Latent Variables in Macroeconomic Models: A Frequency-Domain Investigation

Nikolay Iskrev

Abstract

I propose a formal method for decomposing frequency-domain information about latent variables in dynamic models. These models describe the joint probability distribution of observed and latent variables. Information transfer from observed to latent variables is quantified as the reduction in uncertainty between the prior and posterior distributions of a given latent variable. By employing frequency-domain techniques, the total information transfer is disaggregated into frequency-specific contributions and the contributions of individual observed variables. This spectral decomposition provides researchers with a tool to trace the origins of information about shocks and other latent variables in structural macroeconomic models, thereby enhancing transparency in model estimation. I demonstrate the method's utility through applications to three recent studies.

Keywords: DSGE models, Frequency domain, Information content

JEL classification: C32, C51, C52, E32

Bank of Portugal, email: nikolay.iskrev@bportugal.pt

Additional material supporting this paper can be accessed at https://niskrev.github.io/siga_docs/

There is nothing like a latent variable to stimulate the imagination.
A. Goldberger, quoted by **?**

1 Introduction

The estimation of latent variables is a pervasive challenge in macroeconomic research, requiring the integration of theoretical models with empirical data. Examples abound and include endogenously determined variables such as potential output and natural rates of interest or unemployment, alongside a plethora of exogenous shocks driving business cycle fluctuations in modern macroeconomic models. The inherently unobservable nature of these variables necessitates estimating models that explicitly describe their joint dynamics with observable quantities. Correctly specified and accurately measured latent endogenous variables and structural shocks are key requirements for macroeconomic models to serve as effective tools for policy analysis and credible story-telling devices.

This paper demonstrates how to perform a frequency-domain decomposition of information about latent variables in dynamic economic models. The decomposition reveals both where in the frequency spectrum this information predominantly resides and the relative contributions of individual observed variables. The goal of the analysis is to enhance researchers' understanding of a model's implications regarding the sources of information about unobservable quantities. In doing so, the paper contributes to the emerging literature aimed at improving the transparency of structural estimation in macroeconomic research.

Understanding where information about estimated quantities of interest originates in the data is a key question in the estimation of structural models. Compared to reduced-form estimation, structural models make it particularly challenging to link specific features of the data, on the one hand, to individual estimated parameters or latent variables, on the other. This opacity makes it difficult to understand how modeling assumptions influence estimation results, thereby hindering readers' ability to assess the credibility of the research findings. Enhancing transparency about how information is derived from observed data helps mitigate this challenge, enabling readers who suspect model misspecification along certain dimensions to better understand its implications.[1] The frequency domain perspective is particularly relevant here, as researchers often hold differing views on which data frequencies can be adequately represented by a given model. For example, fitting data contaminated

[1]See **?** and the references therein for a broader perspective on this topic.

by high-frequency noise will distort the estimation of models that fail to account for discrepancies between model variables and observed series. Similarly, models designed to explain business cycle fluctuations may lack mechanisms to account for low-frequency variations in the data used to estimate them, leading to contamination of information from the lower end of the spectrum. A similar issue arises when models that do not account for seasonality are fitted to data exhibiting seasonal patterns.

The existing literature presents different and sometimes contradictory approaches for dealing with these challenges. For instance, there is no consensus among practitioners on whether to allow for measurement errors in commonly used series or how to handle long-term trends when estimating structural macroeconomic models. Greater transparency about how information from various parts of the spectrum contributes to the estimation of latent variables is thus essential. It enables readers, who may hold differing views on the adequacy of a model to represent the data, to assess the potential consequences of misspecification based on the relative importance of frequencies they suspect are contaminated for identifying those variables.

The work most closely related to this paper is **?**, where the question regarding the sources of information about latent variables is treated in the time domain. In that paper, the amount of information from observable variables about latent variables is quantified by comparing prior and posterior probability distributions and employing information-theoretic measures of uncertainty and information gain. Analysis in the time domain preserves information about the temporal order of observable data in relation to latent variables and enables the study of information transfer between variables with arbitrary temporal patterns. Specifically, one can evaluate how each observed variable contributes information from any subperiod of the sample. This facilitates readers' assessment of model misspecification consequences when they suspect certain observed series diverge from their model counterparts during specific sample periods.[2]

The work most closely related to this paper is **?**, where the question regarding the sources of information about latent variables is treated in the time domain. In that paper, the amount of information from observable variables about latent variables is quantified by comparing prior and posterior probability distributions

[2]One example where this could be useful is in estimating monetary models with data from periods when the zero lower bound on interest rates is binding and non-conventional monetary policy measures are in place. One approach in the literature addresses this issue by using the so-called "shadow interest rate" as an observed counterpart of the policy rate in theoretical models (see e.g. **?**). Estimated shadow rates are not constrained by the effective lower bound and typically coincide with observed policy rates when the bound is not binding. A potential concern with this approach is that information from the shadow rate series becomes contaminated during periods of substantial estimation uncertainty.

and employing information-theoretic measures of uncertainty and information gain. Analysis in the time domain preserves information about the temporal order of the observable data in relation to the latent variables and allows to study the transfer of information between variables with arbitrary temporal patterns. In particular, one can evaluate the contribution of information from any observed variable originating in any subperiod of the sample. This can facilitate the assessment of the consequences of model misspecification by readers who suspect that some observed series diverge from the respective model concepts during some part of the sample.[3] The information pertaining to the temporal order of variables is lost completely in the frequency domain. At the same time, it enables the decomposition of uncertainty and information into components at varying frequencies. This decomposition reveals both the extent and spectral location of uncertainty resolution for a given latent variable, as well as the relative contributions from different observed variables at various frequencies. Such insights into the distribution of uncertainty and information are not obtainable in the time domain. While time-domain analysis reveals misspecification with respect to specific sample periods, frequency domain analysis helps researchers understand the implications of using information from specific frequency bands that may be poorly represented by the estimated models. Thus, the time and frequency domain approaches complement each other.

It is important to emphasize that the analysis described in this paper does not require models to be solved, estimated, or transformed from the time to the frequency domain. It can be applied independently of the estimation method. In this respect, it resembles frequency-domain parameter identification analysis (see, e.g., **?**) or spectral variance decompositions (see, among many others, the handbook chapter by **?**). Although less common in the empirical literature, spectral methods for estimating and evaluating macroeconomic models have been advocated in several influential studies, including **?**, **?**, **?**, **?**, **?**, **?**, and **?**.

The paper is also related to a growing literature on the feasibility of recovering structural shocks using reduced form models. Building upon the work of **??** and **??**, most research on this topic has focused on the issue of invertibility (or fundamentalness) in structural vector autoregressions, i.e. whether shocks from general equilibrium

[3]One example of where this could be useful is the estimation of monetary models with data including the period when the zero lower bound on interest rates is binding and non-conventional monetary policy measures are in place. One approach in the literature for dealing with this issue is to use the so-called "shadow interest rate" as an observed counterpart of the policy rate in theoretical models (see e.g. **?**). Estimated shadow rates are not constrained by the effective lower bound and usually coincide with the observed policy rates in the period when the bound is not binding. A possible concern about this approach is that the information from the shadow rate series is contaminated in the part of the sample where there is a substantial estimation uncertainty.

models can be recovered from VAR residuals (see **?** and **?** for useful overviews of this literature). Conditions for invertibility have been analyzed by **?**, **?**, **?**, **?**), while tests for non-invertibility of structural VARs are developed in **?** and **?**. Invertibility issues that are specific to DSGE models with news shocks are discussed in **?** and **?**. More recently, **?** and **?** have proposed measures of the degree of non-invertibility that quantify the discrepancies between true shocks and shocks obtained using non-fundamental VARs.[4] In another recent paper **?** distinguish between invertibility and what they call "recoverability" – the latter defined as the feasibility of recovering structural shocks from all leads and lags of the observables variables. They argue that recoverability is often more relevant for applied research and provide a necessary and sufficient condition for checking shock recoverability in linear models.

Like that literature, the analysis in this paper can be used to assess whether shocks are recoverable from a given set of observed variables. Additionally, as in **?** and especially **?**, it provides a measure of the degree of recoverability of any individual shock or endogenous latent variable. The proposed spectral measures of information gain are defined for each unobserved variable, quantifying how much prior uncertainty about it, within a given frequency band, is resolved by observing specific model variables.

While existing research on invertibility focuses on the usefulness of VAR-based tools for the empirical validation of structural models, the analysis presented here aims to explore how structural macroeconomic models characterize the transfer of information between observed and unobserved variables across different frequencies. Identifying the principal sources of information is of primary interest, rather than measuring the total information about a given shock or endogenous latent variable. To this end, I define and apply measures of frequency band-specific conditional information gains, which quantify the additional information contributed by a subset of variables at specific frequency bands, given the information contained in the remaining observed variables. As demonstrated in the application section, the conclusions can vary significantly depending on the choice of conditional variables.

The remainder of the paper is organized as follows. Section 2 reviews relevant information-theoretic and frequency domain concepts, defines measures of information gains from observables with respect to latent variables, and demonstrates how these measures can be evaluated for linear Gaussian DSGE models. It also shows how the measures can be used to decompose information about latent variables across frequencies and observed variables. This decomposition identifies the primary sources

[4]Simulation evidence that non-invertible VARs may in some cases produce good approximations of the true structural shocks is presented in **?** and **?**.

of information about any latent variable of interest, enhancing the transparency of structural macroeconomic model estimation. Section 3 illustrates the methodology through three applications. The first is a small-scale New-Keynesian model used by **?** to study monetary policy shocks that generate neo-Fisherian dynamics – shocks that move interest rates and inflation in the same direction over the short run. The second is a medium-scale New Keynesian model estimated by **?** to assess the importance of investment shocks in driving business cycle fluctuations. The third considers another medium-scale New Keynesian model presented in **?** to illustrate their method for augmenting macroeconomic models with higher-order belief dynamics. In all applications, I examine the sources of information about structural shocks. These examples illustrate different aspects of the proposed information decomposition while demonstrating its utility in increasing model estimation transparency. Section 4 concludes. An Online Appendix provides additional model specifications and results.

2 Methodology

This section has three objectives. First, it introduces basic information-theoretic concepts and defines a measure of information gain for variables with a multivariate complex Gaussian distribution. Second, it reviews key properties of the spectral representation of a stationary Gaussian vector process and presents frequency domain measures of information gain. Third, it demonstrates how to apply these measures to DSGE models to evaluate the information contributions of observed variables with respect to latent variables across frequencies.

2.1 Quantifying information gains

Consider a $(n_{\boldsymbol{y}} + 1)$-dimensional random vector $\boldsymbol{z} = [\boldsymbol{y}', x]'$ with joint probability density function $f(\boldsymbol{y}, x)$. How much information about x is gained by observing a realization of $\boldsymbol{y}$? Information theory provides the framework and tools to answer such questions. Entropy measures the uncertainty associated with a random variable, while mutual information quantifies the information shared between two random variables. Formally, if $f(x)$ is the marginal probability density function of x with support $\boldsymbol{S}_x$, the entropy $\mathrm{H}(x)$ of $f(x)$ is defined as

$$\mathrm{H}(x) = -\int_{\boldsymbol{S}_x} f(x) \ln\left(f(x)\right) dx = -\,\mathrm{E} \ln f(x). \tag{2.1}$$

The amount of information about x is measured by the reduction in uncertainty – that is, the entropy $\mathrm{H}(x)$ – relative to some base distribution. The mutual information between random variables $\boldsymbol{y}$ and x is defined as

$$\mathrm{I}(\boldsymbol{y}, x) = \int_{\boldsymbol{S_y}} \int_{\boldsymbol{S}_x} f(\boldsymbol{y}, x) \ln \frac{f(\boldsymbol{y}, x)}{f(\boldsymbol{y}) f(x)} d\boldsymbol{y} dx \tag{2.2}$$

where $\boldsymbol{S_y}$ is the support of $\boldsymbol{y}$. The information interpretation of (2.2) follows from the fact that it can be expressed in terms of entropy as

$$\mathrm{I}(\boldsymbol{y}, x) = \mathrm{H}(x) - \mathrm{H}(x|\boldsymbol{y}). \tag{2.3}$$

where $\mathrm{H}(x|\boldsymbol{y}) = -\operatorname{E} \ln f(x|\boldsymbol{y})$ is the entropy of the conditional probability density function of x given $\boldsymbol{y}$. Thus, $\mathrm{I}(\boldsymbol{y}, x)$ represents the reduction in uncertainty about x from observing $\boldsymbol{y}$.[5] As shown in **?**, $\mathrm{H}(x) \geq \mathrm{H}(x|\boldsymbol{y})$ with equality if and only if $f(\boldsymbol{y}, x) = f(\boldsymbol{y}) f(x)$. Therefore, unless $\boldsymbol{y}$ and x are independent, observing $\boldsymbol{y}$ provides information about x. For a partition of $\boldsymbol{y}$ into two sub-vectors $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, the conditional mutual information of x and $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ can be expressed as

$$\mathrm{I}(\boldsymbol{y}_1, x|\boldsymbol{y}_2) = \mathrm{H}(x|\boldsymbol{y}_2) - \mathrm{H}(x|\boldsymbol{y}_1, \boldsymbol{y}_2) \tag{2.4}$$

The conditional mutual information measures the additional reduction in uncertainty about x achieved by observing $\boldsymbol{y}_1$, given that $\boldsymbol{y}_2$ is already observed. Now, let the joint density function $f(\boldsymbol{y}, x)$ be the $(n_{\boldsymbol{y}} + 1)$-dimensional complex Gaussian distribution,

$$\mathcal{N}_{\mathcal{C}} \left(\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_{yy}} & \boldsymbol{\Sigma_{y}}_{x} \\ \boldsymbol{\Sigma}_{x\boldsymbol{y}} & \Sigma_{xx} \end{pmatrix} \right) \tag{2.5}$$

Then, both the marginal distribution of x and the conditional distribution of x given $\boldsymbol{y}$ are univariate complex Gaussian distributions. Their covariances are given by Σ_{xx} and $\boldsymbol{\Sigma}_{x|\boldsymbol{y}} = \Sigma_{xx} - \boldsymbol{\Sigma}_{x\boldsymbol{y}} \boldsymbol{\Sigma}_{\boldsymbol{yy}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{y}x}$, respectively. Using this result, we can show that the mutual information of $\boldsymbol{y}$ and x is

$$\mathrm{I}(\boldsymbol{y}, x) = \mathrm{H}(x) - \mathrm{H}(x|\boldsymbol{y}) = .5 \ln \left(\frac{\Sigma_{xx}}{\boldsymbol{\Sigma}_{x|\boldsymbol{y}}} \right) \tag{2.6}$$

From $\mathrm{H}(x) \geq \mathrm{H}(x|\boldsymbol{y})$ it follows that mutual information is positive $\boldsymbol{y}$ and x are dependent, and zero when they are independent. In the case of perfect dependence,

[5] This is analogous to measuring the information gain in the posterior compared to the prior distribution of parameters in Bayesian analysis, see **?**.

where there exists a one-to-one function g such that $x = g(\boldsymbol{y})$, observing $\boldsymbol{y}$ is equivalent to observing x. This results in $\boldsymbol{\Sigma}_{x|\boldsymbol{y}} = 0$ and $\mathrm{I}(\boldsymbol{y}, x) = \infty$. A common practice is to normalize this measure to the interval $[0,\ 1]$. This can be achieved using the following monotonous increasing transformation (see e.g. **?** or **?**)

$$\mathrm{I}^*(\boldsymbol{y}, \boldsymbol{z}) = 1 - \exp\left(-2\mathrm{I}(\boldsymbol{y}, \boldsymbol{z})\right) \tag{2.7}$$

Applying this transformation to (2.6) results in the following measure of information gain:

$$\mathrm{IG}_{\boldsymbol{y}\to x} = \left(\frac{\Sigma_{xx} - \boldsymbol{\Sigma}_{x|\boldsymbol{y}}}{\Sigma_{xx}}\right) \times 100, \tag{2.8}$$

The interpretation of $\mathrm{IG}_{\boldsymbol{y}\to x}$ is the following: it measures the percent reduction in uncertainty about x from observing vector $\boldsymbol{y}$, relative to the unconditional (prior) uncertainty about x. Similarly, for a partition of $\boldsymbol{y}$ into $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, the conditional information gain of $\boldsymbol{y}_1$ with respect to x given $\boldsymbol{y}_2$ is defined as

$$\mathrm{IG}_{\boldsymbol{y}_1\to x|\boldsymbol{y}_2} = \left(\frac{\boldsymbol{\Sigma}_{x|\boldsymbol{y}_2} - \boldsymbol{\Sigma}_{x|\boldsymbol{y}}}{\Sigma_{xx}}\right) \times 100, \tag{2.9}$$

The interpretation of $\mathrm{IG}_{\boldsymbol{y}_1\to x|\boldsymbol{y}_2}$ is the following: it measures the percent reduction in the remaining uncertainty about x achieved by observing $\boldsymbol{y}_1$ after $\boldsymbol{y}_2$ is known, relative to the unconditional uncertainty about x.

2.2 Information gains in the frequency domain

Let $\boldsymbol{z}_t \in \mathbb{R}^{n_{\boldsymbol{z}}}$ for $t \in \mathbb{Z}$ be $n_{\boldsymbol{z}}-$dimensional stationary Gaussian time series with

$$\begin{aligned} \mathrm{E}\,\boldsymbol{z}_t &= \mathbf{0} \qquad t \in \mathbb{Z} && (2.10)\\ \mathrm{cov}\left(\boldsymbol{z}_t, \boldsymbol{z}_{t-h}\right) &= \boldsymbol{\Gamma}(h) \qquad t, h \in \mathbb{Z} && (2.11) \end{aligned}$$

When $\boldsymbol{Z} = [\boldsymbol{z}_1', \boldsymbol{z}_2', \ldots, \boldsymbol{z}_T']'$ represents a $T \times n_{\boldsymbol{z}}-$dimensional realization the process, the joint, marginal, and conditional distributions of any subset of $\boldsymbol{Z}$ components are Gaussian. Thus, the time-domain information gain measures from the previous section can directly quantify the information gained about any realization of a $\boldsymbol{z}$ component from observing a sample of realizations of other process components (see ?).

In the frequency domain, the information gains analysis proceeds by applying the discrete Fourier transform to the values of $\boldsymbol{Z}$:

$$Z(\omega_j) = (2\pi T)^{-1/2} \sum_{t=1}^{T} \boldsymbol{z}_t e^{-it\omega_j} \tag{2.12}$$

for the Fourier frequencies $\omega_j = 2\pi j/T$, where $j \in \{j \in \mathbb{Z} : -\pi < 2\pi j/T \leq \pi\}$.

The linearity of the discrete Fourier transform preserves joint Gaussianity. Moreover, $Z(\omega_j)$ behave asymptotically as independent complex Gaussian random variables with zero mean and covariance matrix $f(\omega_j)$, where $f_{\boldsymbol{zz}}(\omega) \in \mathbb{C}^{n_{\boldsymbol{z}} \times n_{\boldsymbol{z}}}$ represents the spectral density matrix of $\boldsymbol{z}(t)$ at frequency ω (see **?**, Theorem 4.4.1),

$$f_{\boldsymbol{zz}}(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-ih\omega} \tag{2.13}$$

The asymptotic independence of the Fourier coefficients $Z(\omega_j)$ across frequencies enables frequency-specific information gain analysis. Specifically, there is (asymptotically) no information about a series component at frequency ω_j derived from components at any other frequency $\omega_l,\ l \neq j$. The complex Gaussian distribution further allows information analysis at each frequency ω using the information gain measures from Section 2.1. To be more concrete, consider partitioning $\boldsymbol{z}_t$ into a $n_{\boldsymbol{y}}-$dimensional vector $\boldsymbol{y}_t$ and a scalar x_t, with $\boldsymbol{y}(\omega)$ and $x(\omega)$ representing their respective discrete Fourier transforms at frequency $\omega \in (-\pi, \pi]$. The spectral density matrix of $[\boldsymbol{y}_t', x_t]'$ is given by

$$f_{\boldsymbol{zz}}(\omega) = \begin{bmatrix} f_{\boldsymbol{yy}}(\omega) & f_{\boldsymbol{y}x}(\omega) \\ f_{x\boldsymbol{y}}(\omega) & f_{xx}(\omega) \end{bmatrix} \tag{2.14}$$

and the frequency-specific information gain of $\boldsymbol{y}(\omega)$ with respect to $x(\omega)$ is[6]

$$\mathrm{IG}_{\boldsymbol{y}\to x}(\omega) = \left(\frac{f_{xx}(\omega) - f_{x|\boldsymbol{y}}(\omega)}{f_{xx}(\omega)} \right) \times 100 \tag{2.15}$$

where $f_{x|\boldsymbol{y}}(\omega) = f_{xx}(\omega) - f_{x\boldsymbol{y}}(\omega) f_{\boldsymbol{yy}}^{-1}(\omega) f_{\boldsymbol{y}x}(\omega)$ is the partial spectrum of x given $\boldsymbol{y}$ (**?**).

Similarly, for a partition of $\boldsymbol{y}_t$ into $\boldsymbol{y}_{1t}$ and $\boldsymbol{y}_{2t}$ with respective discrete Fourier transforms $\boldsymbol{y}_1(\omega)$ and $\boldsymbol{y}_2(\omega)$, the frequency-specific conditional information gain from

[6] As in the time domain, there is a natural connection between the measures of spectral information gains and the frequency domain version of mutual information, see **?** and **?**.

$\boldsymbol{y}_1(\omega)$ about $x(\omega)$ given $\boldsymbol{y}_2(\omega)$ is

$$\text{IG}_{\boldsymbol{y}_1\to x|\boldsymbol{y}_2}(\omega) = \left(\frac{f_{x|\boldsymbol{y}_2}(\omega) - f_{x|\boldsymbol{y}}(\omega)}{f_{xx}(\omega)}\right) \times 100 \tag{2.16}$$

The interpretation of $\text{IG}_{\boldsymbol{y}\to x}(\omega)$ and $\text{IG}_{\boldsymbol{y}_1\to x|\boldsymbol{y}_2}$ is the same as before, except that now information is defined in terms of the reduction of uncertainty about x at a given frequency ω due to information in $\boldsymbol{y}$ (or conditionally, in $\boldsymbol{y}_1$) at the same frequency.

In practice, we are usually interested not in a single frequency but rather in a band of frequencies, such as low, business cycle, or high frequencies. Measures of frequency band-specific information gain can be obtained by replacing the frequency-specific spectrum and conditional spectrum of x in (2.15) and (2.16) with their integrated versions,

$$\text{IG}_{\boldsymbol{y}\to x}(\boldsymbol{\omega}) = \left(\frac{f_{xx}(\boldsymbol{\omega}) - f_{x|\boldsymbol{y}}(\boldsymbol{\omega})}{f_{xx}(\boldsymbol{\omega})}\right) \times 100 \tag{2.17}$$

$$\text{IG}_{\boldsymbol{y}_1\to x|\boldsymbol{y}_2}(\boldsymbol{\omega}) = \left(\frac{f_{x|\boldsymbol{y}_2}(\boldsymbol{\omega}) - f_{x|\boldsymbol{y}}(\boldsymbol{\omega})}{f_{xx}(\boldsymbol{\omega})}\right) \times 100 \tag{2.18}$$

where $\boldsymbol{\omega} = \{\omega : \omega \in [\underline{\omega}, \overline{\omega}] \cup [-\overline{\omega}, -\underline{\omega}]\}$ denotes the frequency band of interest, $f_{xx}(\boldsymbol{\omega}) = \int_{\omega\in\boldsymbol{\omega}} f_{xx}(\omega)\text{d}\omega$, and $f_{x|\boldsymbol{y}}(\boldsymbol{\omega}) = \int_{\omega\in\boldsymbol{\omega}} f_{x|\boldsymbol{y}}(\omega)\text{d}\omega$. The interpretation remains the same, except that now the uncertainty and information about x are with respect to the frequency band $\boldsymbol{\omega}$. Note that $\text{IG}_{\boldsymbol{y}\to x}(\boldsymbol{\omega})$ can be written also as

$$\text{IG}_{\boldsymbol{y}\to x}(\boldsymbol{\omega}) = \int_{\omega\in\boldsymbol{\omega}} \text{IG}_{\boldsymbol{y}\to x}(\omega)\frac{f_{xx}(\omega)}{f_{xx}(\boldsymbol{\omega})}\text{d}\omega \tag{2.19}$$

The information gain for a frequency band $\boldsymbol{\omega}$ is thus a weighted sum of the frequency-specific information gains, where the weights equal each frequency's contribution to the total variance of x in $\boldsymbol{\omega}$. Similarly, the conditional information gain (2.18) is a weighted sum of the frequency-specific conditional information gains.

A special case of the band-specific information gain is when $\boldsymbol{\omega}$ spans the full spectrum, with $\underline{\omega} = 0$ and $\overline{\omega} = \pi$. Let $\overline{\boldsymbol{\omega}} = \{\omega : \omega \in [0, \pi] \cup (0, -\pi]\}$. In this case, the information gain takes the form:

$$\text{IG}_{\boldsymbol{y}\to x}(\overline{\boldsymbol{\omega}}) = \left(\frac{\text{var}(x_t) - \text{var}(x_t|\boldsymbol{y}_{t-\tau}, \tau \in \mathbb{Z})}{\text{var}(x_t)}\right) \times 100 \tag{2.20}$$

Thus, beyond its obvious frequency-domain interpretation, this measure also has a time-domain interpretation: it represents the percent reduction of the unconditional variance of x_t from observing the infinite sequence of past, present, and future values of $\boldsymbol{y}_t$. Similarly, when evaluated over the full spectrum, the conditional information gain from $\boldsymbol{y}_1$ about x given $\boldsymbol{y}_2$ is

$$\text{IG}_{\boldsymbol{y}_1 \to x|\boldsymbol{y}_2}(\overline{\boldsymbol{\omega}}) = \left(\frac{\text{var}\,(x_t|\boldsymbol{y}_{2t-\tau}, \tau \in \mathbb{Z}) - \text{var}(x_t|\boldsymbol{y}_{t-\tau}, \tau \in \mathbb{Z})}{\text{var}(x_t)} \right) \times 100 \tag{2.21}$$

The interpretation of (2.21) is the following: it measures the percent reduction in the remaining uncertainty about x_t achieved by observing the infinite sequence of past, present, and future values of $\boldsymbol{y}_1$, given that the corresponding sequence of $\boldsymbol{y}_2$ is already known, relative to the unconditional uncertainty about x_t

Example To fix ideas, consider the following example. A latent variable of interest x_t follows a stationary AR(1) process:

$$x_t = \alpha x_{t-1} + \varepsilon_{xt} \tag{2.22}$$

The observed variables y_{1t} and y_{2t} are noisy measures of x_t, given by

$$y_{1t} = x_t + e_{1t}, \tag{2.23}$$

$$y_{2t} = x_t + e_{2t} \tag{2.24}$$

where $e_{1t} = \beta e_{1t-1} + \sqrt{1-\beta^2}\varepsilon_{1t}$, $e_{2t} = \varepsilon_{2t}$ and ε_{xt}, ε_{1t}, and ε_{2t} are all i.i.d with mean 0 and variance 1.

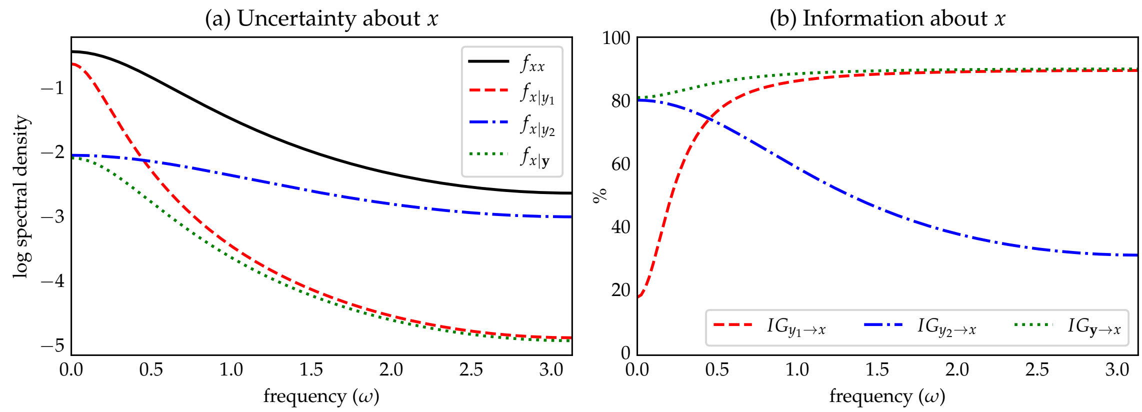

Figure 1: Frequency-specific uncertainty and information about the latent variable x in the system described by (2.22) - (2.24). Panel (a) shows prior and posterior spectral densities of x. Panel (b) shows the respective information gains.

Let $\alpha = .5$, $\beta = .9$. Panel (a) of Figure 1 show the logs of the prior (unconditional) and posterior (conditional) spectral densities of x. Each point on a spectral density curve represents the contribution of the corresponding frequency to the variance of x. The area under each curve represents the total variance of x under different information scenarios: observing nothing, observing either y_1 or y_2, or observing both $\boldsymbol{y} = [y_1, y_2]$. Since both variables are informative about x, the posterior spectral densities lie below the prior density. The lowest uncertainty occurs when both y_1 and y_2 are observed. The area between a prior and posterior spectral density curves represents the reduction of uncertainty – that is, the information about x.[7]

Panel (b) displays the frequency-specific information gains, defined as the percent reduction in uncertainty. The figure reveals that y_1 provides relatively less information about low frequencies compared to y_2, but more information across the rest of the spectrum, particularly at very high frequencies. This pattern stems from the different spectral profiles of the noise terms: while $\text{var}\,(e_{1t}) = \text{var}\,(e_{2t}) = 1$ implies equal total noise in y_1 and y_2, the persistent AR(1) process e_{1t} primarily contaminates very low frequencies, whereas the white noise e_2 contributes uniformly across all frequencies.

The conditional information gains can be determined by comparing the marginal to the joint information gain: $\text{IG}_{y_i \to x|y_j} = \text{IG}_{\boldsymbol{y} \to x} - \text{IG}_{y_j \to x}$ for $i, j \in [1, 2]$. At frequencies close to zero, $\text{IG}_{\boldsymbol{y} \to x} \approx \text{IG}_{y_2 \to x}$, indicating that observing y_1 provides little additional information about the low frequencies of x when y_2 is known. Similarly, given y_1, observing y_2 adds little or no information about high frequencies of x. The smaller conditional information compared to marginal information implies that some information about x is available from either variable – once one is observed, information from parts of the other's spectrum becomes redundant.

For a trivial example of a situation where the conditional information exceeds the marginal one, consider the case when e_1 is observable. Observing e_1 alone provides no information about x. However, observing both y_1 and e_1 is equivalent to observing x directly, yielding 100% information gain. Therefore, the conditional information gains of both y_1 and e_1 exceed their marginal gains.

2.3 DSGE models

A linearized DSGE model can be written as a recursive equilibrium law of motion in the following system of equations:

$$\boldsymbol{y}_t = \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{v}_{t-1} + \boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{u}_t \tag{2.25}$$

$$\boldsymbol{v}_t = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{v}_{t-1} + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{u}_t \tag{2.26}$$

$$\boldsymbol{u}_t = \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{u}_{t-1} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\boldsymbol{\theta})\right) \tag{2.27}$$

[7]Since the figure is in logs, the area represents the log ratio between prior and posterior uncertainty.

where $\boldsymbol{y}_t$ is a $n_{\boldsymbol{y}}$-dimensional vector of observed variables, $\boldsymbol{v}_t$ is a $n_{\boldsymbol{v}}$-dimensional vector of endogenous state variables, $\boldsymbol{u}_t$ is a $n_{\boldsymbol{u}}$-dimensional vector of exogenous state variables, and $\boldsymbol{\varepsilon}_t$ is a $n_{\boldsymbol{u}}$-dimensional vector of exogenous shocks. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{G}$ are functions of the model's structural parameters, collected in the $n_{\boldsymbol{\theta}}$-dimensional vector $\boldsymbol{\theta}$.

In practice, researchers may be interested in various latent variables: endogenous variables like output gap, exogenous shocks such as total factor productivity (TFP), or innovations to exogenous shocks like TFP innovations. Using the notation from sections 2.1 and 2.2, the latent variable x_t corresponds to an element of $\boldsymbol{v}_t$, $\boldsymbol{u}_t$, or $\boldsymbol{\varepsilon}_t$, while $\boldsymbol{y}_t$ represents the vector of observed variables. Computing unconditional and conditional information gains requires the spectral and cross-spectral densities of x_t, $\boldsymbol{y}_t$, and individual elements of $\boldsymbol{y}$. These can be derived from the joint spectral density matrix of $\boldsymbol{z}_t = [\boldsymbol{y}_t', \boldsymbol{v}_t', \boldsymbol{u}_t', \boldsymbol{\varepsilon}_t']'$, given by (see **?**):

$$\boldsymbol{f}_{\boldsymbol{zz}}(\omega) = \frac{1}{2\pi}\boldsymbol{W}(\omega,\boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}(\boldsymbol{\theta})\boldsymbol{W}(\omega,\boldsymbol{\theta})^{*} \tag{2.28}$$

where

$$\boldsymbol{W}(\omega,\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{C}(\boldsymbol{\theta})e^{-i\omega} & \boldsymbol{D}(\boldsymbol{\theta}) & \mathbf{O}_{n_{\boldsymbol{y}},n_{\boldsymbol{u}}} \\ \boldsymbol{I}_{n_{\boldsymbol{v}}} & \mathbf{O}_{n_{\boldsymbol{v}},n_{\boldsymbol{u}}} & \mathbf{O}_{n_{\boldsymbol{v}},n_{\boldsymbol{u}}} \\ \mathbf{O}_{n_{\boldsymbol{u}},n_{\boldsymbol{v}}} & \boldsymbol{I}_{n_{\boldsymbol{u}}} & \mathbf{O}_{n_{\boldsymbol{v}},n_{\boldsymbol{u}}} \\ \mathbf{O}_{n_{\boldsymbol{u}},n_{\boldsymbol{y}}} & \mathbf{O}_{n_{\boldsymbol{u}},n_{\boldsymbol{u}}} & \boldsymbol{I}_{n_{\boldsymbol{u}}} \end{bmatrix} \times \tag{2.29}$$

$$\begin{bmatrix} \left(\boldsymbol{I}_{n_{\boldsymbol{v}}} - \boldsymbol{A}(\boldsymbol{\theta})e^{-i\omega}\right)^{-1}\boldsymbol{B}(\boldsymbol{\theta})\left(\boldsymbol{I}_{n_{\boldsymbol{u}}} - \boldsymbol{G}(\boldsymbol{\theta})e^{-i\omega}\right)^{-1} \\ \left(\boldsymbol{I}_{n_{\boldsymbol{u}}} - \boldsymbol{G}(\boldsymbol{\theta})e^{-i\omega}\right)^{-1} \\ \boldsymbol{I}_{n_{\boldsymbol{u}}} \end{bmatrix}$$

and the asterisk denotes matrix transposition and complex conjugation.

In business cycle research, it is typical to divide the spectrum into three non-overlapping intervals: business cycle frequencies with periodicity between 6 and 32 quarters (as is standard in the literature, for example **?**), and frequencies above and below that interval, labeled as low and high frequencies, respectively. Let $\boldsymbol{\omega}^{BC}$, $\boldsymbol{\omega}^{L}$, and $\boldsymbol{\omega}^{H}$ denote these frequency bands. The total information gain from $\boldsymbol{y}_t$ about x_t can be decomposed as follows:

$$\begin{aligned} \mathrm{IG}_{\boldsymbol{y}\to x}(\overline{\boldsymbol{\omega}}) &= \mathrm{IG}_{\boldsymbol{y}\to x}(\boldsymbol{\omega}^{L})\frac{f_{xx}(\boldsymbol{\omega}^{L})}{f_{xx}(\overline{\boldsymbol{\omega}})} + \mathrm{IG}_{\boldsymbol{y}\to x}(\boldsymbol{\omega}^{BC})\frac{f_{xx}(\boldsymbol{\omega}^{BC})}{f_{xx}(\overline{\boldsymbol{\omega}})} \\ &+ \mathrm{IG}_{\boldsymbol{y}\to x}(\boldsymbol{\omega}^{H})\frac{f_{xx}(\boldsymbol{\omega}^{H})}{f_{xx}(\overline{\boldsymbol{\omega}})} \end{aligned} \tag{2.30}$$

The total information gain is thus a weighted sum of band-specific information gains, where the weights equal each band's contribution to the total variance of x.

Decomposing information gains across frequency bands is feasible because the components within each band are mutually independent. However, due to the correlation among variables in $\boldsymbol{y}$, the overall information about x cannot be decomposed into independent contributions of individual observed variables. Instead, we can measure the marginal contribution from each observed variable y_i, as well as its conditional contribution given the information in other observed variables $\boldsymbol{y}_j \subset \boldsymbol{y}_{-i} \equiv \{\boldsymbol{y} \setminus y_i\}$. For any frequency band $\boldsymbol{\omega}$, the following decomposition holds:

$$\mathrm{IG}_{\boldsymbol{y}\to x}(\boldsymbol{\omega}) = \mathrm{IG}_{y_i \to x|\boldsymbol{y}_{-i}}(\boldsymbol{\omega}) + \mathrm{IG}_{\boldsymbol{y}_{-i}\to x}(\boldsymbol{\omega}) \tag{2.31}$$

The first term on the right-hand side represents information in y_i about x that is not contained in $\boldsymbol{y}_{-i}$. This includes both information unique to y_i (independent from $\boldsymbol{y}_{-i}$) and information that emerges from observing y_i together with $\boldsymbol{y}_{-i}$. Meanwhile, any information about x that is shared between y_i and $\boldsymbol{y}_{-i}$ is captured by the second term in (2.31).

Example (continued) Figure 2 illustrates the decomposition of the prior and

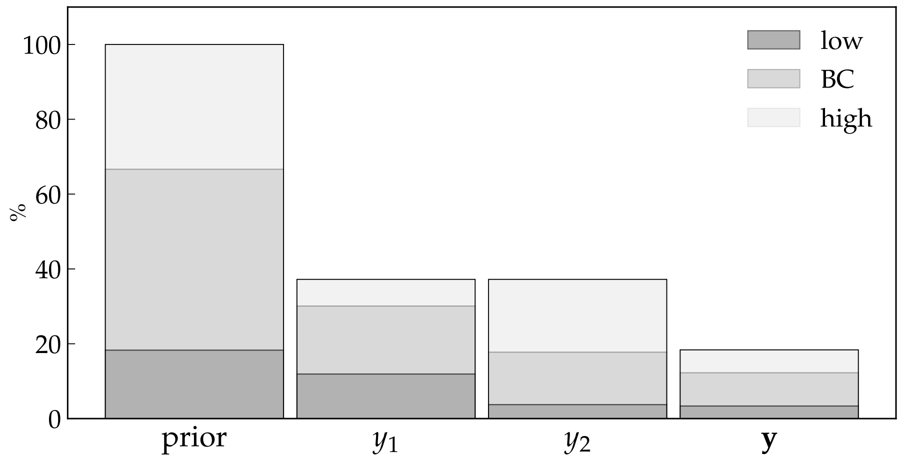

Figure 2: Contributions from the low, business cycle, and high frequencies to the prior and posterior uncertainty about x.

posterior uncertainty about x into contributions from the low, business cycle, and the high frequencies. Observing either y_1 or y_2 reduces uncertainty by the same amount, 63%. However, the contributions from different frequency bands vary significantly. When y_1 is observed, approximately 26% of the information gain originates from the high frequencies, compared to only 14% when y_2 is observed. Conversely, the low frequencies contribute 6% with y_1 and over 14% with y_2. For both variables, the business cycle frequencies provide the largest share of information, contributing 30% and 34% for y_1 and y_2, respectively. When both variables are observed, the total reduction in uncertainty increases to 82%, with contributions of 15%, 40%, and 27% from the low, business cycle, and the high frequencies, respectively.

3 Applications

In this section, I present three examples of application of the proposed method to estimated macroeconomic models. The first two applications involve small- and medium-scale New Kaynesian models taken from **?** and **?**. Considering these models allows me to illustrate different elements of the analysis in a complementary fashion. The model of **?** is much smaller, with only three observed variables, which makes it possible to present fully results regarding information interactions among those variables. This is not practicable in the case of the **?** model, where I present only selected results and leave the rest for the Appendix. Another important difference is that the **?** has more shocks than observables, and finding out how well each shock can be recovered is a relevant dimension of the analysis, in addition to investigating the main sources of information. This is not an issue in the second model, which, with its richer structure, larger number of shocks and observables, is much more representative of the medium-scale New Keynesian framework in the DSGE literature. The last example is another medium-scale New Kaynesian model taken from **?**. The model incorporates many of the features found in other estimated DSGE models, but dispenses with the usual assumption of rational expectations and common information about the state of the economy. Furthermore, in contrast to most of the literature, the model is estimated in the frequency domain using only the business-cycle frequencies. Therefore, it provides an opportunity to discuss the use and usefulness of the proposed methodology in applications where there are concerns about model misspecification in some parts of the spectrum, as is the case in **?**.

3.1 ?

? investigates the nature and empirical importance of monetary policy shocks that produce neo-Fisherian dynamics, i.e. move interest rates and inflation in the same direction over the short run. To that end, the author estimates a standard small-scale New-Keynesian model with price stickiness and habit formation, augmented with seven structural shocks. Full details about the model can be found in the original publication. Here I only describe those of its features that are directly relevant for the analysis which follows.

Firstly, three of the shocks are to monetary policy, which is described by the following policy rule:

$$\frac{1+I_t}{\Gamma_t} = \left[A\left(\frac{1+\Pi_t}{\Gamma_t}\right)^{\alpha_t}\left(\frac{Y_t}{X_t}\right)^{\alpha_y}\right]^{1-\gamma_I}\left(\frac{1+I_{t-1}}{\Gamma_{t-1}}\right)^{\gamma_I} e^{z_t^m}, \tag{3.1}$$

where I_t the nominal interest rate, Y_t is aggregate output, Π_t is the inflation rate, Γ_t is the inflation-target, X_t is a nonstationary productivity shocks, and z_t^m is a stationary interest-rate shock. The inflation target is defined as

$$\Gamma_t = X_t^m e^{z_t^{m2}}, \tag{3.2}$$

where X_t^m and z_t^{m2} are permanent and transitory components of the inflation target. It is assumes that X_t^m and X_t grow at a rates g_t^m and g_t, respectively.

There are two preference shocks affecting the lifetime utility function of the representative household, given by

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t e^{\xi_t} \left\{ \frac{\left[\left(C_t - \delta \tilde{C}_{t-1}\right)\left(1 - e^{\theta_t} h_t\right)^{\chi}\right]^{1-\sigma} - 1}{1-\sigma} \right\}, \tag{3.3}$$

where C_t is consumption, $\tilde{C}_t$ is the cross sectional average of consumption, h_t is hours worked, ξ_t is an intertemporal preference shock, and θ_t is a shock to labor supply.

In addition to X_t, there is also a stationary productivity shock z_t, which affects the production technology according to

$$Y_t = e^{z_t} X_t h_t^{\alpha}, \tag{3.4}$$

The five stationary shocks (ξ_t, θ_t, z_t, z_t^m, and z_t^{m2}) and the growth rates of the two non-stationary shocks (g_t and g_t^m) are all assumed to follow first-order autoregressive processes.

? estimates the model using quarterly US data on three variables: per capita output growth ($\triangle y_t$), the interest-rate-inflation differential ($r_t = i_t - \pi_t$), and the change in the nominal interest rate ($\triangle i_t$). All variables are assumed to be observed with measurement errors, modeled as Gaussian i.i.d. processes. Thus, there are ten independent sources of randomness in the data and only three observables. Clearly, not all, if any, of the latent variables can be recovered fully. The purpose of the remainder of this section is to determine how well each structural shock can be recovered and where in the spectrum most of the information comes from, as well as what are the information contributions of different observed variables overall and across different frequency bands.

3.1.1 Information decomposition across frequency bands

? solves the model by log-linear approximation of the equilibrium conditions around steady state. The linearity of the solution together with the assumption that the structural innovations and the measurement errors are Gaussian, implies that the joint distribution of (any subset of) the innovations, shocks, state and observed variables is also Gaussian. Therefore, the analysis of information gains can be conducted using the measures introduced in Section 2. In the analysis which follows I fix the parameter values at the mean of posterior distribution reported in ?, Table 5.

Table 1 presents the total information gains for the seven shocks and their decompositions into gains from three frequency bands - low, business cycle and high frequencies, with periodicities of more than 32 quarters, between 6 and 32 quarters and less than 6 quarters, respectively. The results show that none of the shocks can be fully recovered from the observed variables. The largest reduction of uncertainty is with respect to the intertemporal preference shock (ξ_t) – by about 93%, and the permanent productivity shock (g_t) – by about 85%. The gains with respect to the three monetary policy-related shocks are between 15% and 18%. The least information is gained with respect to the labor supply (θ_t) and the transitory productivity shocks (z_t), with information gains for both of 1.8%.

Table 1: Information decomposition across frequency bands

		total	low	BC	high
ξ_t	preference	93.2	70.4 = 96.4 × 0.73	19.5 = 88.4 × 0.22	3.2 = 66.0 × 0.05
θ_t	labor supply	1.8	0.2 = 0.5 × 0.33	1.1 = 2.3 × 0.48	0.5 = 2.9 × 0.18
z_t	transitory productivity	1.8	0.2 = 0.5 × 0.32	1.1 = 2.2 × 0.49	0.5 = 2.9 × 0.19
g_t	permanent productivity	83.5	9.3 = 94.9 × 0.10	32.3 = 87.1 × 0.37	42.0 = 78.9 × 0.53
z_t^m	transitory interest rate	15.5	0.1 = 0.9 × 0.12	3.2 = 7.9 × 0.41	12.2 = 25.7 × 0.47
z_t^{m2}	transitory trend inflation	16.5	5.8 = 12.7 × 0.46	9.7 = 23.1 × 0.42	1.0 = 8.3 × 0.12
g_t^m	permanent trend inflation	18.0	7.2 = 69.4 × 0.10	7.0 = 18.3 × 0.38	3.9 = 7.5 × 0.51

Note: Information gain (IG) measures the reduction of uncertainty (variance) about a shock due to observing all three observed variables, as a *percent* of the unconditional uncertainty of the shock. The contribution from each frequency band to the total IG is shown as the product of the IG for that band and the fraction of the total variance of the shock originating in each band. Thus, the units in the last three columns are $\% = \% \times \frac{\text{variance band}}{\text{variance total}}$.

Columns 3 to 5 of the table show the information gain contributions from each frequency band. Following the earlier discussion (see equation (2.30)), the total contribution in each case is shown as the product of two terms: the band-specific information gain, which measures the reduction of uncertainty as a percent of the uncertainty in that band, and the fraction of total uncertainty that originates in the given frequency band.

For six of the seven shocks uncertainty is concentrated in either low and business cycle frequencies, or high and business cycle frequencies. Specifically, in the first groups are the transitory trend inflation, transitory productivity, and the labor supply shocks. And in the second are the permanent productivity, transitory interest rate, and permanent trend inflation shocks. The one exception is the intertemporal preference shock for which the low frequencies are by far the main source of uncertainty. As can be expected, the largest gains generally come from parts of the spectrum where prior uncertainty is larger. There are some notable exceptions, however. In particular, note that even though the low frequency band accounts for only 10% of the uncertainty about the permanent trend inflation shock, the information gain contribution from that band is largest than the business cycle frequency band, which accounts for 38% of the uncertainty, and much larger than the contribution from the high frequency band, which accounts for more than half of the total uncertainty. This is due to the fact that a much larger fraction of the uncertainty in the low frequencies is resolved by information provided by the observed variables than is the case for the higher frequencies. Similarly, note that for the labor supply and transitory productivity shocks, because of the relatively larger information gains from the higher end of the spectrum, the information contributions from there is larger than from the low frequencies, even though the low frequencies account for a significantly larger fraction of the prior uncertainty.

3.1.2 Information contributions by variables

Table 2 shows the conditional information gains for each observed variable for the full spectrum and the three frequency bands. The largest contribution by far is from output growth ($\triangle y_t$) with respect to the permanent productivity shock. Note that the conditional information gain of 83.4% is almost equal to the total gain (all observables) of 83.5% for that shock (see Table 1). This implies that the other two variables - the interest rate-inflation differential (r_t) and the change in the nominal interest rate ($\triangle i_t$) alone reduce the uncertainty about the permanent productivity shock by only 0.1%. This result holds for the full spectrum and the individual frequency bands. Output growth contributes less information for the other shocks, compared to r_t or $\triangle i_t$. The contributions of these variables with respect to the two trend inflation shocks are similar, with r_t being relatively more informative for the

Table 2: Conditional contribution of information

	shock	total			low			BC			high		
		$\triangle y_t$	r_t	$\triangle i_t$	$\triangle y_t$	r_t	$\triangle i_t$	$\triangle y_t$	r_t	$\triangle i_t$	$\triangle y_t$	r_t	$\triangle i_t$
ξ_t	preference	0.3	26.8	7.2	0.0	26.4	0.8	0.1	0.5	4.1	0.1	0.0	2.3
θ_t	labor supply	0.1	0.1	1.1	0.0	0.0	0.0	0.1	0.0	0.6	0.0	0.0	0.5
z_t	transitory productivity	0.1	0.0	1.1	0.0	0.0	0.0	0.1	0.0	0.6	0.0	0.0	0.5
g_t	permanent productivity	83.4	0.8	5.7	9.3	0.0	0.1	32.2	0.6	2.8	41.9	0.2	2.8
z_t^m	transitory interest rate	2.2	1.5	9.0	0.0	0.1	0.0	0.4	0.9	0.4	1.8	0.5	8.5
z_t^{m2}	transitory trend inflation	1.7	13.0	8.2	0.1	5.5	4.3	1.1	7.3	3.7	0.5	0.2	0.2
g_t^m	permanent trend inflation	0.5	10.4	15.6	0.0	4.7	7.0	0.2	5.2	5.6	0.3	0.5	3.0

Note: The conditional contribution of information shows additional reduction of uncertainty about a shock, as a percent of the unconditional uncertainty of the shock, due to observing a variable given that the other two variables are also observed. The variables are: output growth ($\triangle y_t$), interest-rate-inflation differential (r_t), and the change in the nominal interest rate ($\triangle i_t$). Due to rounding in some cases the band-specific contributions do not add up to the total values.

transitory trend inflation shock, while $\triangle i_t$ is more informative for the permanent one. In addition, r_t contributes much more information than either $\triangle y_t$ or $\triangle i_t$ with respect to the preference shock, while $\triangle i_t$ is the most informative observable with respect to the transitory interest rate shock, and, marginally, for the labor supply and transitory productivity shocks.

The ranking of variables in terms of their total information contributions is determined by the relative size of the information gains in the part of the spectrum from where a given shock receives the most total information (see Table 1). In several cases, the ranking changes with the frequency band. For instance, $\triangle i_t$ contributes significantly more information than r_t with respect to the intertemporal preference shock in the BC and high frequencies. At the same time, r_t contributes the most information with respect to the transitory interest rate shock in the low and BC frequencies, in spite of being the least informative variable in the high frequencies and overall. Similarly, $\triangle y_t$ is the least informative variable overall with respect to the transitory trend inflation shock, but has the largest contribution in the high frequency band.

It is worth emphasizing that the information gains shown in Table 2 are from observing a given variable *conditional* on already having observed the other two variables. As the observed variables are obviously not mutually independent, it is conceivable that in some cases the contributions are small because different variables share common information with respect to those shocks. To help find out if and when that is the case, Table 3 shows the unconditional information gains, i.e. the percent reduction of uncertainty about a given shock due to observing only one variable at a

Table 3: Unconditional contribution of information

	shock	total			low			BC			high		
		$\triangle y_t$	r_t	$\triangle i_t$	$\triangle y_t$	r_t	$\triangle i_t$	$\triangle y_t$	r_t	$\triangle i_t$	$\triangle y_t$	r_t	$\triangle i_t$
ξ_t	preference	3.5	84.6	66.0	0.9	69.6	44.0	2.0	14.6	18.9	0.6	0.4	3.1
θ_t	labor supply	0.0	0.6	1.7	0.0	0.1	0.2	0.0	0.5	1.0	0.0	0.0	0.5
z_t	transitory productivity	0.0	0.6	1.6	0.0	0.1	0.1	0.0	0.5	1.0	0.0	0.0	0.5
g_t	permanent productivity	76.7	0.1	0.1	9.0	0.0	0.0	28.7	0.0	0.0	39.1	0.0	0.0
z_t^m	transitory interest rate	0.7	5.8	11.5	0.0	0.1	0.0	0.2	2.5	1.8	0.5	3.2	9.7
z_t^{m2}	transitory trend inflation	2.2	5.3	0.9	0.2	1.1	0.1	1.6	3.8	0.6	0.4	0.4	0.2
g_t^m	permanent trend inflation	1.8	0.4	6.8	0.1	0.0	2.5	0.9	0.3	1.3	0.8	0.1	3.0

Note: see the note to Table 2. The unconditional contribution of information shows the reduction of uncertainty about a shock due to observing a single variable at a time.

time.

The results reveal some notable differences between conditional and unconditional information gains. Most striking is the reduction in the contributions of the three observables with respect to the intertemporal preference shock. In particular, the information gains from r_t and $\triangle i_t$ change from, respectively, 85% and 66% unconditionally, to 27% and 7% conditionally. Similarly, the contribution of $\triangle y_t$ decreases from 3.5% to only 0.3%. This suggests that, to a large extent, the information in either one of the observable variables is not unique to them but is also contained in the other two. In other words, there is a significant degree of redundancy of the information about the intertemporal preference shock. Another, less striking, example of redundancy is the transitory interest rate shock, where the conditional information gains from r_t and $\triangle i_t$ are smaller than the unconditional ones.

Information redundancy is not the only possible consequence of the existing interdependence among observables. In the case of the permanent productivity shock, the conditional information gains for all variables are significantly larger than the unconditional ones. The same is true for the contributions of r_t and $\triangle i_t$ with respect to the permanent and transitory trend inflation shocks, as well as for the contribution of $\triangle y_t$ with respect to the transitory interest rate shock. In all of these cases there is a positive information complementarity instead of information redundancy, that is, information increases when variables are observed together.

Following **?**, the degree of information complementarity between variables can be measured by comparing the joint information gain with respect to a shock to the individual gains. Specifically, the information complementarity between variables y_1

and y_2 conditional on variables $\boldsymbol{y}_3 \subset \{\boldsymbol{y} \setminus \boldsymbol{y}_{12}\}$ at frequency band $\boldsymbol{\omega}$ is defined as:

$$\mathrm{IC}_{\boldsymbol{y}_{12}\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) = \frac{\mathrm{IG}_{\boldsymbol{y}_{12}\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega})}{\mathrm{IG}_{y_1\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) + \mathrm{IG}_{y_2\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega})} - 1. \tag{3.5}$$

Negative values indicate negative complementarity, or information redundancy, between y_1 and y_2, and positive values indicate positive complementarity between the two variables. Since the information gain is non-negative, we have $\mathrm{IC}_{\boldsymbol{y}_{12}\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) \geq -1/2$, with equality when y_1 and y_2 are (conditionally on $\boldsymbol{y}_3$) functionally dependent, in which case $\mathrm{IG}_{\boldsymbol{y}_{12}\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) = \mathrm{IG}_{y_1\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) = \mathrm{IG}_{y_2\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega})$. A lack of information complementarity, i.e. $\mathrm{IC}_{\boldsymbol{y}_{12}\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) = 0$ occurs when y_1 and y_2 are (conditionally on $\boldsymbol{y}_3$) independent, and hence $\mathrm{IG}_{\boldsymbol{y}_{12}\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) = \mathrm{IG}_{y_1\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega}) + \mathrm{IG}_{y_2\rightarrow x|\boldsymbol{y}_3}(\boldsymbol{\omega})$. Note that the conditioning could be with respect to any subset of observables, including the empty set, in which case we have unconditional complementarity between y_1 and y_2.

Figure 3 shows the unconditional and conditional information complementarities between all pairs of variables. The results are shown for the full spectrum as well as the three frequency bands. As already anticipated, the strongest complementarity overall is between r_t and $\triangle i_t$, and is negative for all shocks except the permanent and transitory trend-inflation shocks. Both unconditionally and conditionally the degree of complementarity tends to be significantly lower in the higher frequencies. Conditioning on the third observable in most cases preserves the sign of complementarity and reduces the magnitude. There are some notable exceptions to this pattern, however. For instance, the degree of complementarity between r_t and $\triangle i_t$ increases when conditioning on $\triangle y_t$,

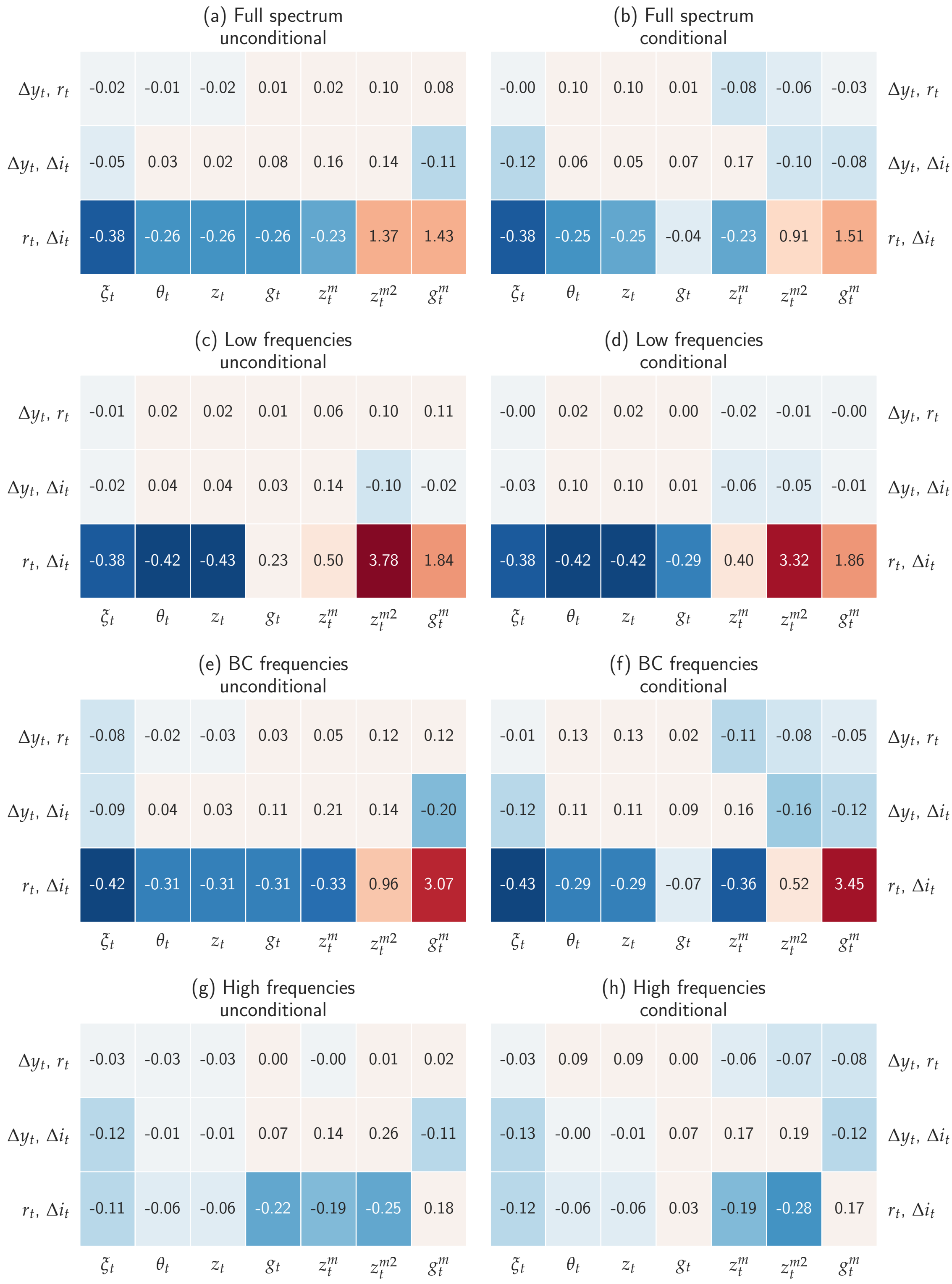

Figure 3: Pairwise information complementarity between observables with respect to shocks.

especially in the business cycle frequencies. Furthermore, the complementarity between the same variables with respect to the transitory productivity shock changes signs when conditioning on y_t, from positive to negative in the low frequencies, and from negative to positive in the high frequencies. At the same time, when evaluated over the full spectrum, the complementarity is strongly negative unconditionally and only weekly so, conditionally.

3.1.3 Information gains in the time domain

The time domain version of the full spectrum information gain measure (see equation (2.20)) is given by:

$$\mathrm{IG}_{\boldsymbol{Y}_T \to x_t} = \left(\frac{\mathrm{var}(x_t) - \mathrm{var}(x_t | \boldsymbol{Y}_T)}{\mathrm{var}(x_t)} \right) \times 100, \tag{3.6}$$

where $1 \leq t \leq T$ and $\boldsymbol{Y}_T = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\}$. The difference between the two measures is that, in the frequency domain, the information for any given x_t stems from the infinite past and future values of the observable variables. Therefore, for a given set of observed variables, the total amount of information is invariant to the temporal location of the latent variable. In contrast, in the time domain, it matters where the location of t is, relative to the beginning and the end of the sample. Thus, the value of time domain measure changes with t and is bounded from above by the value of the full spectrum frequency domain measure.

Figure 4 compares the time and frequency domain information gains for the seven shocks in the model. Specifically, it shows the ratio of the time domain to the frequency domain measure for all values of t in a sample of $T = 255$ observations, which is the sample size in **?**. The results show that for most values of t the time and frequency domain information gains coincide. As anticipated, differences occur only at the beginning and end of the sample. For all shocks except the transitory trend inflation shock, for which convergence is somewhat slower, there are about ten observations or fewer on either end of the sample where the time domain information gains are smaller than the frequency domain ones.

3.1.4 Discussion of the results

As already noted, having more sources of uncertainty than the number of observed variables necessarily implies that the latent variables in the model cannot all be recovered fully. At the same time, as the results presented in Section 3.1.1 show, some shocks in the **?** model are significantly better recoverable than others. The goal

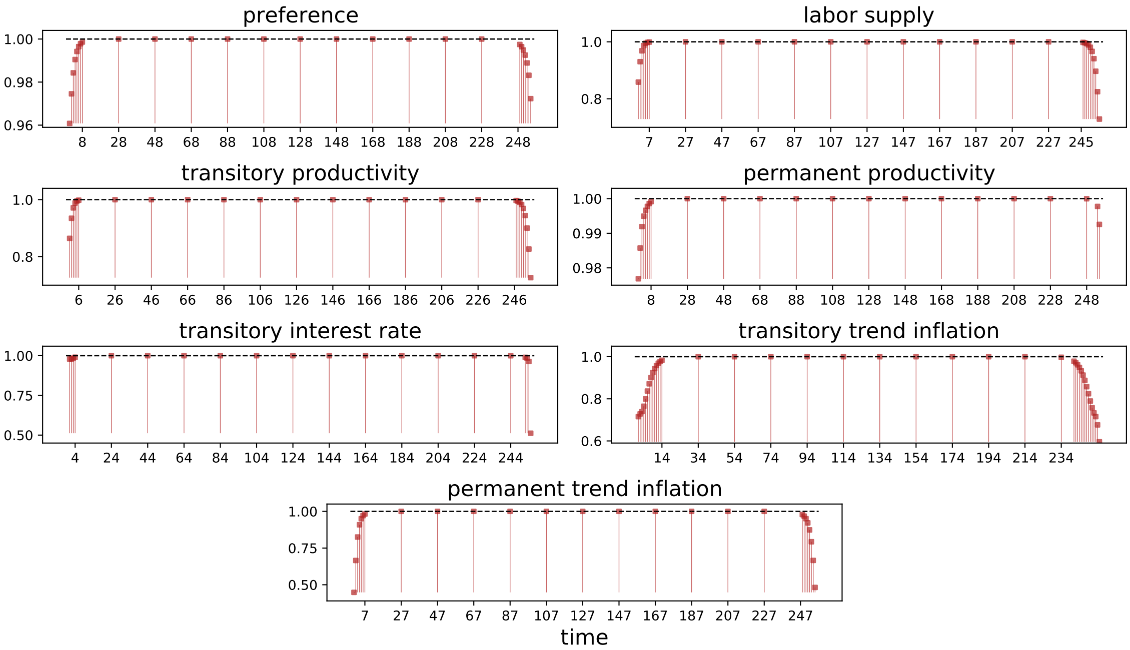

Figure 4: Total information gains in the time domain relative to the frequency domain (full spectrum).

of this section is to develop a further understanding of these findings.

A natural question to ask is: why are the information gains with respect to the intertemporal preference and permanent productivity shocks so much larger than the gains for the remaining shocks, and in particular compared to those with respect to the labor supply and transitory productivity shocks? Intuitively, the amount of information one or more variables contain about another variable depends on the strength of their mutual dependence.[8] Furthermore, an insight gained from the frequency domain perspective is that the interactions need to be strong in the parts of the spectrum that are mainly responsible for the uncertainty of the latent variable. In addition, the extent to which information from multiple sources accumulates, in turn, depends on how interdependent they are among themselves. For instance, variables that are functionally dependent on other observed variables provide no useful information.[9]

Consider the intertemporal preference shock (ξ_t). According to the posterior mean estimates reported in **?**, Table 5, this shock is significantly more persistent and volatile than all other shocks. In particular, its volatility is an order of magnitude larger than the volatilities of all other shocks except the permanent productivity shock

[8]In fact, the mutual information coefficient is commonly used to measure and test for statistical dependence between random variables (see e.g. **?**, **?**, and **?**).

[9]An example of this is output growth in the model estimated by **?**, see **?** for details.

(g_t). The high degree of persistence explains why most of the uncertainty about ξ_t is concentrated in the lower end of the spectrum, as shown in Table 1. Furthermore, as seen from the same table, most of the uncertainty in the low frequencies is resolved by the information contained in the observed variables, which suggests that there are strong interactions between ξ_t and (some of) those variables. Since, in the present context, the variables have a clear causal direction, i.e. from shocks to endogenous variables, a natural way of describing their interactions is in terms of the shocks' impact on the observed variables. A convenient measure of the size of the total impact is each shock's contribution to the total variance of each variable. Figure 5 shows the individual contributions of the shocks as a percent of the total variances of the observables, as well as decompositions of the individual and total contributions in the low, BC, and high frequency bands. Note that the measurement errors also contribute to the variances, which is why the total contributions of the shocks sum up to less than 100%.

The results show that ξ_t drives most of the volatility in two of the observed variables – r_t and $\triangle i_t$, and, in the case of r_t, the contribution is mostly in the low frequencies. This

		$\triangle y$	r	$\triangle i$
all shocks	total	93.7	84.1	96.3
	low	8.7	64.8	8.8
	BC	37.2	17.3	40.7
	high	47.8	2.0	46.8
ξ	total	10.9	77.3	72.4
	low	0.2	62.7	7.3
	BC	4.3	13.8	34.8
	high	6.5	0.8	30.3
θ	total	0.1	0.4	2.4
	low	0.0	0.2	0.1
	BC	0.0	0.2	0.9
	high	0.0	0.0	1.4
z	total	0.1	0.4	2.3
	low	0.0	0.2	0.1
	BC	0.0	0.2	0.9
	high	0.0	0.0	1.4
g	total	76.6	0.0	0.1
	low	8.4	0.0	0.0
	BC	30.1	0.0	0.0
	high	38.2	0.0	0.0
z^m	total	0.7	2.0	11.9
	low	0.0	0.2	0.0
	BC	0.2	0.9	1.7
	high	0.5	0.8	10.1
z^{m2}	total	3.5	3.6	1.6
	low	0.1	1.3	0.0
	BC	1.7	1.9	0.8
	high	1.7	0.4	0.8
g^m	total	1.8	0.4	5.8
	low	0.1	0.2	1.4
	BC	0.9	0.2	1.5
	high	0.8	0.0	2.8
		$\triangle y$	r	$\triangle i$

Figure 5: Total and individual contributions of the shocks as a percent of the variances of the observables in the full spectrum and the low, business cycle, and high frequency bands. The difference to 100% is accounted for by the measurement error variances.

is consistent with the earlier findings that, of the three observed variables, r_t is the most informative and $\triangle y_t$ – the least informative one. Similarly, the second best recoverable shock – to permanent productivity, is responsible for the bulk of the volatility of the third variable – $\triangle y_t$, and particularly in the BC and high frequencies, which, as seen in Table 1, is also where most of the uncertainty of that shock stems from. The variance contributions of the remaining five shocks are significantly smaller, and account for only between 12%, in the case of transitory interest rate shock (z_t^m) with respect to $\triangle i_t$, and $2.3\% - 2.4\%$ in the case of both the labor supply (θ_t) and transitory productivity (z_t) shocks with respect again to $\triangle i_t$.

Equivalence between variance and information decompositions. Variance decompositions in dynamic structural models are typically obtained by shutting-off all shocks but one at a time and then computing the endogenous variables' variances or spectral densities (see for instance **?**, Section 8). This gives the contribution of each shock to the total variances or spectral densities of the endogenous variables. It is easy to see that the same quantities can be obtained using the information gain measures introduced in Section 2. Specifically, a shock's contribution to the variance of a variable is equal to the information gained, i.e. the reduction in variance, with respect to the variable due to knowing that shock. In other words, instead of information from observed variables to shocks, we measure the flow of information in the opposite direction – from shocks to observables. Of course, this only works when the shocks are mutually independent, which is also the assumption behind the standard variance decomposition approach. If shocks are mutually dependent one has to distinguish between conditional and unconditional variance contributions, as in the case of information from observed variables with respect to shocks.

To summarize, as expected, there is a clear link between, on the one hand, the shocks' contributions to the observed variables' volatilities and, on the other hand, the degree to which each shock can be recovered from information contained in those variables. At the same time, it is important to point out that the size of the contributions is not necessarily a good indicator of the variables' importance as sources of information about the shocks. For instance, the intertemporal preference shock contributes similar fractions of the variances of r_t and $\triangle i_t$. Yet, r_t is significantly more informative than $\triangle i_t$ about that shock. As noted earlier, this is due to the fact that the variance contributions are in different parts of the spectrum – the low frequencies in the case of r_t, and the BC and high frequencies, in the case of $\triangle i_t$. Since most of the variance of the preference shock comes from the low frequencies,

r_t is significantly more informative than $\triangle i_t$. In other cases, it is the information interactions among the observed variables that affect their relative importance as sources of information. For instance, as can be seen in Table 2, the conditional contribution of information by $\triangle i_t$ with respect to the transitory trend inflation shock (z_t^{m2}) is much larger than that of $\triangle y_t$, in spite of the significantly larger fraction of the variance of $\triangle y_t$ attributed to that shock, compared to $\triangle i_t$. This is explained by the strong positive complementarity between r_t and $\triangle i_t$ in the BC and especially the low frequencies, which is where most of the uncertainty of that shocks is located. Lastly, small variance contributions of a shock does not necessarily imply that the shock cannot be recovered. In general, having the same number of non-redundant observables as the number of sources of uncertainty means that all shocks are fully recoverable. This is the case in the model I consider next.

3.2 ?

? (henceforth JPT) investigate whether investment shocks are major drivers of business cycle fluctuations. Building on their work in **?**, they estimate a New Keynesian model featuring imperfectly competitive goods and labor markets, along with various nominal and real frictions, including sticky prices and wages, habit formation in consumption, variable capital utilization and investment adjustment costs. As in the previous section, I outline only those model features relevant to the subsequent information decomposition analysis..

The model has eight structural shocks, including three technology shocks, two of which relate to investment. JPT distinguish between final and intermediate consumption, investment, and capital goodsproduced in separate sectors. They incorporate two distinct shocks: one affecting the transformation of consumption into investment goods, and another affecting the transformation of investment goods into productive capital. The investment-specific technology (IST) shock enters through the production function in the investment goods sector:

$$I_t = \Upsilon_t Y_t^I, \tag{3.7}$$

where I_t represents the quantity of investment goods in efficiency units produced using Y_t^I units of the final good. Υ_t denotes the IST and follows a non-stationary random process growing at rate υ_t.

The second investment technology shock enters through the capital goods sector's production technology, where new capital i_t is produced from investment goods according to:

$$i_t = \mu_t \left(1 - S\left(\frac{I_t}{I_{t-1}}\right)\right) \tag{3.8}$$

represents an investment adjustment cost function and μ_t is a stationary shock to the marginal efficiency of investment (MEI), following an AR(1) process.

The third technology shocks affects the production functions in the intermediate good producing sector according to:

$$Y_t(i) = \max\{A_t^{1-\alpha} K_t(i)^\alpha L_t(i)^{1-\alpha} - A_t \Upsilon_t^{\frac{\alpha}{1-\alpha}} F; 0\} \tag{3.9}$$

where $Y_t(i)$, $K_t(i)$, and $L_t(i)$ denote the output produced and the effective capital and labor used by intermediate good producer i. F represents the fixed cost of production, and A_t is a common non-stationary neutral technology process growing at rate z_t.

The final consumption good Y_t is produced by combining a continuum of intermediate goods, according to

$$Y_t = \left[\int_0^1 Y_t(i)^{\frac{1}{1+\lambda_{p,t}}}\right]^{1+\lambda_{p,t}} \tag{3.10}$$

where $\lambda_{p,t}$ is a stationary price markup shock following an ARMA(1,1) process.

Similar to the model in the previous section, there is a shock to the intertemporal preferences of households, whose lifetime utility function is given by

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t b_t \left\{ \log\left(C_t - hC_{t-1}\right) - \varphi \frac{L_t(j)^{1+\nu}}{1+\nu} \right\}, \tag{3.11}$$

where C_t denotes consumption, b_t is the stationary intertemporal preference shock following an AR(1) process. JPT assume a continuum of households $j \in [0, 1]$, each supplying specialized labor $L_t(j)$. The specialized labor is combined into homogenous labor input according to

$$L_t = \left[\int_0^1 L_t(i)^{\frac{1}{1+\lambda_{w,t}}}\right]^{1+\lambda_{w,t}} \tag{3.12}$$

where $\lambda_{w,t}$ is a stationary wage markup shock assumed to follow an ARMA(1,1) process.

The final two shocks affect government fiscal and monetary policy. Public spending

G_t is a time-varying fraction of output,

$$G_t = \left(1 - \frac{1}{g_t}\right) Y_t \tag{3.13}$$

where the government spending shock g_t is a stationary AR(1) process.

Monetary policy sets the nominal interest rate R_t according to the following policy rule:

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \left(\frac{X_t}{X_t^*}\right)^{\phi_X}\right]^{1-\rho_R} \left[\frac{X_t/X_{t-1}}{X_t^*/X_{t-1}^*}\right]^{\phi_{dX}} \varepsilon_{mp,t}, \tag{3.14}$$

where $e_{mp,t}$ represents the monetary policy shock, R denotes the nominal rate's steady state, π_t is the inflation rate, $X_t = C_t + I_t + G_t$ represents actual real GDP, and X_t^* is GDP's level under flexible prices and wages and in the absence of markup shocks.

To summarize, the model has eight shocks: six stationary and two non-stationary. The price and wage markup shocks follow ARMA(1,1) processes, while the monetary policy shock follows an i.i.d process. The government spending, MEI, and intertemporal preference shocks, along with growth rates of IST and neutral technology shocks, follow AR(1) processes. All shock disturbances are Gaussian, resulting in a linear Gaussian state space representation of the solution of the log-linear approximation to the model.

JPT estimate the model using US data on hours worked ($h_t = \log L_t$), inflation (π_t), nominal interest rate (R_t), and the growth rates of GDP ($x_t = \triangle \log X_t$), consumption ($c_t = \triangle \log C_t$), investment ($i_t = \triangle \log I_t$), real wages ($w_t = \triangle \log \frac{W_t}{P_t}$), and relative price of investment ($\pi_t^i = \triangle \log \frac{P_{It}}{P_t}$). Unlike **?**, they do not allow for measurement errors in any of the series. This implies that all eight shocks can be fully recovered from the eight observed variables. In this section, I examine each shock's main sources of information in terms of observed variables and spectral components.

3.2.1 Information decomposition across frequency bands

Table 4 presents the total information gains for the eight shocks and their decompositions across low, BC, and high frequencies. All shocks can be fully recovered from information in the observables, both in the full spectrum and within each frequency band. The band-specific information contributions reflect the fraction of each shock's variance originating in those bands.

For six of the eight shocks uncertainty is distributed monotonically across the frequency bands, increasing or decreasing from low to high frequencies. The gov-

Table 4: Information decomposition across frequency bands

	shock	total	low	BC	high
z	neutral technology	100	$11.2 = 100 \times 0.11$	$40.0 = 100 \times 0.40$	$48.8 = 100 \times 0.49$
g	government	100	$96.1 = 100 \times 0.96$	$3.2 = 100 \times 0.03$	$0.7 = 100 \times 0.01$
υ	IST	100	$8.4 = 100 \times 0.08$	$33.6 = 100 \times 0.34$	$58.0 = 100 \times 0.58$
λ_p	price mark-up	100	$51.7 = 100 \times 0.52$	$16.1 = 100 \times 0.16$	$32.2 = 100 \times 0.32$
λ_w	wage mark-up	100	$5.1 = 100 \times 0.05$	$27.3 = 100 \times 0.27$	$67.6 = 100 \times 0.68$
b	preference	100	$22.8 = 100 \times 0.23$	$49.9 = 100 \times 0.50$	$27.4 = 100 \times 0.27$
ε_{mp}	monetary policy	100	$6.3 = 100 \times 0.06$	$27.1 = 100 \times 0.27$	$66.7 = 100 \times 0.67$
μ	MEI	100	$47.4 = 100 \times 0.47$	$40.8 = 100 \times 0.41$	$11.7 = 100 \times 0.12$

Note: see the note to Table 1.

ernment spending shock's uncertainty concentrates in low frequencies (96% of total variance), while the MEI shock's uncertainty primarily resides in low and BC frequencies. Neutral technology, IST, wage mark-up, and monetary policy shocks show uncertainty mainly in BC and high frequencies. For the intertemporal preference shock, half of the uncertainty resides in the business cycle frequencies, with the remainder distributed almost equally between low and high frequencies. The other shock with a non-monotonic distribution of uncertainty is the price mark-up shock. About half of its variance is attributable to low frequencies, with high frequencies contributing significantly, and the smallest share of uncertainty arising from the business cycle frequencies.

3.2.2 Information contributions by variables

Table 5 presents the conditional information gains for each observed variable across the full spectrum and individual frequency bands. Three contributions stand out, each exceeding 90%: the growth rate of the relative investment price (π^i) for the IST shock (υ), real wage growth (w) for the wage mark-up shock (λ_w), and the nominal interest rate (R) for the monetary policy shock (ε_{mp}). As JPT demonstrate, the price of investment in terms of consumption goods is the inverse of the IST process, enabling full recovery of the IST growth rate from π^i alone. The 97.2% conditional information gain in the full spectrum implies that, without π^i, the remaining variables reduce uncertainty about υ by only 2.8%.

Beyond the IST shock, π^i also contributes to information about the MEI shock, though to a much lesser extent compared to other variables, particularly the investment growth rate, which is the most informative variable for that shock. Output and consumption growth rates are the primary sources of information for the government spending shock, while hours worked strongly informs the neutral technology shock.

Table 5: Conditional contribution of information

	shocks	total								low							
		x	c	i	h	w	π	R	π^i	x	c	i	h	w	π	R	π^i
z	neutral technology	15.6	0.0	0.2	46.4	0.0	0.0	0.0	0.1	0.9	0.0	0.1	1.5	0.0	0.0	0.0	0.0
g	government	45.5	52.8	18.3	0.0	0.0	0.0	0.0	0.0	42.6	49.9	14.8	0.0	0.0	0.0	0.0	0.0
υ	IST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	97.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.1
λ_p	price mark-up	13.7	0.2	0.3	21.4	29.3	32.4	0.2	0.2	11.0	0.0	0.2	9.1	27.3	0.5	0.0	0.1
λ_w	wage mark-up	0.8	0.2	0.4	1.4	93.2	23.3	0.3	0.3	0.1	0.0	0.0	0.1	1.6	1.4	0.1	0.0
b	preference	1.4	28.5	7.3	11.2	2.5	0.7	5.6	0.0	1.0	4.2	6.1	6.6	2.3	0.6	3.4	0.0
ε_{mp}	monetary policy	0.3	3.1	0.2	10.2	0.1	12.1	92.6	0.0	0.1	0.1	0.0	1.5	0.0	4.4	4.8	0.0
μ	MEI	0.1	0.0	8.7	0.4	2.2	0.4	5.2	1.9	0.0	0.0	3.9	0.1	2.0	0.1	3.2	1.2
	shocks	BC								high							
		x	c	i	h	w	π	R	π^i	x	c	i	h	w	π	R	π^i
z	neutral technology	4.5	0.0	0.0	13.4	0.0	0.0	0.0	0.0	10.2	0.0	0.0	31.6	0.0	0.0	0.0	0.0
g	government	2.5	2.6	3.0	0.0	0.0	0.0	0.0	0.0	0.5	0.3	0.5	0.0	0.0	0.0	0.0	0.0
υ	IST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	56.1
λ_p	price mark-up	1.5	0.0	0.0	6.7	1.9	7.9	0.1	0.0	1.2	0.1	0.1	5.7	0.1	24.0	0.1	0.0
λ_w	wage mark-up	0.3	0.1	0.1	0.4	24.6	8.7	0.2	0.2	0.4	0.1	0.2	1.0	67.1	13.2	0.1	0.1
b	preference	0.1	10.0	1.1	2.7	0.2	0.2	1.8	0.0	0.2	14.3	0.1	2.0	0.0	0.0	0.4	0.0
ε_{mp}	monetary policy	0.1	0.8	0.1	4.8	0.0	4.1	24.8	0.0	0.1	2.2	0.2	3.9	0.0	3.6	63.0	0.0
μ	MEI	0.0	0.0	1.7	0.2	0.1	0.2	1.6	0.3	0.0	0.0	3.2	0.2	0.0	0.1	0.4	0.3

Note: see the note to Table 2. The observed variables are: the growth rates of output (y), consumption (c), investment, and wages (w), the inflation rates for consumption (π) and investment (π^i), hours worked (h) and the nominal interest rate (r). Due to rounding in some cases the band-specific contributions do not add up to the total values.

Consumption growth is the most informative variable for the intertemporal preference shock, while inflation is the most informative observable for the price mark-up shock.

Variables that provide the most information overall are typically the most informative within each frequency band, with some exceptions. One exception is wage growth's contribution to the price mark-up shock: it surpasses inflation's contribution in the low frequency band but falls below it in the business cycle and high frequencies, and thus overall. Another notable exception is the intertemporal preference shock: while consumption growth is the most informative variable overall, hours worked and investment growth contribute significantly more in the low frequency band

Table 6 presents the unconditional information gains. As previously discussed, the difference between conditional and unconditional information gains for a given variable and shock indicates information complementarities between that variable and other observables. These complementarities may be positive or negative, depending on whether conditional gains exceed or fall below unconditional ones.

The strongest positive complementarity appears in the government spending shock, where the highest unconditional gain (from R) is below 5%, while conditional

Table 6: Unconditional contribution of information

	shocks	total								low							
		x	c	i	h	w	π	R	π^i	x	c	i	h	w	π	R	π^i
z	neutral technology	17.4	20.1	7.3	24.3	28.7	24.4	9.6	0.0	5.4	6.1	2.7	0.6	7.5	2.0	0.6	0.0
g	government	0.4	3.1	0.1	4.4	0.0	1.7	4.8	0.0	0.1	3.1	0.1	4.2	0.0	1.7	4.7	0.0
υ	IST	1.3	0.1	2.0	0.6	0.1	0.1	0.2	100.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	8.4
λ_p	price mark-up	1.8	0.3	4.4	4.8	18.1	39.3	3.6	0.0	1.3	0.3	3.8	4.3	8.3	6.2	1.5	0.0
λ_w	wage mark-up	0.4	0.4	0.6	1.0	59.7	3.4	0.6	0.0	0.3	0.2	0.2	0.8	0.2	2.1	0.5	0.0
b	preference	7.4	61.9	0.9	6.9	0.0	1.7	9.7	0.0	0.3	3.6	0.2	0.6	0.0	0.6	1.7	0.0
ε_{mp}	monetary policy	3.5	1.2	2.8	3.1	0.0	1.7	57.6	0.0	0.2	0.1	0.1	0.3	0.0	0.4	0.2	0.0
μ	MEI	47.2	10.3	73.1	56.3	3.4	9.7	51.6	0.0	16.5	8.4	28.2	24.9	2.8	4.9	30.4	0.0
	shocks	BC								high							
		x	c	i	h	w	π	R	π^i	x	c	i	h	w	π	R	π^i
z	neutral technology	7.5	8.4	2.7	7.7	16.8	14.3	6.0	0.0	4.5	5.6	1.9	15.9	4.4	8.1	3.0	0.0
g	government	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0
υ	IST	0.2	0.0	0.3	0.1	0.0	0.0	0.1	33.6	1.0	0.0	1.7	0.5	0.0	0.0	0.0	58.0
λ_p	price mark-up	0.3	0.0	0.3	0.3	3.6	7.2	0.6	0.0	0.3	0.0	0.3	0.2	6.2	25.8	1.5	0.0
λ_w	wage mark-up	0.1	0.1	0.2	0.1	10.1	1.2	0.2	0.0	0.0	0.1	0.2	0.0	49.4	0.1	0.0	0.0
b	preference	4.0	34.9	0.5	3.9	0.0	0.9	5.6	0.0	3.2	23.5	0.2	2.3	0.0	0.2	2.4	0.0
ε_{mp}	monetary policy	1.1	0.4	0.8	1.2	0.0	0.7	8.1	0.0	2.2	0.7	1.9	1.6	0.0	0.6	49.3	0.0
μ	MEI	24.6	1.9	34.6	26.6	0.6	4.4	20.2	0.0	6.1	0.1	10.4	4.7	0.0	0.4	1.0	0.0

Note: see the note to Table 3.

gains exceed 45% for c and x, and 18% for i. This results from the tight relationship among x, c, i, and g in the economy's resource constraint. Since g is latent, combined information from pairs of observed resource constraint variables exceeds their individual contributions. This intuition is confirmed by the information complementarity measure (equation (3.5)). The top panel of Figure 6 displays the largest absolute unconditional and conditional information complementarities for the government spending shock. The strongest positive complementarities occur between resource constraint variable pairs. For example, the 3.2 value for x and c indicates that, given the other six variables, their joint observation provides 2.2 times more information about g than their individual contributions combined.

The bottom panel of Figure 6 shows the most significant complementarities for the MEI shock. Comparing Tables 5 and 6 highlights the MEI shock as the strongest case of negative information complementarities. The gain from i, the most informative variable both conditionally and unconditionally, drops from over 70% unconditionally to under 10% conditionally. Similarly, the gains from R, h, and x decrease from about 50% to 5% or less. This suggests that much of the information these variables provide about the MEI shock is shared with other observed variables. As Figure 6 illustrates, combinations of i, R, h, and x show the strongest negative information

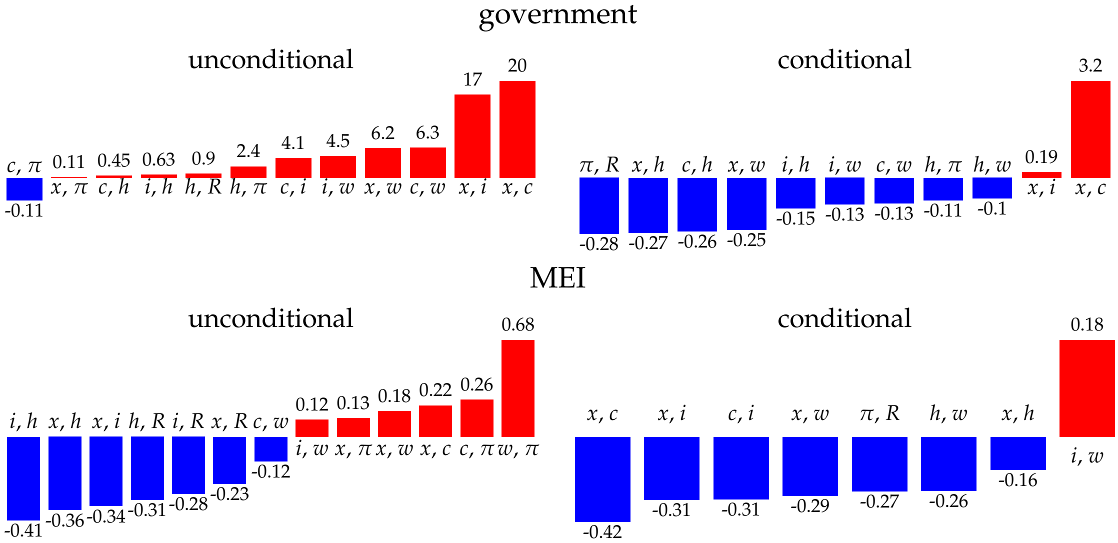

Figure 6: Largest pairwise information complementarities with respect to the government spending and MEI shocks, full spectrum.

complementarity. For i and x, this stems from their strong interdependence with c through the economy's resource constraint. Indeed, negative complementarity should exist between any two resource constraint variables (when the third is among the conditional variables) for all shocks except g.[10] Details appear in the Appendix, which also presents pairwise complementarity coefficients for each frequency band. This frequency analysis explains why overall information complementarity between c and i for the g shock is zero, despite the expected strong mutual complementarity among resource constraint variables. While complementarities among c, i, and x are strong in BC and high frequencies, nearly all information about g resides in low frequencies, where complementarity between c and i vanishes.

JPT conclude that the MEI shock is the key source of business cycle fluctuations whereas the IST shock plays no role. In particular, they demonstrate that the MEI shock accounts for substantial variance in GDP, investment, and hours *at business cycle frequencies*, while IST shock contributions are negligible. Figure 7 shows variance decompositions across frequency bands and overall shock contributions to observed variables' variances.[11]

[10] At the risk of belaboring the obvious, consider the case where g is also observed or where the g shock has zero variance. Due to exact collinearity, information from any resource constraint variable is completely redundant given the remaining resource constraint variables.

[11] JPT's results appear in their Table 3. Several differences exist between their presentation and Figure 7. JPT show variance contributions as fractions of total business cycle frequency variance, while my plot shows fractions relative to the total variances in the full spectrum. For comparable contributions, plot values must be multiplied by the fraction of the total variance of each shock due to the BC frequencies. Additionally, JPT show level decompositions for trending variables (x, c, i,

The results show that the MEI shock accounts for most of the variance in x, i, and h across the full spectrum, not just in the BC frequencies. It also contributes significantly to the volatility of R. This helps explain why R is the second most informative variable about μ (Table 5). The MEI shock contributes 65% and 55% of R's total variance in the low and BC frequencies, respectively, where most uncertainty about μ resides. Unlike the resource constraint variables, a smaller fraction of the information in R is redundant. The small impact of the preference shock on c in the low frequencies explains why c, despite being the most informative variable overall, is less informative in these frequencies. Similarly, the fact that the wage mark-up shock primarily contributes to w's volatility in the high frequencies aligns with w's dominant role in providing information about that shock. Conversely, low frequencies account for most of the uncertainty about the price mark-up shock, contributing more than half of its total variance. Given that this shock's contribution to h's variance is also concentrated in the low frequencies, this explains why h is as important as w and π in the information it contributes about λ_p, despite the much smaller fraction of h's total variance due to that shock.

As seen earlier, the significance of h is even more pronounced in the case of the neutral technology shock, where it provides the largest conditional contribution of information among observables. This insight might not be readily apparent from the variance decomposition results, which indicate that while more than half of z's contribution is to the low frequency component of the variance, only 0.3% of h's total variance comes from the high frequency contribution of that shock. However, the BC and high frequencies account for nearly 90% of the total information about z, with h contributing most of its information within the high frequency band. As detailed in the Appendix, this finding stems from two key relationships: the strong positive complementarity between h and x, and the strong negative complementarities among x, c, and i, as well as between π and w, and π and R. This indicates substantial redundancy in the information about z across variables for which this shock is an important source of volatility. Moreover, while only 1% of h's total variance originates in the high frequency band, z accounts for 30% of this portion, making it the second most important shock after μ for h in the high frequencies.

This observation further reinforces our earlier discussion regarding the **?** model, namely that the magnitude of variance contribution does not necessarily reflect a variable's significance as a source of shock information. Indeed, shocks can be fully

and w), whereas I present growth rates. Finally, the point estimates in JPT are the median values of the posterior distributions of the contributions. I present decompositions at the posterior median of the estimated model parameters.

recoverable even when they contribute modestly to overall volatility. The JPT model illustrates this phenomenon: despite the monetary policy shock accounting for no more than 9.5% of any observable's volatility, and the government spending shock contributing at most 7.3%, both shocks remain fully recoverable.

Shock	Band	x	c	i	h	w	π	R	π^i
all shocks	low	22.4	30.4	19.6	79.8	21.5	46.7	63.4	8.4
all shocks	BC	52.3	45.9	58.9	19.2	33.7	35.4	33.5	33.6
all shocks	high	25.3	23.7	21.5	1.0	44.7	17.9	3.1	58.0
z	total	23.2	30.4	8.4	6.6	33.1	22.2	8.1	0.0
z	low	9.9	17.0	3.0	3.8	14.6	6.3	3.1	0.0
z	BC	11.0	10.7	4.5	2.5	14.9	12.9	4.7	0.0
z	high	2.4	2.7	0.9	0.3	3.6	3.0	0.2	0.0
g	total	7.3	2.2	0.1	2.1	0.0	0.5	1.1	0.0
g	low	0.1	1.2	0.0	1.5	0.0	0.4	0.9	0.0
g	BC	1.8	0.8	0.0	0.4	0.0	0.1	0.3	0.0
g	high	5.4	0.2	0.0	0.1	0.0	0.0	0.0	0.0
v	total	0.7	0.3	1.1	0.4	0.1	0.4	0.9	100.0
v	low	0.1	0.2	0.1	0.3	0.1	0.3	0.8	8.4
v	BC	0.2	0.1	0.4	0.0	0.0	0.0	0.1	33.6
v	high	0.4	0.0	0.6	0.0	0.0	0.0	0.0	58.0
λ_p	total	2.0	0.2	2.2	6.8	21.1	34.6	2.1	0.0
λ_p	low	0.8	0.1	0.9	6.2	4.9	6.4	0.9	0.0
λ_p	BC	1.0	0.0	1.1	0.6	7.5	14.2	1.0	0.0
λ_p	high	0.2	0.0	0.2	0.0	8.7	14.1	0.1	0.0
λ_w	total	1.5	1.8	1.0	26.2	44.0	28.2	10.4	0.0
λ_w	low	1.2	1.6	0.5	26.0	0.6	25.6	9.9	0.0
λ_w	BC	0.3	0.2	0.4	0.2	11.0	2.6	0.5	0.0
λ_w	high	0.0	0.0	0.1	0.0	32.4	0.0	0.0	0.0
b	total	7.3	56.5	1.0	3.3	0.0	2.0	8.4	0.0
b	low	0.3	4.4	0.2	2.0	0.0	1.3	4.8	0.0
b	BC	4.0	31.8	0.6	1.3	0.0	0.7	3.3	0.0
b	high	2.9	20.3	0.2	0.1	0.0	0.1	0.3	0.0
ε_{mp}	total	3.7	1.3	2.8	4.9	0.0	3.9	9.5	0.0
ε_{mp}	low	0.7	0.4	0.5	3.9	0.0	2.5	2.1	0.0
ε_{mp}	BC	2.2	0.6	1.7	1.0	0.0	1.2	5.3	0.0
ε_{mp}	high	0.8	0.3	0.6	0.0	0.0	0.2	2.1	0.0
μ	total	54.1	7.4	83.4	49.6	1.7	8.2	59.7	0.0
μ	low	9.3	5.6	14.3	36.2	1.3	3.9	40.9	0.0
μ	BC	31.7	1.6	50.1	13.0	0.3	3.7	18.4	0.0
μ	high	13.1	0.1	19.0	0.4	0.0	0.5	0.3	0.0

Figure 7: Total and individual contributions of shocks to the variances of observables, expressed as percentages, across the full spectrum and within the low, business cycle, and high-frequency bands

3.3 ?

The primary contribution of **?** (henceforth ACD) lies in demonstrating how to incorporate higher-order belief dynamics into a broad class of macroeconomic models in a tractable manner. This approach is illustrated through several applications, including a medium-scale New Keynesian model, which I analyze in this section. Apart from the incorporation of higher-order beliefs, this model shares many features with the JPT model from the previous section and other medium-scale DSGE models in the literature: habit persistence in consumption, investment adjustment costs, variable capital utilization, price stickiness under Calvo pricing, monetary policy governed by a Taylor rule, and a variety of shocks, including permanent and transitory TFP shocks, permanent and transitory investment-specific shocks, a discount-rate shock, a news shock about future productivity, a government-spending shock, and a monetary policy shock.

From a modeling perspective, what distinguishes ACD from most of the literature is their departure from the assumptions of rational expectations and common information about the state of the economy. Specifically, their framework introduces autonomous variations in agents' beliefs about other agents' expectations (higher-order beliefs) through what they call a "confidence shock". This shock creates a divergence between different forms of beliefs, generating short-term fluctuations in agents' expectations of economic outcomes while leaving their medium- and long-term expectations unaffected. Notably, these fluctuations do not alter expectations about exogenous fundamentals at any horizon. ACD argue that incorporating this mechanism into standard DSGE models enhances their ability to replicate observed patterns in macroeconomic data.

In the remainder of this section, I evaluate the sources and distribution of information about the shocks in the ACD model. The model has nine shocks, most of which – except for the confidence shock – are defined similarly to those in the JPT model. A key difference, however, is that the level of TFP in the ACD model consists of both permanent and transitory components, whereas the JPT model includes only a permanent component. Additionally, ACD introduce a one-quarter-ahead news component to the permanent TFP term, modeled as an exogenous stationary AR(1) process.

The model equations closely parallel those presented in Section 3.2, and a detailed exposition is deferred to the Appendix. While the incorporation of higher-order beliefs alters the model's solution approach compared to standard rational expectations models, this modification primarily manifests in the presence of two types of expectations within standard equilibrium conditions. Specifically, some decisions are based on agents' beliefs that others' expectations are biased, while other decisions are made after the true state of nature and the realized values of economic activity become publicly known. The perceived bias in expectations is captured by the commonly observed confidence shock. For further details, see Section C in the Appendix and the original work of **?**.

ACD estimate the model using six quarterly US data series: GDP, consumption, investment, hours worked, the inflation rate, and the federal fund rate. The estimation is conducted in the frequency domain, focusing exclusively on business cycle frequencies. This approach represents another departure from the common practice in the literature, where time-domain methods are predominantly used. The authors justify their choice by emphasizing that their model is specifically designed to capture business cycle phenomena and, therefore, lacks the features and mechanisms necessary to explain the lower and higher frequency components of the data.

In the next section, I will examine the implications of this estimation approach and explore possible interpretations of the information decomposition in this context.

3.3.1 Information decomposition across frequency bands and observables

Table 7: Information decomposition across frequency bands

	shock	total	low	BC	high
a^p	permanent TFP	99.7	98.6 = 99.9 × 0.99	1.0 = 90.2 × 0.01	0.1 = 62.5 × 0.00
a^n	news	43.2	5.8 = 59.1 × 0.10	18.3 = 49.3 × 0.37	19.1 = 36.0 × 0.53
a^τ	transitory TFP	30.2	0.7 = 4.6 × 0.15	10.7 = 23.5 × 0.45	18.8 = 47.3 × 0.40
ζ^{IP}	permanent investment	91.5	91.5 = 92.7 × 0.99	0.1 = 6.3 × 0.01	0.0 = 6.6 × 0.00
ζ_t^{IT}	transitory investment	92.5	10.3 = 76.2 × 0.14	42.1 = 96.0 × 0.44	40.1 = 94.0 × 0.43
ζ^c	discount factor	98.6	64.9 = 99.4 × 0.65	27.5 = 98.0 × 0.28	6.2 = 93.9 × 0.07
ζ^g	fiscal	93.2	39.7 = 90.6 × 0.44	41.3 = 95.8 × 0.43	12.3 = 93.5 × 0.13
ζ^m	monetary policy	97.4	25.5 = 92.9 × 0.27	49.2 = 98.7 × 0.50	22.7 = 99.7 × 0.23
ξ	confidence	96.7	51.8 = 98.8 × 0.52	36.1 = 96.2 × 0.38	8.8 = 87.4 × 0.10

Note: see the note to Table 1.

Consider Table 7 which presents the results of applying the information decomposition to the shocks in the ACD model. These results can be interpreted in multiple ways, depending on one's perspective regarding the model's ability to represent the empirical data. First, assuming the model is correctly specified across all frequencies of the observed time series, the table illustrates how information about the shocks is distributed across the low, business cycle, and high-frequency bands, as well as the total information obtained about each shock. This interpretation mirrors the earlier analysis of Tables 1 and 4 for the Uribe and JPT models, respectively. Such a perspective enables conclusions about the sources of identification for the shocks in the ACD model, similar to the approach taken for those models.

Second, one may adopt the perspective, as ACD do, that the model is misspecified outside the BC frequencies and focus solely on those frequencies, disregarding the rest of the spectrum. From this viewpoint, the key question is how much of the uncertainty associated with the shocks originating in the BC frequencies is resolved by the information provided by the observed variables within that frequency band. The answer lies in the information gains at BC frequencies, shown in the middle of the fourth column of Table 7. Although no shock's prior uncertainty is completely resolved, the information gains exceed 90% for six shocks, reaching 98% for the discount factor and monetary policy shocks. In contrast, the least amount of information, only 6%, is obtained for the permanent investment-specific technology shock, followed by the transitory TFP and news shocks, with information gains of approximately 24% and 49%, respectively.

It is worth noting that for the latter two shocks, the information gains at BC frequencies are comparable to those over the full spectrum – 32% and 43%, respectively. However, for the permanent investment-specific technology shock, the total information gain is much higher, at 92%. This discrepancy reflects the fact that 99% of the uncertainty associated with this shock originates in the low-frequency band, where the information gain is nearly 93%.

Table 8: Conditional information gains

shock		all						BC					
		Y	C	I	N	π	R	Y	C	I	N	π	R
a^p	permanent TFP	0.9	0.5	1.1	1.2	0.0	0.1	26.6	0.1	1.0	37.3	0.0	0.3
a^n	news	10.7	2.2	0.7	11.4	0.1	2.7	15.0	0.5	0.9	19.4	0.0	1.4
a^τ	transitory TFP	2.0	0.3	2.4	11.5	0.3	1.7	1.0	0.4	2.6	8.6	0.2	2.3
ζ^{IP}	permanent investment	2.1	10.1	15.3	0.7	0.3	3.1	0.8	1.0	2.2	0.8	0.1	1.0
ζ_t^{IT}	transitory investment	0.2	10.1	32.7	1.2	0.8	1.2	0.3	12.4	24.8	0.1	0.5	1.6
ζ^c	discount factor	1.6	2.1	0.2	11.3	13.0	8.7	2.3	2.9	0.2	20.0	23.0	17.8
ζ^g	fiscal	65.5	53.9	70.2	10.1	0.1	0.7	62.0	55.9	71.9	9.8	0.0	0.1
ζ^m	monetary policy	3.2	3.4	0.4	19.1	71.6	67.4	1.6	1.8	0.1	13.0	70.1	66.7
ξ	confidence	5.4	19.1	0.1	10.4	3.2	3.4	4.7	23.2	0.1	9.2	3.3	4.6

Note: The conditional information gain measures the marginal reduction of uncertainty about a shock due to observing a variable relative to observing the other five variables, as a percent of the unconditional uncertainty of the shock. The observed variables are: output (Y), consumption (C), investment (I), hours worked (N), inflation (π) and the nominal interest rate (R).

The relative importance of observed variables for each shock can be determined by examining their respective information gains across both BC frequencies and the full spectrum, as shown in Table 8. Unlike Tables 2 and 5, the BC panel in this table reports how much of the uncertainty originating specifically in the BC frequencies is resolved by the information contained in the observed variables, rather than showing BC frequencies' contributions to total information. It is the appropriate metric when the model is deemed suitable for explaining only the BC frequencies of the data.

Comparing the two sets of information gains, the rankings of the most informative variables for each shock are nearly identical in the full spectrum and the BC frequencies. A notable exception is the permanent TFP shock (a^p), where output (Y) is significantly more informative than investment (I) at BC frequencies, whereas I is more informative in the full spectrum. For this shock, as well as for the transitory TFP (a^τ) and news (a^n) shocks, hours worked (N) provides the largest contribution of information. N also contributes substantially to the discount factor (ζ^c) and confidence (ξ) shocks, providing the second largest contributions for these shocks after inflation (π) and consumption (C), respectively. For the government spending shock (ζ^g), Y, C, and I are the three most significant sources of information. This pattern, also observed in the JPT model, reflects the resource constraint equation linking these variables to ζ^g.

Inflation (π), followed by the nominal interest rate (R), are the most informative variables for the monetary policy shock (ζ^m). This result is consistent with the JPT model (see Table 5), although in that model, R provides the largest contribution by a considerable margin. Lastly, as expected, I is the most informative variable for the two investment-specific shocks (ζ^{IP} and ζ_t^{IT}).

As previously discussed, the rationale for excluding frequencies outside the BC range is to avoid biased parameter estimates due to contaminated information from parts of the spectrum where the model is misspecified.[12] The same reasoning applies to the estimation of shocks and other latent variables: using contaminated information leads to distorted estimates of these variables, even when the true values of model parameters are known.

The extent and nature of these distortions depend on what mechanisms operate at the low and high frequencies of the data but are absent from the theoretical model. These factors are model- and data-specific. At the same time, it is clear that the greater the reliance on contaminated information during estimation, the more severe the impact of misspecification will be on the results. For instance, Table 7 shows that model misspecification at the lower end of the spectrum would have a greater impact on the estimates of shocks with larger contributions from the lower frequencies, such as the permanent investment-specific shock, compared to shocks

[12] **?** and **?** make similar arguments and develop band-spectral estimation methods. See also **?**, **?**, and **?**.

with smaller contributions, like the transitory investment-specific shock. Conversely, misspecification at high frequencies would have a greater impact on the transitory investment-specific shock than on the permanent one.

4 Concluding Comments

I have shown how to decompose the frequency domain information observables provide with respect to latent variables in dynamic macroeconomic models. Through this analysis, researchers can determine where in the spectrum information about latent variables predominantly comes from, and evaluate the relative contributions of individual observed variables. The examples I have presented illustrate how reporting the results from such analysis can make the estimation of shocks and other latent variables more transparent. Researchers often disagree on the specific model features needed to adequately represent the data. In particular, there is no consensus on which data frequencies macroeconomic models should aim to represents, or are capable of representing. Whilst much of the empirical literature is focused on explaining business cycle phenomena, models are usually estimated in the time domain, which is tantamount to using the full spectrum. The Uribe and JPT models I have considered are only two cases in point. Even if not explicitly stated, the time domain approach implicitly assumes that models are capable of representing all frequencies in the data. Presenting readers, who may have divergent views on model adequacy, with information on the relative importance of different frequencies will allow them to assess the potential consequences of using contaminated information due to model misspecification.

Another issue over which researchers may disagree concerns the extent to which observed time series adequately represent theoretical variables in macroeconomic models. The existence of multiple empirical counterparts to variables like output, inflation, wages, hours worked, etc. suggests that each should be treated as a noisy indicator of the underlying theoretical concept. This supports the argument in favor of treating these variables as measured with errors.[13] Yet, measurement errors are not universally present in estimated models, as demonstrated by the JPT and ACD examples. Such an omission could be interpreted by some readers as a reason to suspect that models are misspecified with respect to particular observed variables. Based on the perceived nature of the errors, one can draw a conclusion about which

[13] An earlier statement of this argument was made by **?** who proposed incorporating structural macroeconomic models into a dynamic factor framework where multiple imperfectly measured indicators correspond to each model concept.

frequencies are most affected. Knowing how important those frequencies are as a source of information can help readers better understand the consequences of the failure to account for the imperfect match between theoretical concepts and empirical time series. For instance, pure measurement errors are often modeled as white noise processes, and therefore the contamination is concentrated in the higher end of the spectrum. As a result, estimates that rely more heavily on information from the high frequencies will be compromised more severely. Similarly, perceived failure to adequately account for low frequency variations in some series would cause some readers to be sceptical of estimates that are more dependent on information from the lower end of the spectrum.[14]

Lastly, the methodology described in this paper can help researchers who develop and estimate structural macroeconomic models by revealing, in cases of information deficiency, what type of information is needed to better recover unobserved variables of interest. Having well-identified structural shocks and unobserved endogenous variables, such as potential output or natural rate of interest, is a key requirement for macroeconomic models to be useful as tools for policy analysis and to be credible as story-telling devices.

[14]One commonly cited example of this is the series for aggregate hours worked, which contains significant low-frequency variations attributed to demographics and other structural developments in the labor market that are absent from most business cycle models. See the discussion of Figure 5 in **?**.

Appendix

A Uribe (2021) model

Table A1: Parameter values, Uribe (2021) model

	parameter	posterior mean
ϕ	price stickiness	146.000
α_π	coeff inflation in monetary policy rule	2.320
α_y	coeff output in monetary policy rule	0.188
γ_m	backward-looking component in inflation	0.606
γ_I	coeff lagged interest rate in monetary policy rule	0.242
δ	habit formation	0.258
ρ_ξ	AR preference	0.915
ρ_θ	AR labor supply	0.708
ρ_z	AR transitory productivity	0.700
ρ_g	AR permanent productivity	0.221
ρ_{gm}	AR permanent trend inflation	0.248
ρ_{zm}	AR transitory interest rate	0.306
ρ_{zm2}	AR transitory trend inflation	0.796
σ_ξ	std. preference	0.0287
σ_θ	std. labor supply	0.00164
σ_z	std. transitory productivity	0.00122
σ_g	std. permanent productivity	0.00758
σ_{gm}	std. permanent trend inflation	0.000848
σ_{zm}	std. transitory interest rate	0.000832
σ_{zm2}	std. transitory trend inflation	0.00131
σ_1^{me}	std. measurement error$\triangle y_t$	4.46e-06
σ_2^{me}	std. measurement errorr_t	4.55e-06
σ_3^{me}	std. measurement error $\triangle i_t$	1.74e-07

B ?

Table B1: Parameter values, JPT (2011) model

	parameter	posterior median
α	capital share	0.169
ι_p	price indexation	0.113
ι_w	wage indexation	0.102
h	consumption habit	0.864
λ_p	SS mark-up goods prices	0.177
λ_w	SS mark-up wages	0.166
ν	inverse frisch elasticity	5.162
ξ_p	Calvo prices	0.783
ξ_w	Calvo wges	0.773
χ	Elasticity capital utilization cost	5.491
S'	Investment adjustment costs	3.017
ϕ_π	Taylor rule inflation	1.735
ϕ_Y	Taylor rule output	0.059
ρ_R	Taylor rule smoothing	0.863
ρ_z	AR neutral technology growth	0.286
ρ_g	AR government spending	0.990
ρ_ν	AR IST growth	0.148
ρ_p	AR price mark-up	0.978
ρ_w	AR wage mark-up	0.968
ρ_b	intertemporal preference	0.583
θ_p	MA price mark-up	0.793
θ_w	MA wage mark-up	0.990
ϕ_{dy}	Taylor rule output growth	0.199
ρ_μ	AR MEI	0.807
σ_{mp}	std. monetary policy	0.216
σ_z	std. neutral technology growth	0.943
σ_g	std. government spending	0.362
σ_ν	std. IST growth	0.634
σ_p	std. price mark-up	0.222
σ_w	std. wage mark-up	0.310
σ_b	std. intertemporal preference	0.038
σ_μ	std. MEI	5.691

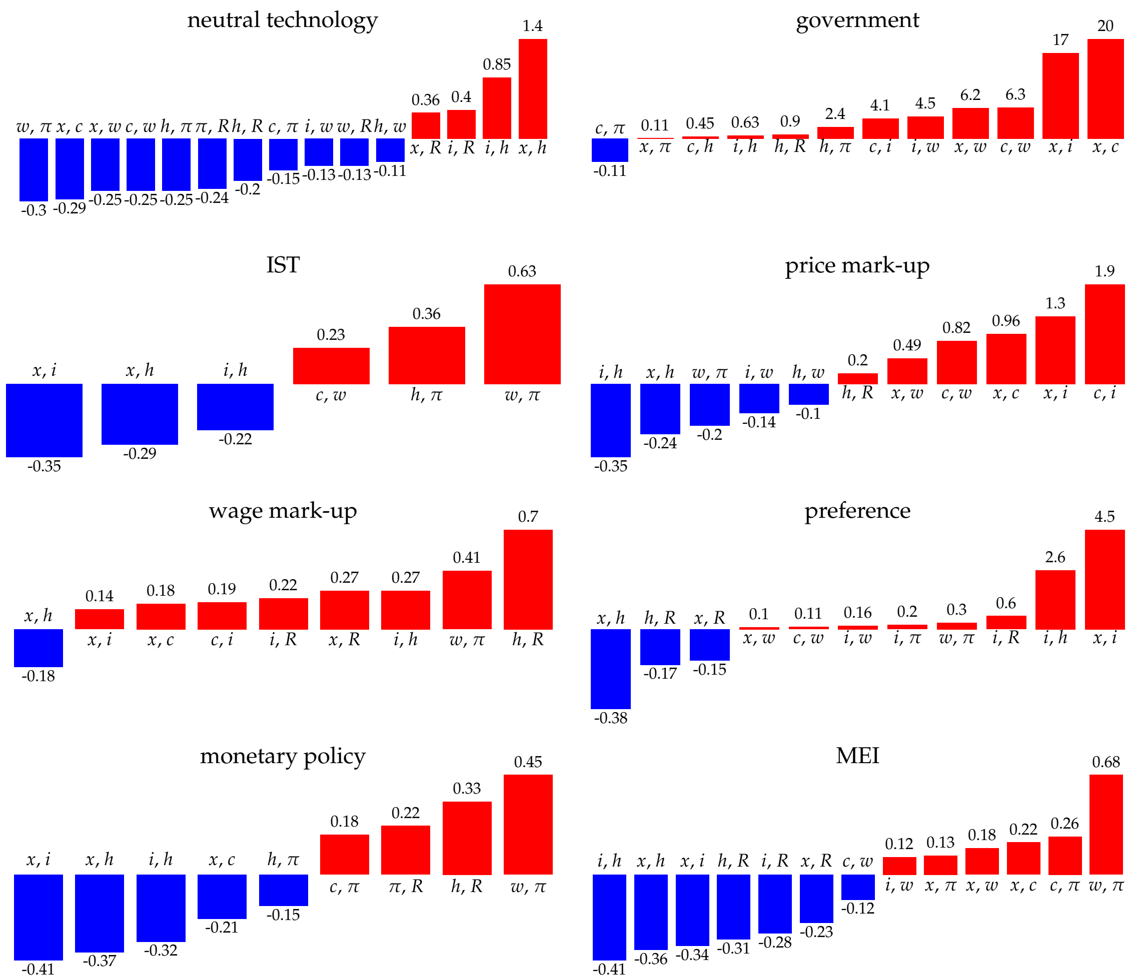

Figure B1: Largest unconditional pairwise information complementarities, all frequencies.

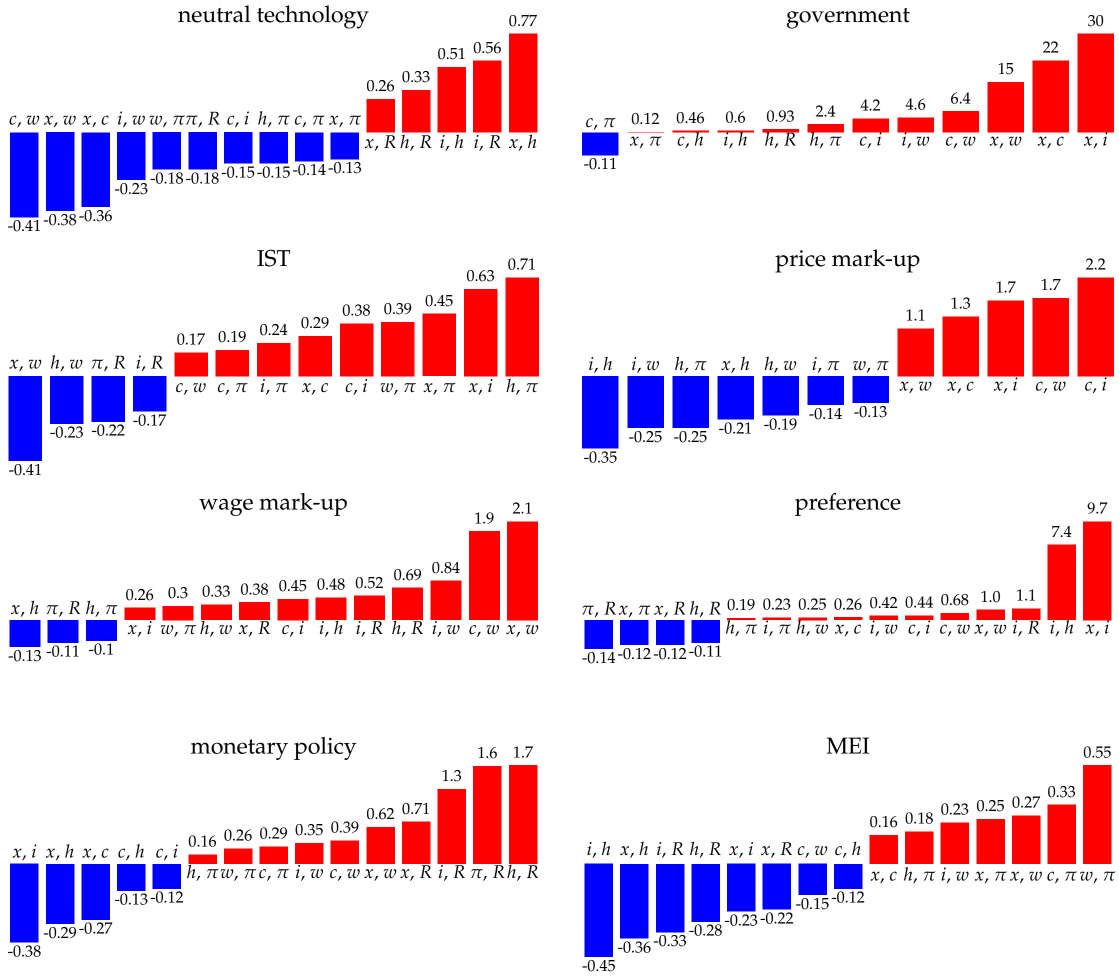

Figure B2: Largest unconditional pairwise information complementarities, low frequencies.

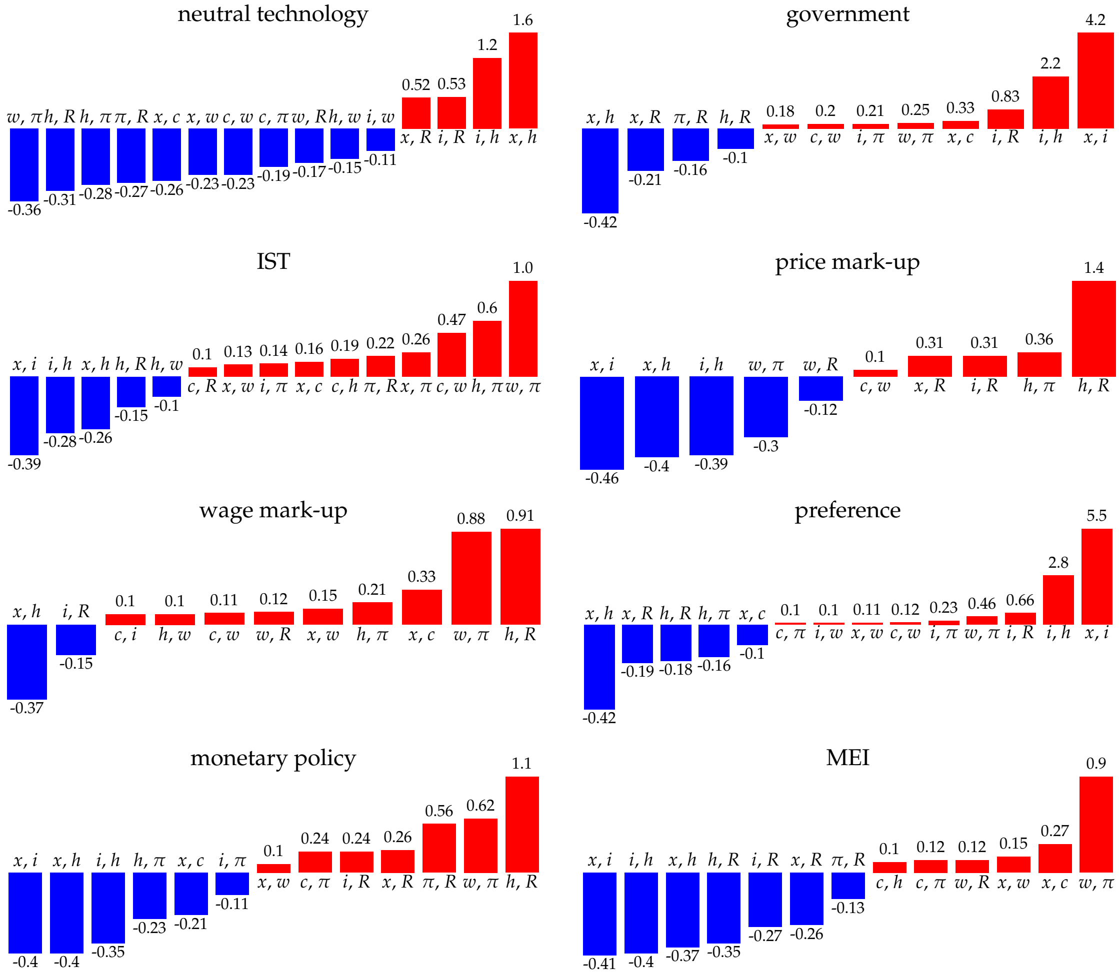

Figure B3: Largest unconditional pairwise information complementarities, BC frequencies.

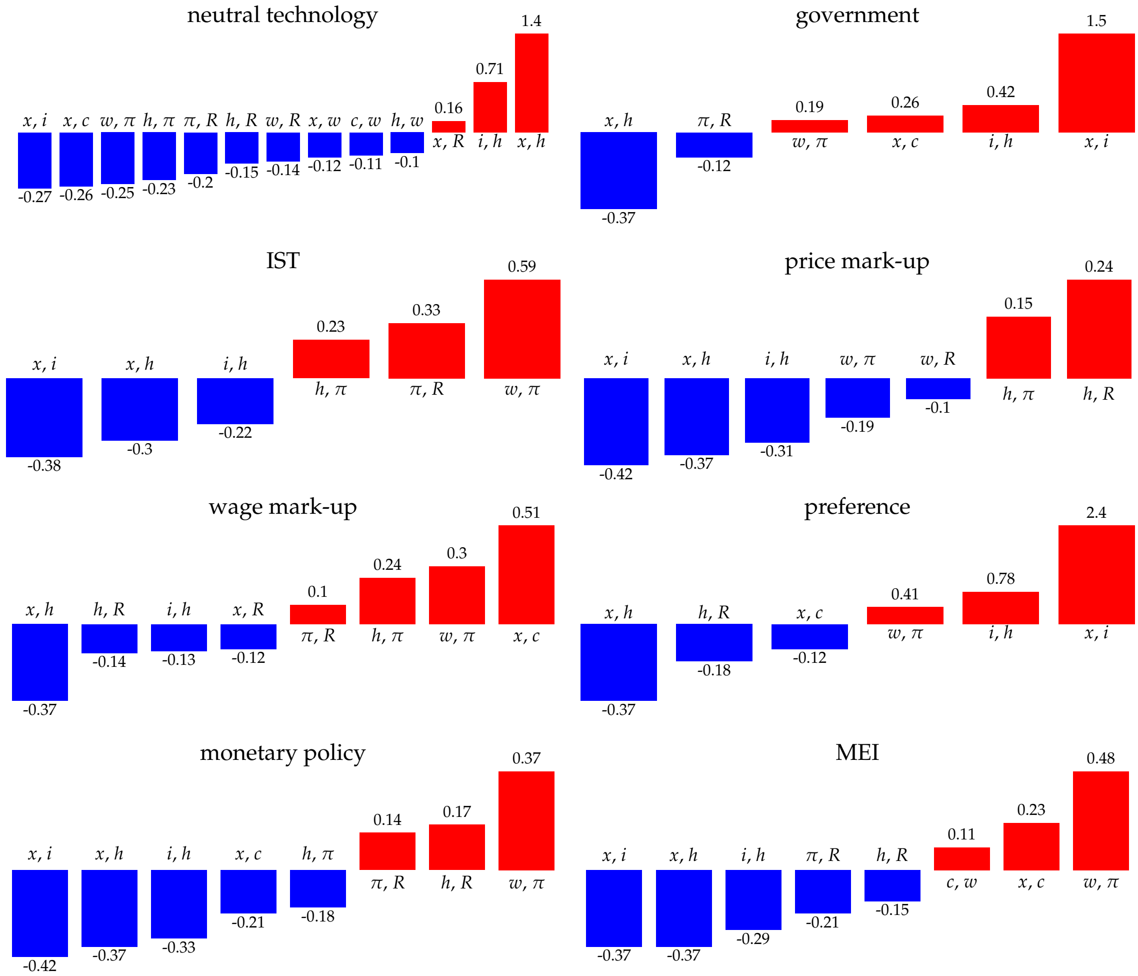

Figure B4: Largest unconditional pairwise information complementarities, high frequencies.

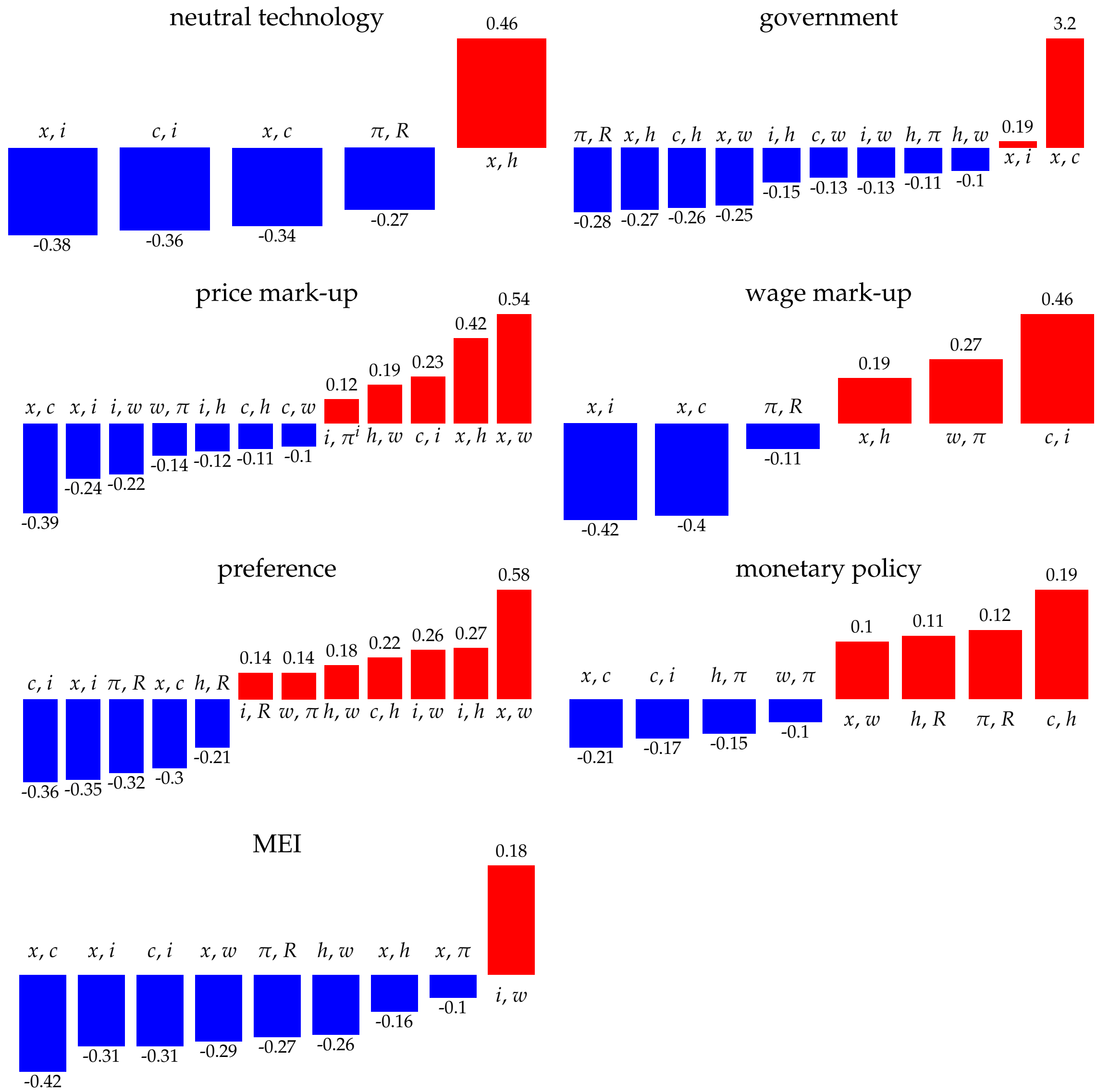

Figure B5: Largest conditional pairwise information complementarities, full spectrum.

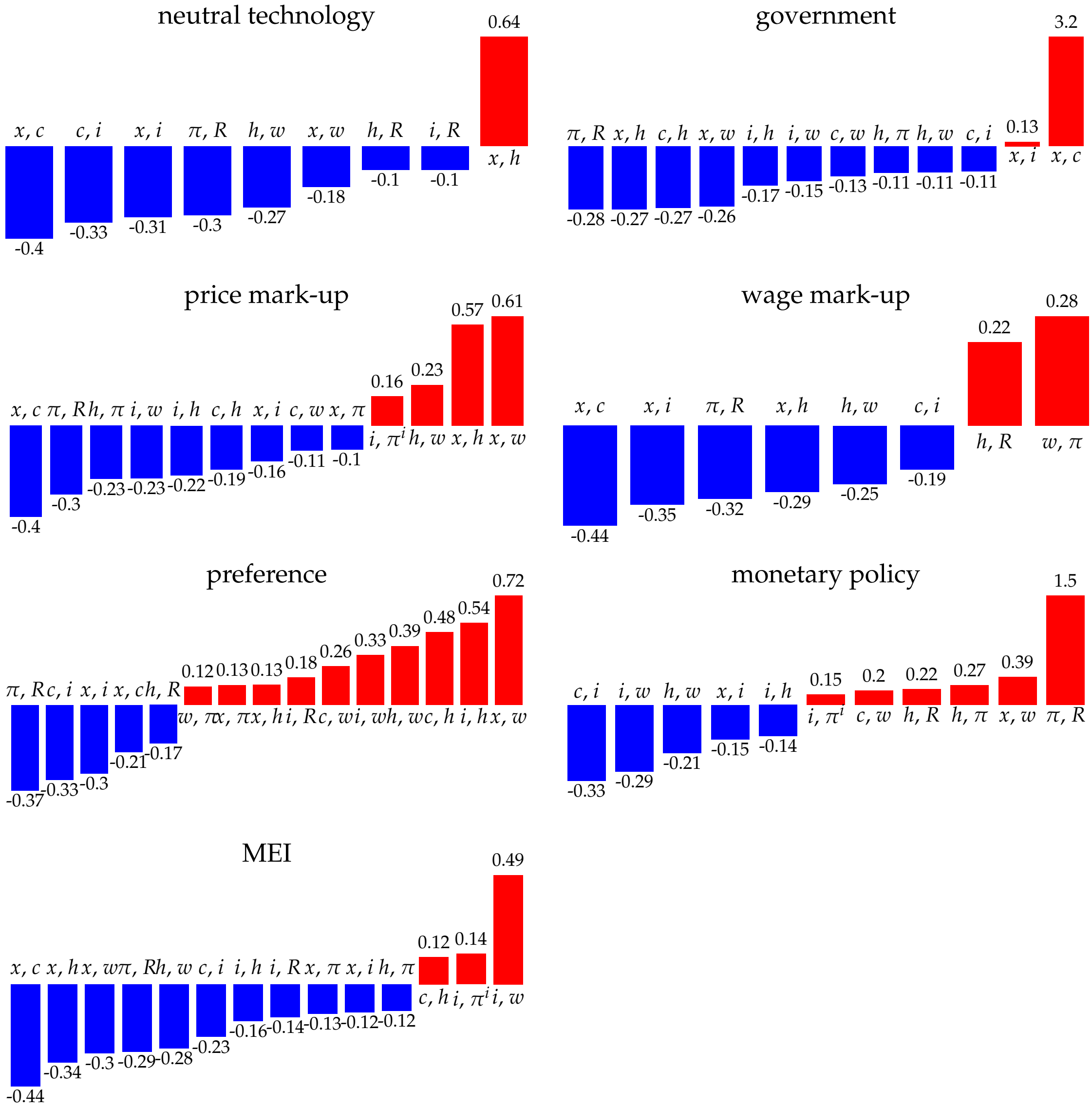

Figure B6: Largest conditional pairwise information complementarities, low spectrum.

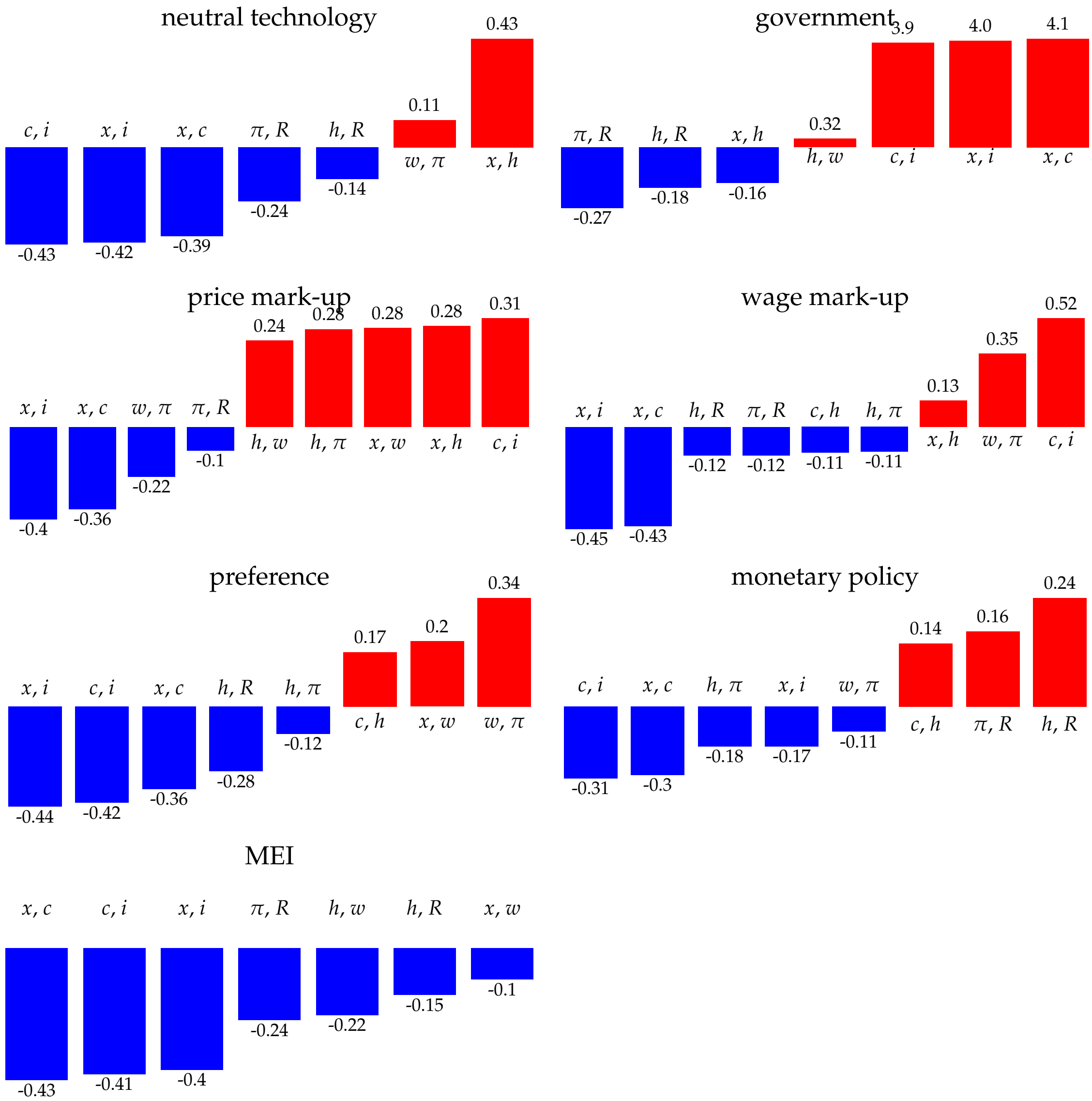

Figure B7: Largest conditional pairwise information complementarities, BC spectrum.

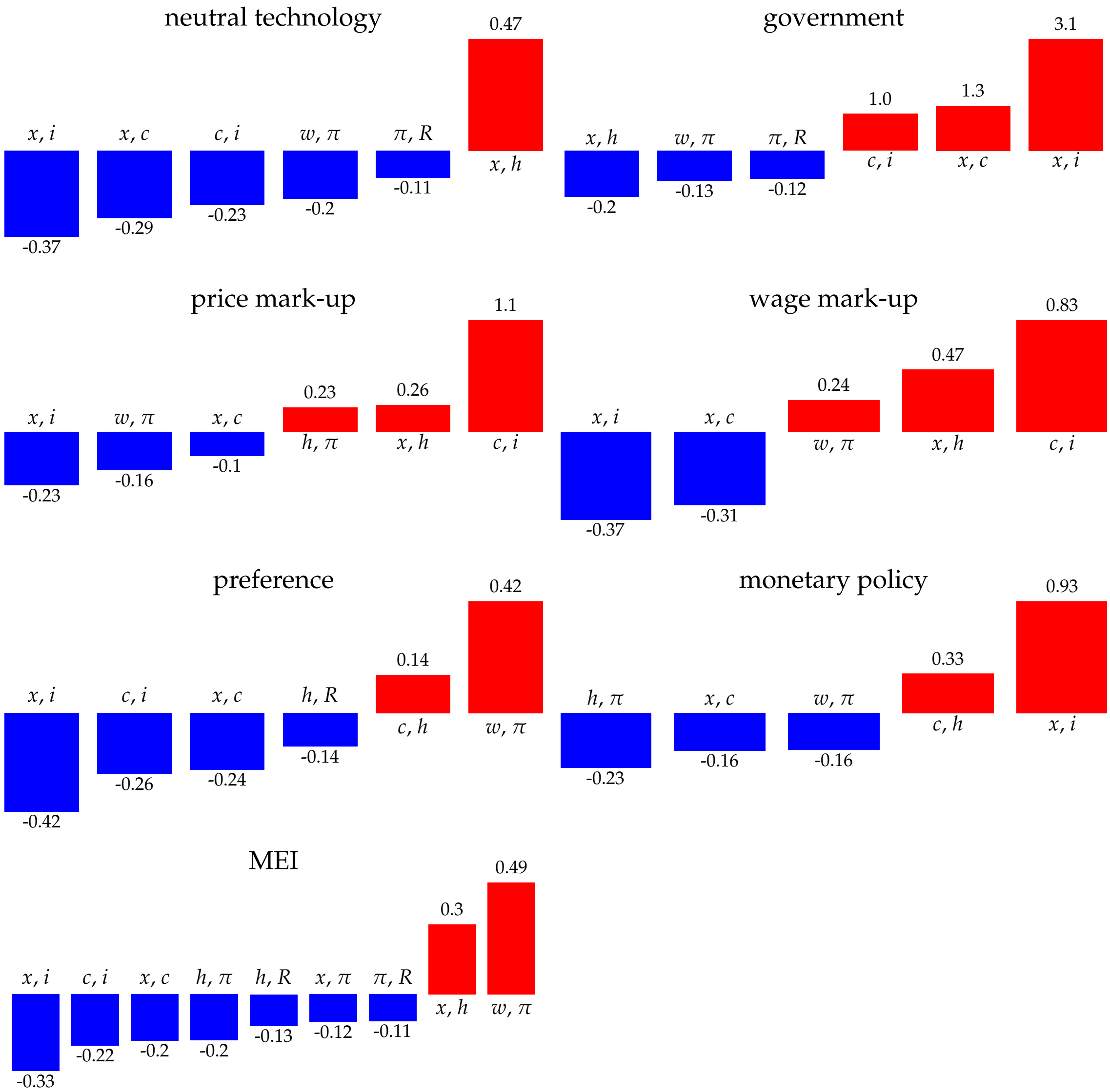

Figure B8: Largest conditional pairwise information complementarities, high spectrum.

C ?

C.1 Linearized equilibrium conditions

The economy consists of a continuum of islands and a mainland. Each island contain a representative household and a continuum of monopolistically competitive firms producing a differentiated commodity using labor and capital provided by the household. These commodities are combined through a CES aggregator into an island-specific composite good, which in turn enters the production of the final good in the mainland through another CES aggregator. The final good is used for consumption and investment. The log-linearized equilibrium conditions with variables presented as log-deviations from their steady-state values are summarized as follows:

Optimal consumption allocation

$$\mathrm{E}_{it}\left[\zeta_t^c + \nu n_{it}\right] = \zeta_t^c - \frac{c_{it} - bC_{t-1}}{1-b} + \mathrm{E}_{it}\left[s_{it} + \varrho Y_t + (1-\varrho)y_{it} - n_{it}\right], \quad \text{(C.1)}$$

where c_{it} and C_t are consumption on island i and aggregate consumption, y_{it} and Y_t are the quantity of the final good produced in island i and aggregate output, n_{it} is hours worked, s_{it} denotes the realized markup in island i, and ζ_t^c is a preference shock. The parameter ν determines the inverse labor supply elasticity, and the parameters b and ϱ denote the degree of habit persistence, and the degree of substitutability across the islands' composite goods in the production of the production of the final good, respectively.

Optimal investment decision

$$\begin{aligned}\mathrm{E}_{it}\left[\lambda_{it} + q_{it}\right] = \ & \mathrm{E}_{it}\left[\lambda_{it+1} + \beta(1-\delta)q_{it+1} + (1-\beta(1-\delta))(s_{it+1} + \varrho Y_{t+1}\right. \\ & \left. + (1-\varrho)y_{it+1} - u_{it+1} - k_{it+1})\right] \quad \text{(C.2)}\end{aligned}$$

where q_{it} is the price of capital, u_{it} is the rate of capital utilization, and λ_{it} is the marginal utility of consumption, given by

$$\lambda_{it} = \zeta_t^c - \frac{c_{it} - bC_{t-1}}{1-b} \quad \text{(C.3)}$$

The parameter β is the intertemporal discount rate in the utility function of the households, and δ is the depreciation rate.

Optimal bond holdings decision

$$R_t = \zeta_t^c - (1+\nu)n_{it} - s_{it} - \varrho Y_t - (1-\varrho)y_{it} - \mathrm{E}'_{it}[\lambda_{it+1} - \pi_{it+1}] \quad \text{(C.4)}$$

where R_t is the nominal interest rate and π_{it} is the inflation rate in island i.

Equilibrium price of capital

$$q_{it} = (1+\beta)\varphi\iota_{it} + \varphi\iota_{t-1} - \beta\varphi\,\mathrm{E}'_{it}\,\iota_{it+1} + \zeta_t^{IP} - \zeta_t^{IT} \quad \text{(C.5)}$$

where i_{it} denotes the level of investment, ζ_t^{IP} is non-stationary investment-specific technology shock, ζ_t^{IT} is a stationary shock shifting the demand for investment, and φ is a parameter governing the size of investment adjustment costs.

Production function

$$y_{it} = \zeta_t^A + \alpha(u_{it} + k_{it}) + (1-\alpha)n_{it} \tag{C.6}$$

where k_{it} is the local capital stock, ζ_t^A is the level of aggregate TFP, and α is the share of capital in the production function. The capital accumulation equation is

$$k_{it+1} = (1-\delta)k_{it} + \delta(\zeta_t^{IT} + \iota_{it}), \tag{C.7}$$

and level of TFP is the sum of a permanent (a_t^p) and a transitory (a_t^τ) component:

$$\zeta_t^A = a_t^p + a_t^\tau, \tag{C.8}$$

Resource constraint

$$\varrho y_t + (1-\varrho)y_{it} = x_{it} + \alpha u_{it}, \tag{C.9}$$

where x_{it} denotes GDP on island i, given by

$$x_{it} = s_c c_{it} + (1 - s_c - s_g)(\zeta_t^{IP} + \iota_{it}) + s_g G_t, \tag{C.10}$$

and G_t, s_c and s_g denote the level of government spending and the steady-state ratios of consumption and government spending to output. To ensure the existence of a balanced growth path, government spending is defined as

$$G_t = \zeta_t^g + \frac{1}{1-\alpha}a_t^p - \frac{\alpha}{1-\alpha}\zeta_t^{IP} \tag{C.11}$$

where ζ_t^g a government spending shock.

Equilibrium utilization

$$\zeta_t^{IP} + \frac{1}{1-\psi}u_{it} = s_{it} + \varrho y_t + (1-\varrho)y_{it} - k_{it}, \tag{C.12}$$

where ψ is a capital utilization elasticity parameter.

Inflation rate

$$\pi_{it} = \frac{(1-\chi)(1-\beta\chi)}{\chi\,(1+\chi(1-\beta))}s_{it} + \frac{\beta\chi(1-\chi)\pi_t + \beta\chi\,\mathrm{E}'\,\pi_{it+1}}{\chi\,(1+\chi(1-\beta))}, \tag{C.13}$$

where Π_{it} is the aggregate inflation rate, and $(1-\chi)$ is the probability that a firm resets its price in a given period.

Monetary policy rule

$$R_t \;=\; \kappa_R R_{t-1} + (1-\kappa_R)(\kappa_\pi \pi_{it} + \kappa_y(x_{it} - x_{it}^F)) + \zeta_t^m \tag{C.14}$$

where x_{it}^F denotes the GDP that would be attained in a flexible-price allocation, ζ_t^m is a monetary policy shock, κ_π and κ_y are parameters determining the policy rate reaction to inflation and the output gap and κ_{Ri} controls the degree of interest-rate smoothing. The flexible-price allocations are obtained from equations (C.1) – (C.12) by setting the realized markup to zero ($s_{it} = 0$) and replacing R_t in (C.4) with the real interest rate.

It is worth pointing out that there are two different subjective expectation operators E_{it} and E'_{it} in the above conditions. In the model, each time period t is divided into two stages: in stage 1, the inhabitants of each island receive an unbiased signal about the level of TFP in that period, and form beliefs that firms and households on other islands receive a signal that is biased by the confidence shock ξ_t, which is also observed. In stage 2, the true state of nature and the realized value of economic activity is publicly revealed. ACD discuss two protocols for the timing of decisions of firms and households, depending on whether supply is determined first and prices adjust to make demand meet supply, or whether demand is determined first and supply adjusts to meet demand. The model presented above is estimated under the second assumption, as seen by the use of stage 1 expectations in the optimality conditions for consumption and saving in equations (C.1), (C.2), and stage 2 expectations in equations (C.4), (C.5), (C.13).

There are nine shocks in the model: a permanent (a_t^p) and a transitory (a_t^τ) TFP shock; a permanent (ζ_t^{IP}) and a transitory (ζ_t^{IT}) investment-specific shock; a news shock regarding future productivity (a_t^n); a discount-rate shock (ζ_t^c); a government-spending shock (ζ_t^g); a monetary policy shock (ζ_t^m); and a confidence shock (ξ_t). The later shock is an exogenous random variable observed in stage 1 of each period, representing the perceived bias in the other islands' signals about the level of TFP in that period. The permanent TFP shock is given by

$$a_t^p \;=\; a_{t-1}^p + a_{t-1}^n + \varepsilon_t^p, \tag{C.15}$$

and the permanent investment-specific shock follows a random walk

$$\zeta_t^{IP} \;=\; \zeta_{t-1}^{IP} + \varepsilon_t^{IP}, \tag{C.16}$$

where ε_t^p and ε_t^{IP} are i.i.d. innovations. All remaining shocks are stationary AR(1) processes.

Table B1: Parameter values, ACD (2018) model

	parameter	posterior median
ψ	utilization elasticity	0.500
ν	inverse labor supply elasticity	0.282
α	capital share	0.255
φ	investment adjustment costs	3.312
b	habit persistence	0.758
χ	Calvo parameter,	0.732
κ_R	Taylor rule smoothing,	0.198
κ_π	Taylor rule inflation,	2.271
κ_y	Taylor rule output,	0.121
ρ_m	AR mon. policy	0.647
ρ_a	AR transitory TFP component	0.412
ρ_n	AR news	0.224
ρ_i	AR transitory investment-specific technology	0.374
ρ_c	AR preference	0.888
ρ_g	AR government spending	0.786
ρ_ξ	AR confidence	0.833
σ_a^P	std. permanent TFP component	0.406
σ_a^T	std. transitory TFP component	0.347
σ_n	std. news	0.378
σ_i^P	std. permanent investment-specific technology	0.610
σ_i^T	std. transitory investment-specific shocks	5.805
σ_c	std. preference	0.357
σ_g	std. government spending	1.705
σ_ξ	std. confidence	0.613
σ_m	std. mon. policy	0.313