics and Statistics*, 85(4): 1063–1070.

**Leduc, Sylvain, and Zheng Liu.** 2015. "Uncertainty Shocks are Aggregate Demand Shocks." Federal Reserve Bank of San Francisco Working Paper 2012-10.

**Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang.** 2008. "Fiscal Foresight: Analytics and Econometrics." National Bureau of Economic Research, Inc NBER Working Papers 14028.

**Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang.** 2013. "Fiscal Foresight and Information Flows." *Econometrica*, 81(3): 1115–1145.

**Levinsohn, James, and Amil Petrin.** 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2): 317–341.

**Lintner, John.** 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *The Review of Economics and Statistics*, 47(1): 13–37.

**Lucas, Robert E.** 1972. "Expectations and the neutrality of money." *Journal of Economic Theory*, 4(2): 103 – 124.

**Lucas, Robert E.** 1976. "Econometric policy evaluation: A critique." *Carnegie-Rochester Conference Series on Public Policy*, 1(1): 19–46.

**Maćkowiak, Bartosz, and Mirko Wiederholt.** 2009. "Optimal Sticky Prices under Rational Inattention." *The American Economic Review*, 99(3): 769–803.

**Mankiw, N. Gregory, and Ricardo Reis.** 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *The Quarterly Journal of Economics*, 117(4): 1295–1328.

**Manski, Charles F.** 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531–542.

**Manski, Charles F.** 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329–1376.

**Markowitz, Harry.** 1952. "Portfolio selection." *The Journal of Finance*, 7(1): 77–91.

**McDonald, Robert, and Daniel Siegel.** 1986. "The Value of Waiting to Invest." *The Quarterly Journal of Economics*, 101(4): 707–728.

**Milani, Fabio, and Ashish Rajrhandari.** 2012. "Observed Expectations, News Shocks, and the Business Cycle." University of California-Irvine, Department of Economics Working Papers 121305.

**Muth, John F.** 1961. "Rational Expectations and the Theory of Price Movements." *Econometrica*, 29(3): 315–335.

**Nguyen, Thuy Lan, and Wataru Miyamoto.** 2014. "News shocks and Business cycles: Evidence from forecast data." Society for Economic Dynamics 2014 Meeting Papers 259.

**Olley, G. Steven, and Ariel Pakes.** 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–1297.

**Pigou, A.** 1927. *Industrial Fluctuations.* MacMillan, London.

**Pindyck, Robert S.** 1984. "Risk, Inflation, and the Stock Market." *American Economic Review*, 74(3): 335–51.

**Pratt, John W.** 1964. "Risk Aversion in the Small and in the Large." *Econometrica*, 32(1/2): 122–136.

**Roberts, John M.** 1995. "New Keynesian Economics and the Phillips Curve." *Journal of Money, Credit and Banking*, 27(4): 975–984.

**Rubin, Donald B.** 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, 66(5): 688–701.

**Ryngaert, Jane.** 2017. “What Do (and Don’t) Forecasters Know about U.S. Inflation?” University of Texas at Austin.

**Schwert, G. William.** 1989. “Why Does Stock Market Volatility Change Over Time?” *The Journal of Finance*, 44(5): 1115–1153.

**Sharpe, William F.** 1964. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *The Journal of Finance*, 19(3): 425–442.

**Shawky, Hany A., and Achla Marathe.** 1995. “Expected stock returns and volatility in a two-regime market.” *Journal of Economics and Business*, 47(5): 409 – 421.

**Shiller, R. J.** 2015. *Irrational Exuberance.* Princeton University Press.

**Sims, Christopher A.** 1980. “Macroeconomics and Reality.” *Econometrica*, 48(1): 1–48.

**Sims, Christopher A.** 2003. “Implications of rational inattention.” *Journal of Monetary Economics*, 50(3): 665 – 690.

**Sims, Christopher A., and Tao Zha.** 2006. “Were there regime switches in U.S. monetary policy?” *The American Economic Review*, 96(1): 54–81.

**Sims, Eric R.** 2012. “News, Non-Invertibility, and Structural VARs.” In *Nathan Balke, Fabio Canova, Fabio Milani, Mark A. Wynne (ed.) Advances in Econometrics, Volume 28, DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*. 81–135. Emerald Group Publishing Limited.

**Uhlig, Harald.** 2004. “Do Technology Shocks Lead to a Fall in Total Hours Worked?” *Journal of the European Economic Association*, 2(2-3): 361–371.

**Uhlig, Harald.** 2005. “What are the effects of monetary policy on output? Results from an agnostic identification procedure.” *Journal of Monetary Economics*, 52(2): 381–419.

**Vuong, Quang, and Haiqing Xu.** 2017. “Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity.” *Quantitative Economics*, 8(2): 589–610.

**Watson, Mark W.** 1986. “Vector autoregressions and cointegration.” In *Handbook of Econometrics*. Vol. 4 of *Handbook of Econometrics*, , ed. R. F. Engle and D. McFadden, Chapter 47, 2843–2915. Elsevier.

**Woodford, Michael.** 2003. “Imperfect Common Knowledge and the Effects of Monetary Policy.” In *Knowledge, Information, and Expectations in Modern Macroeconomics : In Honor of Edmund S. Phelps*. , ed. Philippe Aghion, Roman Frydman, Joseph Stiglitz and Michael Woodford, 25–58. Princeton, N.J.: Princeton University Press.

**Wu, Guojun.** 2001. “The Determinants of Asymmetric Volatility.” *Review of Financial Studies*, 14(3): 837–859.