Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

DISSERTATION

Essays on monetary policy transmission

Author: **Marek Rusnák**

Advisor: **prof. Roman Horváth, Ph.D.**

Academic Year: **2017/2018**

Thank you

Coming up with good idea to write a paper is challenging. Estimating a model and writing up a paper into a publishable form is even harder. Rewriting a paper to satisfy numerous - sometimes contradicting - comments and suggestions from referees and editors seems impossible at times. During my graduate studies I was lucky to be able to write and eventually publish a good number of papers.

I doubt this would be ever possible without the support of my advisor, Roman Horváth, and my colleague and friend, Tomáš Havránek. Roman encouraged me to do research ever since my undergraduate studies and has given me many great pieces of advice. Tomáš and his ability to write clearly and succinctly, together with his enthusiasm about economic research, was source of constant inspiration. For this I would like to thank them.

Further, I would like to thank co-authors of papers not included in this dissertation for the opportunity to cooperate with them and learn from them: Jan Babecký, Martin Feldkircher, Michal Franta, Marc Joy, Zuzana Havránková, David Havrlant, Jakub Matějů, Kateřina Šmídková, Anna Sokolova, Bořek Vašíček, and Jan Zápal.

I would also like to express my gratitude to the Institute of Economic Studies for providing me with the opportunity to pass the knowledge accumulated while doing these papers by allowing me to teach the course on Bayesian econometrics and state space models named *Tools for Modern Macroeconometrics.*[1]

I would also like to thank all members of my PhD committee for their invaluable suggestions in their referee reports and during the pre-defense.

Finally, I would like to dedicate this dissertation to my wife Dita and my son Jonáš.

Typeset in LaTeX 2_ε using the IES Thesis Template.

Bibliographic Record

Rusnák, Marek: *Empirical essays on monetary policy transmission.* Dissertation thesis. Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, 2018, 247 pages. Advisor: prof. Roman Horváth, Ph.D.

[1] This dissertation also benefited from the European Union's Horizon 2020 Research and Innovation Staff Exchange programme under the Marie Sklodowska-Curie grant agreement No 681228.

Contents

Abstract

The dissertation consists of three papers presenting applications of meta-analysis in macroeconomics and two papers dealing with real-time data properties and forecasting. The first two papers examine the effect of monetary policy on price level, while the third paper investigates the habit formation in consumption. Fourth paper presents a model to nowcast Czech GDP in real time, while the last paper looks at the properties of data revisions to Czech national accounts.

In the first paper we investigate a meta-analysis of the effect of monetary policy on price level, focusing on the so-called price-puzzle. We collect and examine about 1,000 point estimates of impulse responses from 70 articles that use vector autoregressions to study monetary transmission in various countries. We find that the puzzle is created by model misspecifications: especially by the omission of commodity prices, neglect of potential output, and reliance on recursive identification. Our results also suggest that the strength of monetary policy depends on the country's openness, phase of the economic cycle, and degree of central bank independence.

The transmission of monetary policy to the economy is generally thought to have long and variable lags. In the second paper we quantitatively review the modern literature on monetary transmission to provide stylized facts on the average lag length and the sources of variability. We collect 67 published studies and examine when prices bottom out after a monetary contraction. The average transmission lag is 29 months, and the maximum decrease in prices reaches 0.9% on average after a one-percentage-point hike in the policy rate. Transmission lags are longer in developed economies (25–50 months) than in post-transition economies (10–20 months). We find that the factor most effective in explaining this heterogeneity is financial development: greater financial development is associated with slower transmission.

In the third paper we examine 597 estimates of habit formation reported in 81 published studies. In contrast to previous results for most fields of empirical economics, we find no publication bias in the literature. The mean reported strength of habit formation equals 0.4, but the estimates vary widely both within and across studies. We use Bayesian model averaging to assign a pattern to this variance while taking into account model uncertainty. Studies employing macro data report consistently larger estimates than micro studies: 0.6 vs. 0.1 on average. The difference remains 0.5 when we control for 30 factors that reflect the context in which researchers obtain their estimates, such as data frequency, geographical coverage, variable definition, estimation approach, and publication characteristics. We also find that evidence for habits strengthens when researchers use lower data frequencies, employ log-linear approximation of the Euler equation, and utilize open-economy DSGE models. Moreover, estimates of habits differ systematically across countries.

In the fourth paper we employ a Dynamic Factor Model (DFM) to nowcast Czech GDP. Using multiple vintages of historical data and taking into account the publication lags of various monthly indicators, we evaluate the real-time performance of the DFM over the 2005–2012 period. The main result of this paper is that the accuracy of model-based nowcasts is comparable to that of the nowcasts of the Czech National Bank (CNB). Moreover, combining the DFM and the CNB nowcasts results in more accurate performance than in the case of the individual nowcasts alone. Our results also suggest that foreign variables are crucial for the accuracy of the model, while omitting financial and confidence indicators does not worsen the nowcasting performance.

Frequent revisions to the GDP and its components cause policymakers to face considerable uncertainty about the current state of the economy. In the fifth paper we provide stylized facts about the magnitude of revisions to the Czech national accounts. Using data over the 2003–2012 period, we find that the revisions are rather large. We investigate whether the revisions could have been predicted using the information available at the time of announcement. We find evidence for in-sample predictability for most of the variables, suggesting that the first releases of these variables are not efficient predictors of the actual values. In a real-time out-of-sample exercise, however, we find that the revisions to real GDP, gross fixed capital formation and government consumption are not predictable. Only revisions to GDP deflator can be predicted with substantial gains relative to zero revisions forecasts.

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation contains five separate papers. Nonetheless, all of the papers can be linked to the monetary policy transmission mechanism, and are thus of great importance for central bankers. In particular, reliable estimates about the magnitude of the effect the monetary policy on the price level are crucial for optimal decision making. These issues are addressed in the first three papers. The first paper was published in the *Journal of Money, Credit and Banking*, while the second one was published in the *International Journal of Central Banking*. The third paper was published in the *European Economic Review*.

In practice, the optimal decision making is only possible if one has good idea about the initial conditions of the economy. Hence the knowledge about the current state of the economy in real time is equally important. The final two papers of this dissertation therefore tackle the real-time data properties and our ability to forecast the current state of the economy in real time. The fourth paper was published in the *Economic Modelling*, while the fifth one appeared in the *Czech Journal of Economics and Finance (Finance a uver)*.

This dissertation uses meta-analysis as a main tool to systematically analyze the empirical literature on the effects of monetary policy transmission mechanism. Meta-analysis was developed in medical research to synthesize costly clinical trials and over time it proliferated to the social sciences, including economics (Stanley & Jarrell 1989). Unlike narrative literature surveys, meta-analysis allows for a more structured discussion concerning the effect that different methods have on the results. To answer the question what is the empirical literature telling us about the effect of interest, one cannot simply average the collected estimates as this has two major shortcomings. The meta-analysis provides tools to address these shortcomings and to come up with the best estimate suggested by the literature as a whole.

First, the simple average ignores possible publication selection. If some results are more likely to get published than others, the average becomes a biased estimator of the underlying impulse response. For this reason, most meta-analyses test—and, if necessary, correct—for so-called publication bias. Brodeur *et al.* (2016) collect 50,000 p-values reported in economics and document widespread publication bias. Ioannidis *et al.* (2017) survey meta-analyses conducted in economics and find that most fields suffer from the bias, as editors, referees, or authors themselves prefer statistically significant results that have an intuitive sign.

Second, the simple average ignores heterogeneity in the results of the primary studies. Since different researchers use different data and methods, and the studies are of different quality, it is unrealistic to assume that all estimates are drawn from the same population. Meta-analysis attempts to take these factors into account. In particular, in addition to controlling for publication bias, meta-analysis aims at capturing the role that structural-, data-, estimation-, and publication characteristics might play in the variation of empirical estimates. As there are many such characteristics, the question of how to decide which of the potential explanatory variables should be included in the final model. This question, however, is not particular to meta-analysis and is encountered by virtually all applied econometricians.

There are many different approaches to variable selection. The prime example is sequential t-testing (sometimes called the "general-to-specific approach"), which is often used to decide which variables belong to the underlying model. Nevertheless, this approach is not statistically valid and gives rise to the possibility of excluding relevant variables (Koop 2003). Similarly, handpicking the variables based on how well the estimates conform to expected signs is commonly used approach that, however, risks the possibility that wrong model is arbitrarily selected (see, for example, Gross & Poblacion 2015). In this regard, a recent survey among the members of the European Economic Association, Necker (2014), reveals that a third of economists in Europe admit that they have engaged in presenting empirical findings selectively so they confirm their arguments and in searching for control variables until they get a desired result. In case of the large number of potential variables thus brings about problems related to model uncertainty that could result in severely erroneous inference. Other alternatives focused on variable selection include shrinkage approaches (e.g. ridge regression, lasso, elastic net), but these are not yet commonly used, although becoming more popular over time, mainly in forecasting applications (Korobilis 2013; Kim & Swanson 2014; Li *et al.* 2015; Chan-Lau 2017).

Our preferred method to tackle the problem of many potential explanatory variables and resulting model uncertainty is model averaging, in particular its Bayesian implementation - Bayesian model averaging (BMA). Inference in BMA is based on a weighted average of individual regressions that include different combinations of explanatory variables; the weights reflect the posterior model probabilities of the corresponding individual specifications. Posterior model probabilities can be thought of as a Bayesian analogy of information criteria used in frequentist econometrics (at least under certain assumptions, such as that model shocks are Gaussian). Researchers typically want to check the robustness of their results by estimating several regressions that include different combinations of explanatory variables; BMA generalizes this approach in a formal manner.

In contrast to sequential testing, BMA does not require selecting one individual specification and by averaging models allows to account for model uncertainty, thus insuring against selecting a wrong model. It makes sure that the model uncertainty is not ignored and does not lead to overestimation of the precision of estimates (Claeskens & Hjort 2008). In addition, conceptual connection with meta-analysis can be made in two ways. First, meta-analysis attempts to correct for publication bias that arises from preference for statistically significant results that are in line with economic theory. Using model averaging mitigates the possibility that the bias would arise in meta-regression estimation. Second, meta-analysis integrates information from many models, data, and publications in an attempt to provide more robust and complete estimates. In similar spirit, model averaging uses information from many models to come up with the robust estimate.

Nevertheless, several issues might arise when BMA is used. First, BMA does not discriminate among models that may or may not be satisfactory, e.g. estimates might be of incredible sign (not aligned with the theory) or a particular model might not satisfy statistical criteria (e.g. assumptions on the homoscedasticity) beyond using posterior model probabilities. The issue of expected signs can be in general addressed by adjusting priors (models with particular set of signs are omitted i.e. obtain zero posterior model probability). Second, taking into account a large number of various models has a consequence that the method is more complicated than estimation of a single regression. Modern algorithms, however, are able to alleviate this issue, and are able to approximate the whole model space in a feasible manner using Monte Carlo Markov Chain Methods. Third, model averaging relies on the assumption that one of the models in the list is the true model. When this assumption does not hold and all the models are only approximations of the true mode, the interpretation of posterior model probabilities is less clear. This issue is, nevertheless, present also in traditional model selection techniques. The computational complexity render more rigorous averaging of various model classes (linear, non-linear, panel regressions, regime-switching) infeasible, although attempts exist (e.g. application to full instrumental variables estimation Koop *et al.* 2012) or heuristic approach in form of combination of forecasts from different model classes (Bjornland *et al.* 2012). Finally, various model and parameter priors might lead to different posterior inclusion probabilities, so one needs to perform sensitivity analysis to ensure that the results are not sensitive to the choice of priors.

Finally, as regards to the relationship between model averaging and best practice, it is crucial to obtain most robust estimates given data, this in turn allows for reliable estimates of the best practice. In principle, one could consider averaging the best practice estimate by assuming all sensible combinations. However, while in meta-regression we often do not have string prior about what might systematically drive variation in estimates or have many different options, in best practice there often is an indication from the literature about the appropriateness of using certain data or methods. In any case, this could be an interesting avenue for future research.

The first paper of the dissertation focuses on the effects of monetary tightening on the price level. One of the major peculiarities of vector autoregressions, the dominant framework for the empirical analysis of monetary policy, is the counterintuitive rise in prices often reported in these models following a monetary contraction. The so-called price puzzle is encountered by about half of all empirical studies, and in many of them the puzzle is even statistically significant. In the first paper we collect 70 published studies using vector autoregressions to examine the effects of monetary policy. Employing meta-regression analysis, a quantitative method of research synthesis, we investigate which aspects of methodology systematically contribute to reporting the price puzzle. The meta-regression analysis also shows how the characteristics of the countries examined influence the reported shape of the impulse responses and thus help explain the cross-country heterogeneity in monetary transmission.

We evaluate the reported graphs of impulse responses at five time horizons (representing the short, medium, and long run) and for each horizon extract the numerical value of the impulse response. In this way we collect more than 1,000 estimates, 210 on average for each horizon; the estimates summarize evidence from 31 countries and were produced by 103 researchers. We present a method of research synthesis suitable for graphical results such as impulse responses and employ modern meta-analysis methods to examine the extent of publication selection bias (the preference of authors, editors, or referees for some particular results based on significance or consistency with theory).

Our results indicate some evidence of publication selection against the price puzzle, and the selection seems to strengthen for responses with longer horizons after monetary tightening. The finding is in line with Doucouliagos & Stanley (2013), who suggest that publication selection is likely to be stronger for research areas with less theory competition. In macroeconomics, agreement exists about the effects of monetary policy on prices in the long run: prices should eventually decrease after a contraction. On the other hand, a smaller consensus arises regarding the exact effects of monetary policy in the short run because of the uncertainty caused by transmission lags. Published results often exhibit the price puzzle for the short run; on the contrary, results showing the price puzzle for the long run would be difficult to publish.

The reported impulse responses are systematically affected by study design and country-specific characteristics. Study design is important in particular for the short run: the reported short-run increase in prices after a tightening is well explained by the effects of commonly questioned aspects of methodology, such as the omission of commodity prices, the omission of potential output, or the use of recursive identification. When these aspects of methodology are filtered out, the average impulse-response function inferred from the entire literature becomes hump-shaped with no evidence of the price puzzle. Based on such "best-practice" impulse response the maximum decrease in prices following a one percentage-point increase in the interest rate is 0.33% and occurs already half a year after the tightening.

Our results suggest that heterogeneity between countries is important for the long-run response of prices to monetary policy action. Structural characteristics such as GDP growth, average inflation, and openness, as well as institutional characteristics such as financial development and central bank independence, determine the strength of transmission.

For central bankers, an important thing to know is the time after which changes in the policy rate reach the maximum influence on the price level. For example, if the central bank intends to curb inflation, it needs information on how long it takes before the price level is fully affected by the hike in the interest rate. This delay between the monetary policy action and the maximum effect on the economy is called the transmission lag of monetary policy.

The transmission lag of monetary policy is usually estimated using the vector-autoregression framework, which produces graphs of the evolution of the price level in response to a change in the interest rate. These graphs are called impulse response functions and form the basis of the empirical investigation of monetary policy transmission. Yet the transmission lags estimated by different vector-autoregression models vary greatly.

In the second paper we collect the reported impulse-response functions from 67 comparable studies corresponding to many different countries and explore three problems. First, we would like to know whether study design influences the reported transmission lag in a systematic way. Some aspects of study design are considered misspecifications by many researchers (for example the omission of commodity prices, the neglect of potential output, and the reliance on recursive identification) and have been found to affect the reported strength of monetary policy (Rusnak *et al.* 2013). Second, we investigate whether transmission lags vary across countries. If the lags are country-specific, we would like to find out which country characteristics are associated with the heterogeneity. Third, we are interested in the average transmission lag identified in the literature.

To examine these three problems we employ meta-analysis, the quantitative method of literature surveys. Our results suggest that, first, study design matters for the reported transmission lag of monetary policy. For example, we find that the use of monthly data instead of quarterly data makes researchers report faster transmission.

Second, transmission lags are highly heterogeneous across countries. In developed economies the lags vary between 25 and 50 months, while in post-transition countries the lags are much shorter: between 10 and 20 months. We find that the country characteristic that is the most effective in explaining this heterogeneity is financial development: in developed countries financial institutions have more opportunities to hedge against surprises in monetary policy stance, causing greater delays in the transmission of monetary policy shocks. Third, the average transmission lag, corrected for misspecifications in the literature, is 29 months. In other words, for an average country in our sample the price level bottoms out about two and a half years after a monetary contraction.

The third paper of the dissertation focuses on habit formation in consumption, as it is a key component of the modern structural models used by central banks around the world to evaluate the effects of various policy measures. As shown by Fuhrer (2000), the observed inflation dynamics are consistent with a large habit formation coefficient. Furthermore, habit formation helps explain various empirical regularities: the risk-free rate puzzle (Campbell & Cochrane 1999), the equity premium puzzle (Abel 1990), and the happiness puzzle (Choudhary *et al.* 2012).

Habits in consumption can assume two forms: internal and external. Internal habit formation arises when a consumer becomes accustomed to a certain level of consumption, comparing current consumption with consumption in the previous period. In other words, the consumer's utility is no longer a function of current consumption, but one of consumption growth, with past consumption reducing present utility: more food today makes the consumer hungrier tomorrow. In contrast, external habit formation describes "keeping up with the Joneses": the consumer's utility depends on the difference between her consumption and the consumption of a reference group (such as people in the town where she lives).

Dozens of researchers have attempted to estimate the strength of habit formation, but their results vary widely and it is not clear what values should be used for the calibration of stylized models. In the third paper we collect the published estimates and perform a quantitative review of the literature. We find that the average reported estimate is close to 0.4, which is consistent with moderate habit formation, but does not suffice to explain some of the major puzzles in economics, such as the equity premium puzzle. Remarkably, the literature does not seem to be plagued with publication bias. Our results suggest that micro estimates of habit formation tend to be substantially smaller than macro estimates—by about 0.5.

The difference remains 0.5 when we control for 30 factors that reflect the context in which researchers obtain their estimates, such as data frequency, geographical coverage, variable definition, estimation approach, and publication characteristics. We also find that evidence for habits strengthens when researchers use lower data frequencies, employ log-linear approximation of the Euler equation, and utilize open-economy DSGE models. Moreover, estimates of habits differ systematically across countries.

Because of sizeable publication delays in the release of new data on gross domestic product (GDP), timely estimates of current-quarter GDP (so-called nowcasts) are crucial for policymakers assessing the state of the economy in real time. Obtaining these nowcasts is not straightforward because of the peculiar structure of real-time data, as characterized by unbalanced datasets at the end of the sample, data sampled at different frequencies, and substantial data revisions.

The recently developed nowcasting framework of Giannone *et al.* (2008) can deal with these real-time issues by casting a dynamic factor model in a state-space framework. In addition to the ability of the framework to deal with unbalanced datasets and mixed frequencies, it can utilize a potentially large set of variables by summarizing macroeconomic comovements by a few common factors.

In the fourth paper, we evaluate the performance of the dynamic factor model when applied to nowcasting Czech GDP over the 2005–2012 period, using multiple vintages of real-time data. The model utilizes 28 headline macroeconomic variables. In addition to so-called hard data covering the production, sales, labor, and trade sectors of the economy, we include a handful of financial variables and confidence indicators. These are potentially useful because of their timeliness. Furthermore, we add several foreign variables to account for the fact that the Czech Republic is a small open economy.

Our results suggest that the dynamic factor model can compete successfully with the nowcasts of the Czech National Bank (CNB). Furthermore, the results indicate that the dynamic factor model provides useful additional information relative to the nowcasts of the CNB, since combining the two nowcasts results in smaller forecasting errors on average. We also find that the inclusion of foreign variables is crucial for the accuracy of the model. On the other hand, excluding financial variables and confidence indicators does not result in a substantial deterioration of the nowcasting accuracy.

We also show how one can decompose changes in the nowcasts into different news coming from newly published data. Moreover, we show that the dynamic factor model can be used successfully to nowcast other variables, such as expenditure components of the Czech national accounts. Finally, we find that the forecasting performance of the DFM at longer horizons (up to six quarters ahead) is comparable to that of the official CNB predictions.

Frequent revisions to the GDP and its components cause policymakers to face considerable uncertainty about the current state of the economy. In the fifth paper, we provide stylized facts about the magnitude of revisions to the Czech national accounts. Using data over the 2003–2012 period, we find that the revisions are rather large.

Revisions to real GDP growth are on average 1.4 for annualized quarterly growth rate and 0.7 percentage points for annual growth rate. Revisions to other variables are even larger: the average size of revisions range from 1 to 12 percentage points for annualized quarterly growth rates and from 0.5 to 4 percentage points for annual growth rates. We investigate whether the revisions could have been predicted using the information available at the time of announcement.

We find evidence for in-sample predictability for most of the variables, suggesting that the first releases of these variables are not efficient predictors of the actual values. In a real-time out-of-sample exercise, however, we find that the revisions to real GDP, gross fixed capital formation and government consumption are not predictable. Only revisions to GDP deflator can be predicted with substantial gains relative to zero revisions forecasts.

References

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching Up with the Joneses." *American Economic Review* **80(2)**: pp. 38–42.

Bjornland, H. C., K. Gerdrup, A. S. Jore, C. Smith, & L. A. Thorsrud (2012): "Does Forecast Combination Improve Norges Bank Inflation Forecasts?" *Oxford Bulletin of Economics and Statistics* **74(2)**: pp. 163–179.

Brodeur, A., M. Le, M. Sangnier, & Y. Zylberberg (2016): "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.

Campbell, J. Y. & J. Cochrane (1999): "Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* **107(2)**: pp. 205–251.

Chan-Lau, J. A. (2017): "Lasso Regressions and Forecasting Models in Applied Stress Testing." *Technical Report 17/108*, IMF Working Papers.

Choudhary, A., P. Levine, P. McAdam, & P. Welz (2012): "The happiness puzzle: analytical aspects of the Easterlin paradox." **64(1)**: pp. 27–42.

Claeskens, G. & N. L. Hjort (2008): *Model Selection and Model Averaging*. Number 9780521852258 in Cambridge Books. Cambridge University Press.

Doucouliagos, H. & T. D. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Fuhrer, J. C. (2000): "Habit Formation in Consumption and Its Implications for Monetary-Policy Models." *American Economic Review* **90(3)**: pp. 367–390.

Giannone, D., L. Reichlin, & D. Small (2008): "Nowcasting: The real-time informational content of macroeconomic data." *Journal of Monetary Economics* **55(4)**: pp. 665–676.

Gross, M. & G. F. J. Poblacion (2015): "A false sense of security in applying handpicked equations for stress test purposes." *Working Paper Series 1845*, European Central Bank.

Ioannidis, J. P. A., T. D. Stanley, & H. Doucouliagos (2017): "The power of bias in economics research." *The Economic Journal* **127(605)**: pp. F236–F265.

Kim, H. H. & N. R. Swanson (2014): "Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence." *Journal of Econometrics* **178(P2)**: pp. 352–367.

Koop, G. (2003): *Bayesian Econometrics.* John Wiley & Sons.

Koop, G., R. Leon-Gonzalez, & R. Strachan (2012): "Bayesian model averaging in the instrumental variable regression model." *Journal of Econometrics* **171(2)**: pp. 237–250.

Korobilis, D. (2013): "Hierarchical shrinkage priors for dynamic regressions with many predictors." *International Journal of Forecasting* **29(1)**: pp. 43–59.

Li, J., I. Tsiakas, & W. Wang (2015): "Predicting Exchange Rates Out of Sample: Can Economic Fundamentals Beat the Random Walk?" *Journal of Financial Econometrics* **13(2)**: pp. 293–341.

Necker, S. (2014): "Scientific misbehavior in economics." *Research Policy* **43(10)**: p. 1747–1759.

Rusnak, M., T. Havranek, & R. Horvath (2013): "How to Solve the Price Puzzle? A Meta-Analysis." *Journal of Money, Credit and Banking* **45(1)**: pp. 37–70.

Stanley, T. D. & S. B. Jarrell (1989): "Meta-Regression Analysis: A Quantitative Method of Literature Surveys." *Journal of Economic Surveys* **3(2)**: pp. 161–70.

Chapter 2

How to Solve the Price Puzzle? A Meta-Analysis

Abstract

The short-run increase in prices following an unexpected tightening of monetary policy constitutes a puzzle frequently reported in empirical studies. Yet the puzzle is easy to explain away when all published models are quantitatively reviewed. We collect and examine about 1,000 point estimates of impulse responses from 70 articles that use vector autoregressions to study monetary transmission in various countries. We find that the puzzle is created by model misspecifications: especially by the omission of commodity prices, neglect of potential output, and reliance on recursive identification. Our results also suggest that the strength of monetary policy depends on the country's openness, phase of the economic cycle, and degree of central bank independence.

The paper was co-authored with Tomas Havranek and Roman Horvath and published in the *Journal of Money, Credit and Banking* [2013, 45(1), pp. 37–70]. We are grateful to Adam Elbourne, Bill Gavin, Jakob de Haan, Tomas Holub, Sergey Slobodyan, Tom Stanley, Harald Uhlig, and two anonymous referees for providing data or comments on previous versions of the manuscript. The paper benefited from discussions at the Meta-Analysis of Economics Research Colloquium 2010, Conway; the Scottish Economics Society Annual Conference 2011, Perth; the International Conference on Macroeconomic Analysis and International Finance 2011, Rethymno; and the Annual Congress of the European Economic Association 2011, Oslo. All remaining errors are ours. We acknowledge financial support from the Czech Science Foundation (grant #P402/11/1487) and the Grant Agency of Charles University (grant #267011). An online appendix is available at `meta-analysis.cz/price_puzzle`.

2.1 Introduction

How does monetary policy affect the price level? This fundamental question of monetary economics still ranks among the most controversial when it comes to empirical evidence. Although intuition and stylized macro models suggest that prices should decrease following a surprise increase in interest rates, empirical findings often challenge the theory. About 50% of modern studies using vector autoregressions (VARs) to investigate the effects of monetary policy report that after a tightening prices actually increase—at least in the short run. Beginning with Sims (1992), many different solutions to the "price puzzle" have been proposed, varying from alleged misspecifications of VARs (Giordani 2004; Bernanke *et al.* 2005) to theoretical models that try to justify the observed rise in prices (Barth & Ramey 2002; Rabanal 2007).

Depending on the point of view, the price puzzle casts serious doubt on either the ability of VAR models to correctly identify monetary policy shocks, or the ability of central banks to control inflation in the short run, or both. Since macroeconomists have produced a plethora of empirical research on the topic, it seems natural to ask what general effect the literature implies. The method designed to answer such questions is meta-analysis, a quantitative method of research synthesis commonly used in economics (Smith & Huang 1995; Stanley 2001; Disdier & Head 2008; Card *et al.* 2010; Chetty *et al.* 2011). In contrast to narrative literature surveys, meta-analysis takes into account possible publication selection: the preference of authors, editors, or referees for results that are statistically significant or consistent with the theory, a bias that has become a great concern in empirical economic research (DeLong & Lang 1992; Card & Krueger 1995; Ashenfelter & Greenstone 2004; Havranek & Irsova 2011).

Meta-analysis enables researchers to examine the systematic dependencies of reported results on study design and to separate the wheat from the chaff by filtering out the effects of misspecifications. Meta-analysis can create a synthetic study with ideal parameters, such as the maximum breadth of data or a consensus best-practice methodology, and, in our case, estimate the underlying effect of monetary policy corrected for misspecification and other biases. Furthermore, meta-analysis makes it possible to investigate how the strength of monetary transmission depends on the characteristics of the countries examined. In this paper we attempt to collect all published studies examining monetary transmission within a VAR framework and extract point estimates of impulse responses together with the corresponding confidence bounds. We investigate the degree of publication selection, the role of model misspecification for the occurrence of the price puzzle, and the factors underlying the heterogeneity of price responses to monetary shocks across countries and over time.

Based on the mixed-effects multilevel model we illustrate how meta-analysis is able to disentangle various factors causing researchers to encounter the price puzzle.

We show that when best practice is followed, the researcher is likely to find that prices decrease significantly soon after a tightening of monetary policy. Our results thus suggest that the puzzle stems from model misspecification rather than from what really happens in the economy. In addition, the results indicate publication selection in favor of the negative responses of prices to a monetary contraction. Finally, our analysis of the determinants of transmission heterogeneity suggests that monetary policy has a stronger effect on prices in more open economies, in countries with a more independent central bank, and during economic downturns.

The remainder of the paper has the following structure. Section 2.2 describes how we collected the estimates from VAR models. Section 2.3 reviews the suggested solutions to the price puzzle. Section 2.4 tests for publication selection bias and for the underlying effect of monetary tightening on prices. Section 2.5 examines the method and structural heterogeneity among impulse responses. Section 2.6 concludes. Appendix A provides additional robustness checks, and Appendix B lists the studies used to construct the data set.

2.2 The Impulse Responses Data Set

Ever since the seminal contribution of Sims (1980), VARs have been the dominant empirical tool for investigating monetary transmission. Researchers using VARs to examine the impact of monetary policy usually assume that the economy can be described by the following dynamic model:

$$AY_t = B(L)Y_{t-1} + \varepsilon_t, \tag{2.1}$$

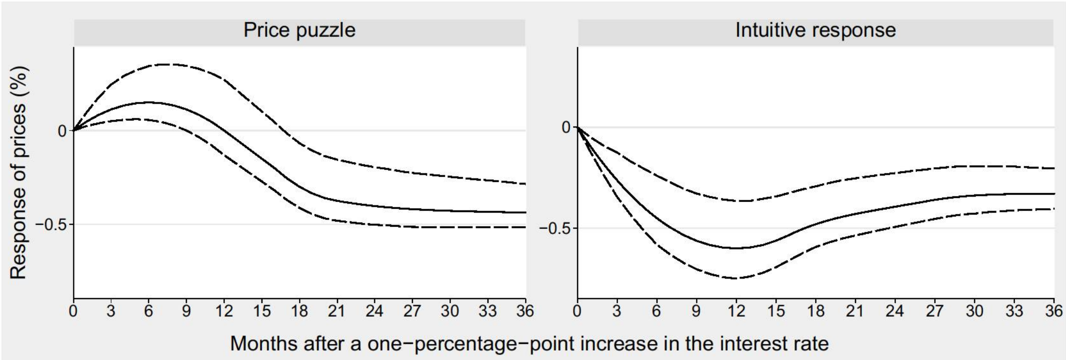
where Y_t is a vector of endogenous variables typically containing a measure of output, prices, interest rates, and, in the case of a small open economy, the exchange rate. Matrix A describes contemporaneous relationships between endogenous variables, $B(L)$ is a matrix lag polynomial, and ε_t is a vector of structural shocks with the variance-covariance matrix $E(\varepsilon_t \varepsilon_t') = I$. The system is called the structural-form VAR. In order to estimate it, researchers rewrite the system to its reduced form:

$$Y_t = C(L)Y_{t-1} + u_t, \tag{2.2}$$

where the elements of matrix $C(L)$ are the convolutions of the elements of matrices A and B, and u_t is a vector of reduced-form shocks with the variance-covariance matrix $E(u_t u_t') = \Sigma$; the relationship between structural shocks and reduced-form residuals is $\varepsilon_t = Au_t$. The dynamic responses of endogenous variables to structural shocks are described by impulse-response functions.

Figure 2.1 presents two stylized types of the price level's impulse responses to a monetary tightening. The left panel demonstrates the price puzzle: prices increase significantly in the short run. In contrast, the right panel shows a response that corresponds with the mainstream prior: the price level declines soon after a tightening.

Figure 2.1: Stylized impulse responses



The first step of meta-analysis is to select the studies to be included. While some meta-analysts use both published and unpublished studies, others confine their sample to journal articles (for instance, Abreu *et al.* 2005). Including working papers and mimeographs in meta-analysis does not help alleviate publication bias: if journals systematically prefer certain results, rational authors will already adopt the same preference in the earlier stages of research as they prepare for journal submission. Indeed, empirical evidence suggests no difference in the magnitude of publication bias between published and unpublished studies (see the meta-analysis of 87 meta-analyses by Doucouliagos & Stanley 2013). Even if there was a difference, modern meta-regression methods not only identify but also filter out the bias. Therefore, as a preliminary and simple criterion of quality, we only consider articles published in peer-reviewed journals or in handbooks (such as the Handbook of Macroeconomics).

The following literature search strategy was employed. First, we examined two narrative surveys (Stock & Watson 2001; Egert & MacDonald 2009) and set up a search query able to capture most of the relevant studies; we searched both the EconLit and RePEc databases. Next, we checked the references of studies published in 2010 and the citations of the most widely cited study in the VAR literature, Christiano *et al.* (1999). After going through the abstracts of all the identified studies, we selected 195 that showed any promise of containing empirical estimates of impulse responses and examined them in detail. The search was terminated on September 15, 2010.

To be able to use meta-analysis methods fully, we exclude the studies that omit to report confidence intervals around impulse responses. Unfortunately, we thus have to exclude some seminal articles such as Sims (1992) or a few recent studies that estimate time-varying-parameter VARs. To obtain a more homogeneous sample we only focus on studies that define a monetary policy shock as a shock in the interest rate. A number of studies investigate the change in the monetary base; since Bernanke & Blinder (1992) and Sims (1992), however, the majority of the literature investigates interest rate shocks because most central banks now use the interest rate as their main policy instrument. We only include studies examining the response of the price level; a minority of studies examine the responses of the inflation rate. These inclusion criteria leave 70 studies in our database. The full list of studies included in the data set can be found in Appendix B, and the list of excluded studies is presented in the online appendix at `meta-analysis.cz/price_puzzle`.

Considering the richness and heterogeneity of the empirical evidence on the effects of monetary policy, it is surprising there has been no quantitative synthesis using modern meta-regression methods.[1] One reason is that the results are typically presented in the form of graphs instead of numerical values, and the graphs contain estimates for many time horizons following the monetary policy shock. Researchers usually investigate up to 36- or 48-month horizons when using monthly data and up to 20 quarters when using quarterly data; it is unclear which horizon should be chosen to summarize the effect.

Our meta-analysis is designed in the following way. We extract responses at 3- and 6-month horizons to capture the short-run effect, at 12- and 18-month horizons to capture the medium-run effect, and at the 36-month horizon to capture the long-run effect. We enlarge the graphs of impulse responses and using pixel coordinates we measure the response and its confidence bounds. The graphs of all impulse responses as well as the extracted values are available in the online appendix. The resulting measurement error is random, similar to the rounding error in numerical outcomes, and thus inevitable in a meta-analysis.

The extracted values must be transformed into a common metric to ensure that the estimates are comparable. To standardize the estimates so as they represent the effect of a one-percentage-point increase in the interest rate, we divide the responses by the magnitude of the monetary policy shock used in the study. (When we were uncertain about the magnitude of shock used in the primary study, we contacted the authors.) In the case of factor-augmented VAR (FAVAR) studies, where the

[1] To our knowledge, there has been one unpublished meta-analysis on the impact of monetary policy on prices (de Grauwe & Storti 2004) and it focused solely on heterogeneity in the reported estimates; that is, it did not filter out publication bias and misspecifications to estimate the underlying impulse response. We also use four times more point estimates of impulse responses and three times more variables to explain heterogeneity.

responses are usually given in standard-deviation units, we normalize the responses by the standard deviation of the particular time series.

Since the confidence intervals around the estimates of impulse responses are often asymmetrical (confidence intervals are usually computed by the Bayesian Monte Carlo integration method; see Sims & Zha 1999), the standard errors of the estimates cannot be obtained directly. In this case we approximate the standard error by the distance from the point estimate to the confidence bound closer to zero; that is, we take the lower confidence bound for positive responses and the upper bound for negative responses. This bound determines significance and would be associated with potential publication selection. Should we use the average of the distance to both confidence bounds, the inference would remain similar; these additional results are available in the online appendix. When the reported confidence interval is presented in standard-deviation units (for example, two standard deviations on both sides), we can immediately approximate the standard error. Otherwise, we proceed as if the estimates were symmetrically distributed and assume that, for example, the 68% confidence interval represents an interval of one standard error around the mean.

Following the recent trend in meta-analysis (Disdier & Head 2008; Havranek & Irsova 2011), we use all reported estimates from the 70 primary studies. Arbitrarily selecting the "best" estimate or using the average reported estimate would discard a great deal of useful information about the differences in methods within one study.

The number of impulse responses collected for each of the horizons is approximately 210, which in total amounts to more than 1,000 point estimates. More specifically, we collect 208 estimates for the 3-month horizon, 215 for the 6- and 12-month horizons, 217 for the 18-month horizon, and 205 for the 36-month horizon. For comparison, consider Nelson & Kennedy (2009), who review 140 economic meta-analyses and report that the median analysis only uses 92 point estimates from 33 primary studies. The oldest study in our sample was published in 1992 and the median study was published in 2006, the data set covers evidence from 31 countries, and we build upon the work of 103 researchers that produced the impulse responses. The median time span of the data used by the primary studies is 1980–2002. All studies in the sample combined receive approximately 800 citations in Google Scholar per year, indicating the influence of VARs in monetary economics.

2.3 Collecting the Pieces of the Puzzle

To motivate the selection of explanatory variables in the multivariate meta-regression analysis (Section 2.5), we now briefly review the methodological solutions to the price puzzle that have been proposed in the literature. Most of these remedies have proven to alleviate the puzzle in some cases; none of them, though, has been fully successful in solving it. Table 2.1 demonstrates that from the 208 estimates collected for the

3-month horizon, exactly half exhibit the price puzzle, and in 15% of the estimates the puzzle is even statistically significant at the 5% level. The table summarizes the effectiveness of the different solutions to the puzzle. Even in the case of the most effective solution, 24% of specifications still exhibit the puzzle (except for sign restrictions, which in some cases represent a tautological solution). Clearly no single misspecification is responsible for the price puzzle. But perhaps the puzzle is associated with a combination of bad method choices. In the following paragraphs we describe why some methods are thought to be better than others and what may help explain the reported puzzle.

Table 2.1: Effectiveness of the suggested solutions to the price puzzle

	Methodology used in the estimation						
	All	Commodity	Trend/Gap	FAVAR	SVAR	Sign	Single
No. of responses estimated	208	125	33	11	60	31	64
Price puzzle present	104	61	8	8	20	3	24
Price puzzle significant	32	16	1	3	6	0	5

Note: Commodity = Commodity prices are included in the VAR, Trend/Gap = time trend or output gap is included, FAVAR = a factor-augmented VAR is estimated, SVAR = non-recursive identification is used, Sign = shocks are identified by imposing sign restrictions, Single = the VAR is estimated on the sample containing a single monetary policy regime.

2.3.1 Omitted Variables

Commodity Prices According to Sims (1992), researchers observe the price puzzle because central banks are forward-looking and react to the anticipated future movements of inflation by raising the interest rate. When researchers omit information about future inflation in their VAR system, the examined shocks become combinations of true monetary policy shocks and endogenous reactions to expected inflation.[2] If the central bank does not fully accommodate the expected inflation, the data show that an increase in the interest rate, mistakenly recognized as a monetary policy shock, is followed by an increase in the price level. Sims (1992) finds that including commodity prices into the VAR mitigates the price puzzle. Nevertheless, the evidence from the entire literature summarized in Table 2.1 suggests that the inclusion of commodity prices helps little by itself. Almost 50% of VAR models with commodity prices still report the puzzle.

Output Gap Giordani (2004) argues that the use of GDP in the VAR system without controlling for the potential output of the economy can bias the estimates and

[2]Recent contributions to the study of monetary transmission mechanism stress the importance of forward looking nature of monetary policy (Cloyne & Hürtgen 2016; Wolf 2016).

cause the price puzzle. He claims that the inclusion of commodity prices alleviates the puzzle mostly because commodity prices contain useful information about the output gap, not just because they are a good predictor of future inflation. Indeed, Hanson (2004) finds little correlation between the ability to solve the price puzzle and the ability to forecast inflation. Approximately 16% of the studies in our sample use the output gap (or add a time trend), but some of them still find the puzzle.

Factor-augmented VAR To address the major shortcomings of standard small-scale VARs, Bernanke *et al.* (2005) introduce the factor-augmented VAR approach. They argue that policymakers take into account hundreds of variables when deciding about monetary policy. Standard VAR models with three to six variables may therefore suffer from omitted-variable bias. The FAVAR approach, on the other hand, makes use of additional information by extracting principal components from many time series and, as Bernanke *et al.* (2005) argue, should solve the price puzzle. But evidence from the literature (Table 2.1) indicates that FAVAR is ineffective in explaining the puzzle away.

2.3.2 Identification

While some researchers stress the role of omitted variables, others argue that the puzzle arises from implausible identification of monetary policy shocks. The usual recursive identification, which assumes that monetary policy affects output and prices only with a lag, is, for example, not consistent with the New-Keynesian class of theoretical models (Carlstrom *et al.* 2009).

Non-recursive Identification The main idea of a non-recursive identification of shocks, going back to Bernanke (1986) and Blanchard & Watson (1986), is that the matrix contemporaneously linking structural shocks and reduced-form residuals is no longer lower triangular, but that it assumes a general form indicated by theory: the rows of the matrix have a structural interpretation. The restrictions presented by Kim & Roubini (2000), for example, are elicited from the structural stochastic equilibrium model developed by Sims & Zha (1998). Although non-recursive identification is theory-consistent, Table 2.1 suggests that in almost 33% of the responses computed using this strategy the price puzzle still occurs.

Sign Restrictions Canova & Nicolo (2002) and Uhlig (2005) present a novel identification approach that assigns a structural interpretation to orthogonal innovations by imposing sign restrictions on the responses to shocks. The method is attractive since sign restrictions can be derived directly from structural theoretical models. The identifying assumptions are clearly stated and the shocks can be given the structural interpretation without imposing zero restrictions.[3] As Table 2.1 documents, VARs estimated with sign restrictions rarely encounter the price puzzle.

2.3.3 Monetary Policy Regime

Another stream of literature suggests that the price puzzle is historically limited to periods of passive monetary policy or that it emerges when researchers mix data for different monetary policy regimes (Elbourne & de Haan 2006; Borys *et al.* 2009).[4] For example, if a researcher assumes that the central bank uses the interest rate to target inflation, although for some part of the sample monetary or exchange rate targeting was in place, monetary policy shocks in the VAR system become incorrectly identified. Table 2.1 shows that most researchers who evaluate monetary transmission in a period of a single monetary policy regime do not report the price puzzle.

The previous paragraphs illustrate that the quality of studies included in our sample varies. Some of the studies are obviously misspecified. Will not the misspecified studies bias the research synthesis? Indeed, this has been an objection to meta-analysis, and an alternative approach called best-evidence synthesis has been proposed (Slavin 1986). Proponents of best-evidence synthesis argue that we should not include bad studies when we are interested in the average effect. If misspecifications have a systematic influence on the results, then the simple average produced by meta-analysis will be biased. The problem with best-evidence synthesis is the definition of best evidence. For example, should we discard all VAR models that omit commodity prices? In that case we would have 125 observations for the 3-month horizon. But if we additionally threw away all studies that neglect potential output, mix monetary policy regimes, and resort to recursive identification, we would be left with a handful of observations.

The empirical literature on monetary transmission is rich in method choices that the researcher must make. When more and more aspects of methodology become a subject of scrutiny, best-evidence synthesis boils down to selecting the best study from the literature. But this denies the purpose of research synthesis—to provide robust results and explain the differences between the findings of individual studies. Meta-analysis, in contrast, enables us to test explicitly whether misspecifications of primary studies affect the reported results in a systematic way. If so, we can define what we think constitutes best practice and estimate the average impulse response conditional on such best practice without throwing away any information. Because

[3] The way how sign restriction are incorporated might matter a lot: it might be important to distinguish between cases where sign restrictions are imposed on the price level response and cases where the response of price level is not restricted.

[4] Recently, contributions by Coibion (2012) and Barakchian & Crowe (2013) suggest that the fact whether periods of disinflations are included in the sample might be more important than just controling for the monetary policy regime.

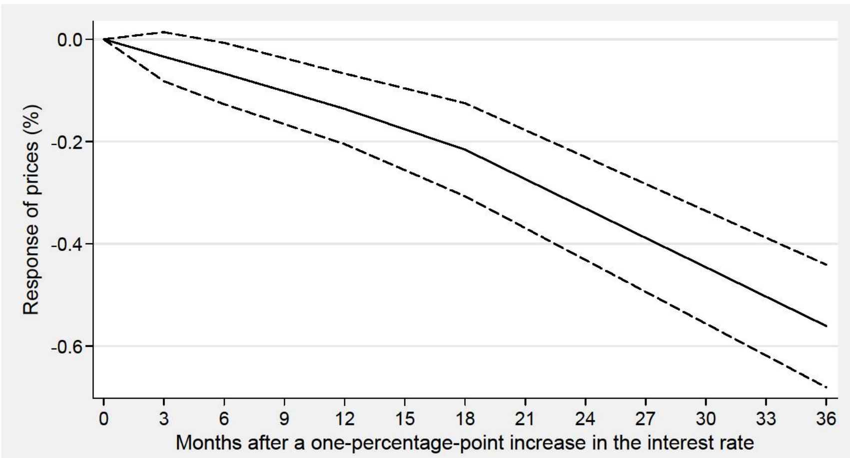
best practice is subjective, we will try several alternative definitions. Moreover, we want to explain what makes researchers report the price puzzle. If misspecifications cause the price puzzle, we need misspecified studies as well.

2.4 Consequences of Publication Selection

After we have collected about 1,000 estimates of the response of prices to monetary tightening, a natural question arises: what general impulse response does the literature suggest? Meta-analysis was originally developed in medicine to combine many small studies into a large one, and therefore to boost the number of degrees of freedom. Clinical trials are costly, and meta-analysis thus became the dominant method of taking stock of medical research. Estimating a VAR model may be less expensive, but the degrees of freedom in macroeconomics are limited. Hence, the original purpose of meta-analysis is useful even here since it combines information from many countries and time periods: when recomputed into quarters the primary studies in our sample taken together use 2,452 unique observations.

Taking a simple mean of all point estimates for each of the five horizons implies the impulse-response function depicted in Figure 2.2. This average impulse response shows a relatively intuitive short-run reaction of prices to a one-percentage-point increase in the interest rate: prices decline already in the short run, the decrease becomes significant in the medium run and reaches 0.56% after 36 months. Nevertheless, the response shows no sign of bottoming out.

Figure 2.2: Average impulse response implied by the literature



Note: Confidence bands are constructed as +/- one standard error.

Simply averaging the collected impulse responses has two major shortcomings. First, it ignores possible publication selection. If some results are more likely to get published than others, the average becomes a biased estimator of the underlying impulse response. Second, it ignores heterogeneity in the results of the primary studies. Since different researchers use different data and methods, and the studies are of different quality, it is unrealistic to assume that all estimates are drawn from the same population. In addition, as discussed in Section 2.3, some VAR models are misspecified, and if misspecifications have a systematic influence on the results, it is possible to improve upon the average response by filtering out the misspecifications. We address publication selection in this section and heterogeneity and misspecification issues in Section 2.5.

Stanley (2008), among others, points out that publication selection is of major concern for empirical research in economics. When there is little theory competition for what sign the underlying effect should have, estimates inconsistent with the predominant theory will be treated with suspicion or even be discarded. An illustrative example can be found in the literature on the effect of a common currency on trade (Rose & Stanley 2005): it is hard to defend negative estimates of the trade effect of currency unions. The negative estimates most likely result from misspecification, and researchers may be correct in not stressing them. On the other hand, it is far more difficult to identify excessively large estimates of the same effect that also arise from misspecifications. No specific threshold exists above which the estimate would become suspicious. If researchers include the large positive estimates but omit the negative ones, the inference will be on average biased toward a stronger effect.

A similar selection, perhaps of lower intensity, may be taking place in the VAR literature on monetary transmission as well (Uhlig 2010, p. 17, provides anecdotal evidence).[5] Some researchers treat the price puzzle as a clear indication of a misspecification error and try to find an intuitive impulse response for interpretation. Statistical significance is also important. Significant impulse responses are more convenient for interpretation, and especially researchers in central banks may be interested in reporting a well-functioning monetary transmission with a significant reaction of prices to a change in monetary policy. The selection for significance does not distort the average estimate from the literature if the true underlying effect equals zero, but otherwise it creates a bias, again in favor of a stronger effect, since estimates with the wrong sign are less likely to be significant.

A common way to detect publication selection is an informal examination of a so-called funnel plot (Stanley & Doucouliagos 2010). The funnel plot depicts the esti-

[5]Uhlig (2010, p. 17) writes: "At a Carnegie-Rochester conference a few years back, Ben Bernanke presented an empirical paper, in which the conclusions nicely lined up with a priori reasoning about monetary policy. Christopher Sims then asked him, whether he would have presented the results, had they turned out to be at odds instead. His half-joking reply was, that he presumably would not have been invited if that had been so. There indeed is the danger (or is it a valuable principle?) that a priori economic theoretical biases filter the empirical evidence that can be brought to the table in the first place."

mates on the horizontal axis against their precision (the inverse of the standard error) on the vertical axis. If there is no heterogeneity or misspecification, the estimates with the highest precision will be close to the true underlying effect. In the absence of publication selection the funnel is symmetrical: the reported estimates are dispersed randomly around the true effect. The asymmetry of the funnel plot suggests publication bias; for example, if estimates with a positive sign are less likely to be selected for publication, estimates on the right side of the funnel will be underrepresented.

The funnel plots for all five horizons are depicted in Figure 2.3. The plots resemble funnels commonly reported in economic meta-analyses, which indicates that the employed approximation of standard errors is plausible. As expected, the left part of all funnels is clearly heavier, suggesting publication selection against the price puzzle and in favor of the more negative (that is, stronger) effects of monetary tightening on prices. Nevertheless, the interpretation of funnel plots is subjective, and we need a more formal test of publication bias.

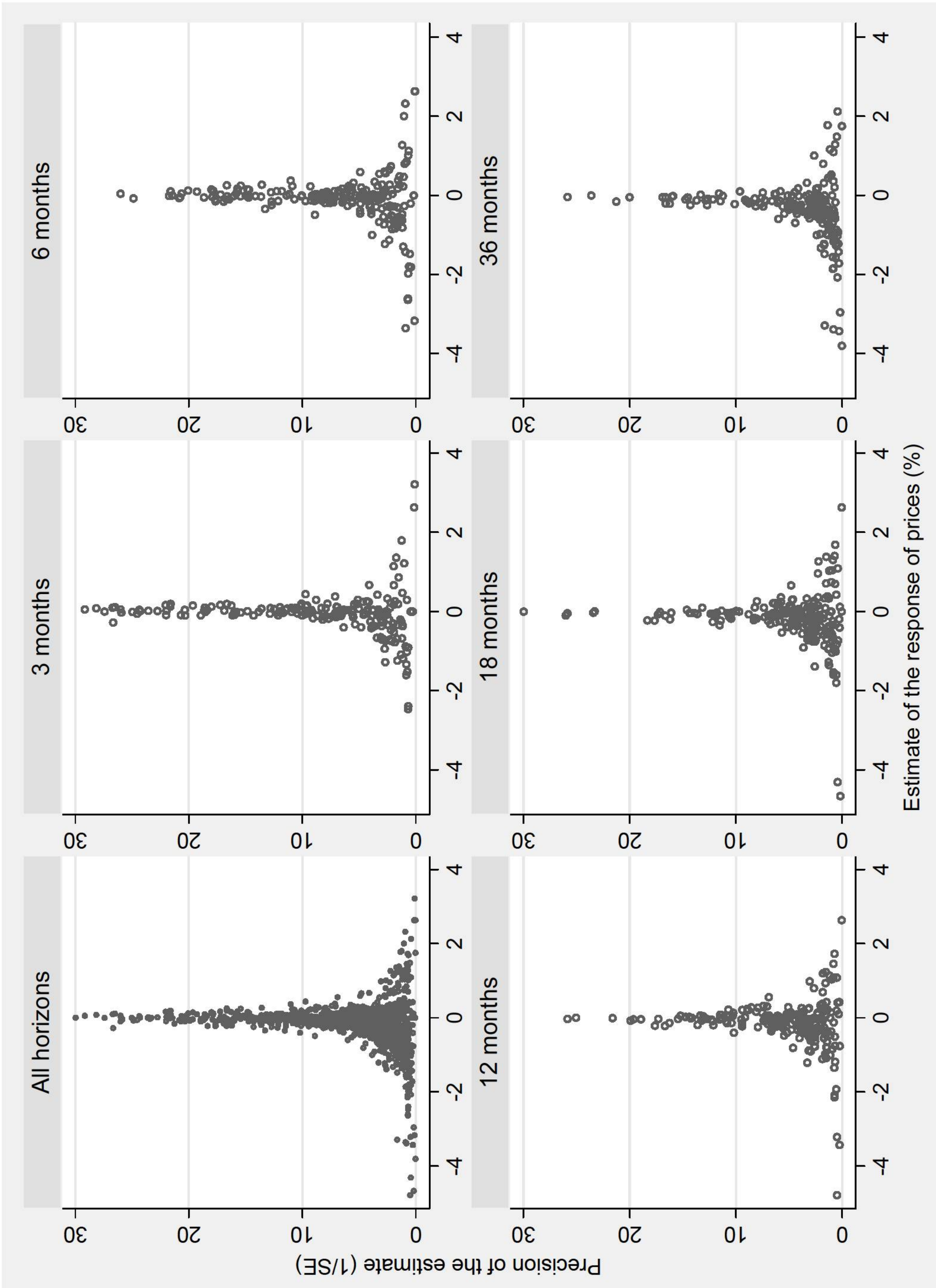
Given small samples, authors wishing to obtain significant results may be tempted to try different specifications until they find estimates large enough to offset the standard errors. In contrast, with large samples even tiny estimates might be statistically significant, and authors therefore have fewer incentives to conduct a specification search. If publication selection is present, we should observe a relationship between an estimate and its standard error (or the square root of the number of observations). The following regression formalizes the idea (Card & Krueger 1995):

$$\hat{\beta}_j = \beta + \beta_0 SE_j + e_j, \tag{2.3}$$

where β denotes the true underlying effect, $\hat{\beta}_j$ denotes the effect's j-th estimate, β_0 denotes the magnitude of publication bias, SE_j denotes the standard error of $\hat{\beta}_j$, and e_j denotes a disturbance term.

Specification (2.3) has become the cornerstone of modern meta-analysis in medicine and the social sciences, including economics. The question is whether the method is suitable for summarizing graphical results such as impulse responses. In order for this meta-analysis method to be valid, the distribution of empirical effects needs to be symmetrical *in the absence of publication bias* [usually it is assumed that the disturbance term in (2.3) is normally distributed]. But impulse responses are nonlinear functions of the coefficients estimated in the VAR system; as discussed in Section 3.2, the confidence intervals around the individual estimates are often asymmetrical. If the pattern of asymmetry is not random across the individual estimates, the distribution of the impulse responses will not be symmetrical even in the absence of publication bias, and the test for publication bias will be invalid.

Figure 2.3: Funnel plots show publication selection against the price puzzle



Systematic asymmetry of the distribution of impulse responses would manifest as a significant difference between the average distance from the point estimate of the impulse response to the lower and upper confidence bound. We select the 68% confidence bound (34% on both sides of the estimate), which for a symmetrical distribution would imply a distance of one standard error on both sides of the mean. The difference of the distances is significant at the 5% level for only one out of five horizons (the 12-month horizon), and even there the difference is small: the average lower confidence interval is 11.6% farther from the mean than is the average upper confidence interval. It is unlikely that such a small difference could explain the degree of asymmetry apparent from Figure 2.3. It cannot explain the asymmetry of the collected point estimates of the impulse responses at the 12-month horizon, where the distance from the 16th percentile to the mean is 53.1% larger than the distance from the mean to the 84th percentile. For this reason, we employ the standard meta-analysis methodology—bearing in mind that the results concerning publication bias must be interpreted with some caution.[6]

In practice, meta-analysts rarely estimate specification (2.3) directly since it suffers from heteroscedasticity by definition (the explanatory variable is a sample estimate of the standard deviation of the response variable). Instead, weighted least squares are used to gain efficiency, and they require that specification (2.3) be divided by SE_j, the measure of heteroscedasticity (Stanley 2008):

$$\frac{\hat{\beta}_j}{SE_j} \equiv t_j = \beta_0 + \beta \left(\frac{1}{SE_j} \right) + \xi_j, \quad \xi_j | SE_j \sim N(0, \sigma^2), \tag{2.4}$$

where t_j denotes the approximated t-statistic of the estimate and the new disturbance term ξ_j has constant variance. Note that the intercept and the slope are now reversed: the slope measures the true effect and the intercept measures publication bias. In addition to removing heteroscedasticity, specification (2.4) gives more weight to more precise results, which represents a common approach in meta-analysis. Testing the significance of β_0 in this specification is analogous to testing the asymmetry of the funnel plot—it follows from rotating the funnel plot and dividing the values on the new vertical axis by SE_j. Testing the significance of β constitutes a test for the true underlying effect of monetary tightening on prices, corrected for publication selection.

The intercept of specification (2.4), which in our case measures the degree of publication bias, has an alternative interpretation that is sometimes used in economics meta-analyses. Since the response variable is the t-statistic, the intercept represents the average t-statistic that the literature reports for the effect in question. The av-

[6] Additionally, the asymmetry of funnel plots may partly reflect small-sample bias in the estimated VAR coefficients. A similar limitation was found in a meta-analysis of unemployment hysteresis (Stanley 2004).

erage is, however, conditional on precision (that is, the inverse of standard error). If precision was not included in specification (2.4), such as, for example, in the meta-analysis by Görg & Strobl (2001), the intercept would represent the unconditional average t-statistic. In that case, however, publication bias could not be separated from the true effect.

Our specification controls for precision, which means that the intercept corresponds to the average t-statistic conditional on precision being close to zero (or, alternatively, on the standard error of the estimated coefficient being close to infinity). The true effect has no relation to the observed t-statistic as precision goes to zero; in other words, the precision term in (2.4) filters out any underlying effect. When precision is zero, the average t-statistic should be zero as well. If it is not, something is wrong with the literature, and we observe signs of publication bias (or any other bias that causes the asymmetry of funnel plots). A more detailed treatment of this problem is available in Stanley (2008).

Since we use all reported impulse responses we need to account for the potential dependence of estimates within one study (Disdier & Head 2008); in such a case, (2.4) would be misspecified. As a remedy, researchers typically employ the mixed-effects multilevel model (Doucouliagos & Stanley 2009; Havranek & Irsova 2011):

$$t_{ij} = \beta_0 + \beta \left(\frac{1}{SE_{ij}} \right) + \alpha_j + \epsilon_{ij}, \quad \alpha_j | SE_{ij} \sim N(0, \psi), \quad \epsilon_{ij} | SE_{ij}, \alpha_j \sim N(0, \theta), \tag{2.5}$$

where i and j denote estimate and study subscripts, respectively. The overall error term now consists of study-level random effects and estimate-level disturbances ($\xi_{ij} = \alpha_j + \epsilon_{ij}$), and its variance is additive since both components are assumed to be independent: $\mathrm{Var}(\xi_{ij}) = \psi + \theta$, where ψ denotes between-study variance and θ within-study variance. If ψ approaches zero the benefit of using the mixed-effect estimator instead of ordinary least squares (OLS) dwindles. To put the magnitude of these variance terms into perspective the within-study correlation is useful: $\rho \equiv \mathrm{Cor}(\xi_{ij}, \xi_{i'j}) = \psi/(\psi + \theta)$, which expresses the degree of dependence of estimates reported in the same study, or equivalently, the degree of between-study heterogeneity.

The mixed-effects multilevel model is analogous to the random-effects model commonly used in panel-data econometrics. We follow the terminology from multilevel data modeling, which calls the model "mixed effects" since it contains a fixed (β) as well as a random (α_j) part. For the purposes of meta-analysis the multilevel framework is more suitable because it takes into account the unbalancedness of the data (the restricted maximum likelihood estimator is used instead of generalized least squares), allows for nesting multiple random effects (study-, author-, or country-level), and is thus more flexible (Nelson & Kennedy 2009).

Table 2.2: Test of true effect and publication bias

	Mixed-effects multilevel				
Horizon	3 months	6 months	12 months	18 months	36 months
Intercept (bias)	0.058	-0.088	-0.176	-0.325**	-0.806***
	(0.167)	(0.166)	(0.145)	(0.128)	(0.126)
$1/SE$ (effect)	0.009	0.007	-0.014	-0.019	-0.009
	(0.009)	(0.011)	(0.014)	(0.012)	(0.010)
Within-study correlation	0.43	0.56	0.46	0.41	0.14
Observations	208	215	215	217	205
Studies	69	70	70	70	63

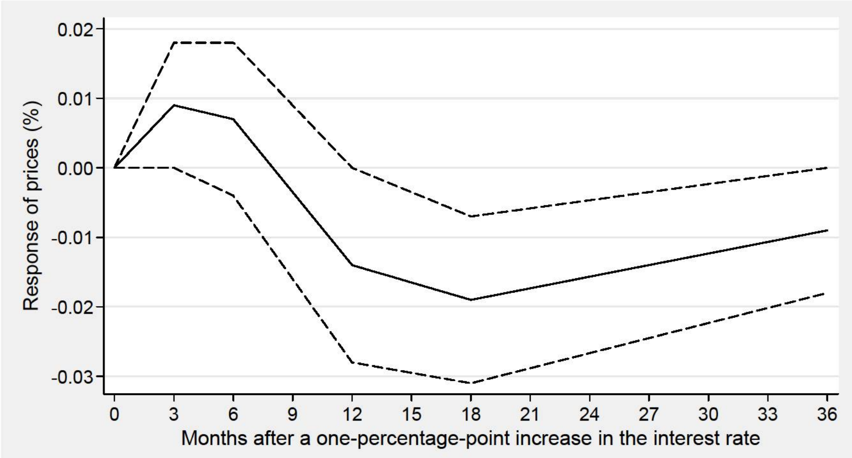
Note: Standard errors in parentheses. Response variable: the approximated t-statistic of the estimate of the percentage response of prices to a one-percentage-point increase in the interest rate.
***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The outcomes of the mixed-effects estimator are presented in Table 2.2. OLS with standard errors clustered at the study level are reported in the Appendix: Table 2.6 gives even more significant results for publication bias. The within-study correlation is large, indicating that the mixed-effects estimator is more appropriate, which is confirmed by likelihood-ratio tests.[7] Compared with the simple average, the response of prices corrected for publication bias is more positive (that is, weaker), corroborating evidence for publication selection in favor of the stronger responses of prices to monetary policy contraction. Moreover, the magnitude of publication bias increases with the time horizon after the shock. This result is in line with Doucouliagos & Stanley (2013), who find stronger publication selection for research questions with weaker theory competition. For the short run, some disagreement occurs regarding the effects of monetary policy on prices because of the cost channel. (Since firms depend on credit to finance production, their costs rise when the central bank increases the interest rate, and they may increase prices.) On the other hand, a consensus emerges about the long-run effect: prices should eventually decrease after monetary policy tightening; estimates showing the opposite would be difficult to publish.

The impulse-response function corrected for publication bias is depicted in Figure 2.4: it exhibits the price puzzle. In the short run prices increase, but in the medium run they decrease and bottom out 18 months after the tightening. The

[7] We experimented with several nested mixed-effects models, but they yield qualitatively similar outcomes. Additionally, we collected data from unpublished manuscripts appearing in the working-paper series of NBER, OECD, IMF, European Commission, and all central banks listed in the Bank for International Settlements Central Bank Research Hub, and ran regression (2.5) using this new sample. The working papers show a pattern of publication bias very similar to that presented in Table 2.2. These robustness checks are available in the online appendix.

Figure 2.4: Impulse response corrected for publication bias exhibits the puzzle



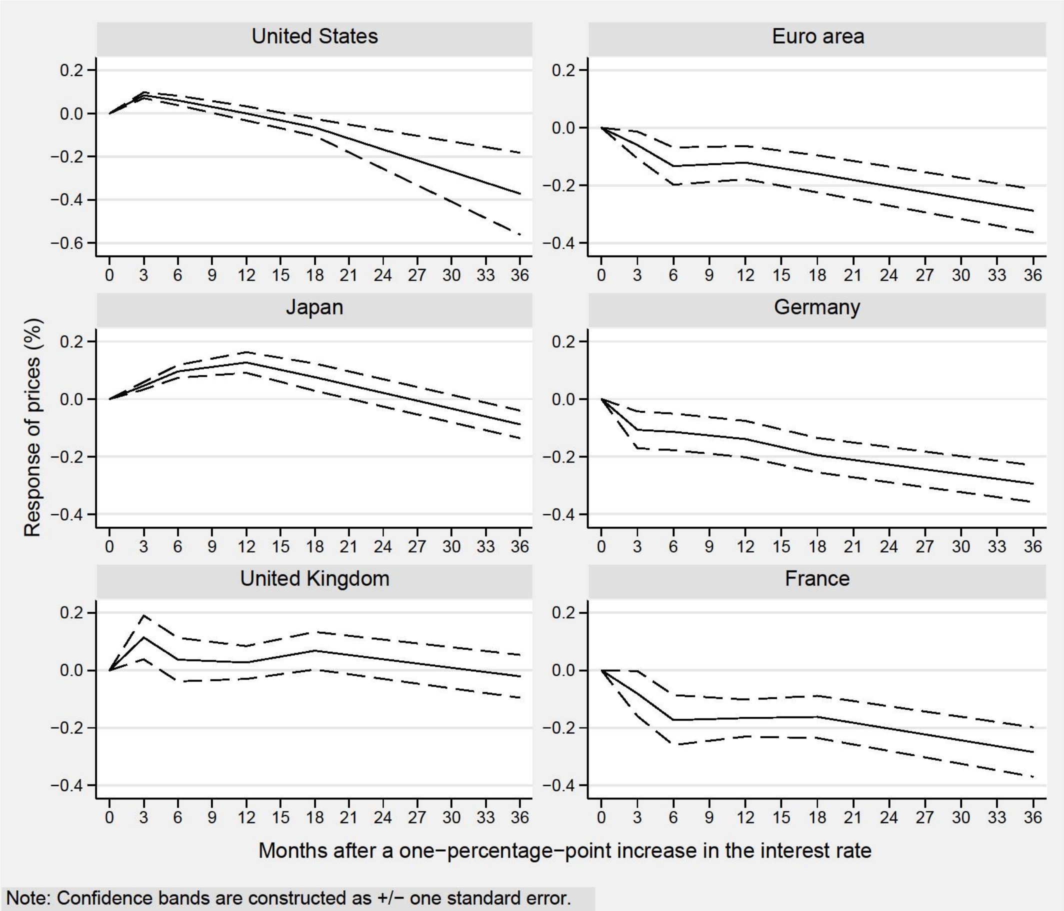
Note: Confidence bands are constructed as +/- one standard error.

maximum decrease in the price level, however, is negligible: only 0.02%. Compared to the average response reported in Figure 2.2, now the function shifts upwards—especially in the long run, because publication bias is filtered out. Figure 2.4 would be our best estimate of the underlying impulse response if all heterogeneity between studies was random; the estimate is unconditional on the characteristics of the countries examined and on the methodology used. In the next section we relax the assumption of random heterogeneity and explain the differences in the reported estimates. In particular, we are interested in the average impulse response conditional on best-practice methodology.

2.5 What Explains Heterogeneity

As motivation for the empirical investigation of structural heterogeneity consider Figure 2.5, which depicts the differences in monetary transmission among selected countries. We use a simple random-effects meta-analysis to compute impulse-response functions. Simple meta-analysis weights each estimate by its precision and adds an estimate-specific random effect; it does not correct for publication bias. We use simple meta-analysis for estimation by countries since it requires fewer degrees of freedom than meta-regression. Figure 2.5 shows that the impulse responses for the United States, the United Kingdom, and Japan exhibit the price puzzle, but that monetary transmission in euro area countries seems to work intuitively and prices decline soon after a tightening. Nevertheless, a part of these differences may arise from the use of diverse methods since some countries are examined only in a few studies.

Figure 2.5: Aggregate impulse responses for selected countries suggest heterogeneity



To account for heterogeneity we extend the meta-regression (2.5) to the following multivariate version:

$$t_{ij} = \beta_0 + \frac{\beta}{SE_{ij}} + \sum_{k=1}^{K} \frac{\gamma_k Z_{ijk}}{SE_{ij}} + \alpha_j + \epsilon_{ij}, \tag{2.6}$$

where Z denotes explanatory variables assumed to affect the reported estimates. The exogeneity assumptions become $\alpha_j | SE_{ij}, Z_{ijk} \sim N(0, \psi)$ and $\epsilon_{ij} | SE_{ij}, \alpha_j, Z_{ijk} \sim N(0, \theta)$.

Table 2.3 presents descriptions and summary statistics of all the explanatory variables we consider. In principle, they can be divided into five groups: variables capturing the fundamental characteristics of the economy (structural heterogeneity), data characteristics controlling for differences in the data used, specification characteristics controlling for differences in the basic design of the estimated models, estimation characteristics controlling for differences in econometric techniques, and publication characteristics controlling mainly for differences in quality not captured by other variables.

Table 2.3: Description and summary statistics of regression variables

Variable	Description	Mean	Std. dev.
Response (3M)	The percentage response of prices 3 months after a tightening.	-0.034	0.692
Response (6M)	The percentage response of prices 6 months after a tightening.	-0.067	0.883
Response (12M)	The percentage response of prices 12 months after a tightening.	-0.136	1.012
Response (18M)	The percentage response of prices 18 months after a tightening.	-0.216	1.327
Response (36M)	The percentage response of prices 36 months after a tightening.	-0.561	1.714
$1/SE$	The precision of the estimate of the response (all horizons).	6.805	7.821
Structural heterogeneity			
GDP per capita	The logarithm of the country's real GDP per capita.	9.881	0.414
GDP growth	The average growth rate of the country's real GDP.	2.668	1.035
Inflation	The average inflation of the country.	7.748	14.26
Inflation volatility	The standard deviation of the difference between the country's inflation and its Hodrick-Prescott-filtered inflation trend.	6.234	33.43
Financial development	The financial development of the country measured by (domestic credit to private sector)/GDP.	0.837	0.414
Openness	The trade openness of the country measured by (exports + imports)/GDP.	0.460	0.401
CB independence	A measure of central bank independence (Arnone *et al.* 2009).	0.774	0.143
Data characteristics			
Monthly	=1 if monthly data are used.	0.630	0.483
Time span	The number of years of the data used in the estimation.	18.83	10.44
No. of observations	The logarithm of the number of observations used.	4.889	0.675
Average year	The average year of the data used (2000 as a base).	-8.926	7.881
Specification characteristics			
GDP deflator	=1 if the GDP deflator is used instead of the consumer price index as a measure of prices.	0.177	0.382
Single regime	=1 if the VAR is estimated over a period of a single monetary policy regime.	0.296	0.457
No. of lags	The number of lags in the model, normalized by frequency: lags/frequency	0.610	0.370

Continued on the next page

Table 2.3: Description and summary statistics of regression variables (continued)

Variable	Description	Mean	Std. dev.
Commodity prices	=1 if a commodity price index is included.	0.607	0.489
Money	=1 if a monetary aggregate is included.	0.529	0.499
Foreign variables	=1 if at least one foreign variable is included.	0.441	0.497
Time trend	=1 if a time trend is included.	0.126	0.332
Seasonal	=1 if seasonal dummies are included.	0.146	0.354
No. of variables	The logarithm of the number of endogenous variables included in the VAR.	1.741	0.383
Industrial production	=1 if industrial production is used as a measure of economic activity.	0.430	0.495
Output gap	=1 if the output gap is used as a measure of economic activity.	0.028	0.165
Other measures	=1 if another measure of economic activity is used (employment, expenditures).	0.119	0.324
Estimation characteristics			
BVAR	=1 if a Bayesian VAR is estimated.	0.144	0.352
FAVAR	=1 if a factor-augmented VAR is estimated.	0.051	0.221
SVAR	=1 if non-recursive identification is employed.	0.295	0.456
Sign restrictions	=1 if sign restrictions are employed.	0.144	0.352
Publication characteristics			
Study citations	The logarithm of [(Google Scholar citations of the study)/(age of the study) + 1].	1.882	1.279
Impact	The recursive RePEc impact factor of the outlet.	0.888	2.274
Central banker	=1 if at least one co-author is affiliated with a central bank.	0.451	0.498
Policymaker	=1 if at least one co-author is affiliated with a Ministry of Finance, IMF, OECD, or BIS.	0.055	0.228
Native	=1 if at least one co-author is native to the investigated country.	0.446	0.497
Publication year	The year of publication (2000 as a base).	5.032	3.886

Structural heterogeneity When constructing the variables that capture structural heterogeneity, we use the average values which correspond with the sample employed in the estimation of the impulse response. For instance, in the case of inflation: When the impulse response comes from a VAR model estimated on the 1990:1–1999:12 Italian data, we use the average inflation rate in Italy for the period 1990–1999. This approach increases the variability in regressors and describes the estimates more precisely than using the same year of structural variables for all extracted impulse responses. The variable GDP per capita reflects the importance of the degree of economic development of the economy for monetary transmission. To investigate whether the strength of transmission depends on the phase of the economic cycle, we include the variable GDP growth in the meta-regression. The underlying reason is related to credit market imperfections, which could amplify the propagation of monetary policy shocks during bust periods (Bernanke & Gertler 1989).

Next, we examine the variables implied by the various channels of the transmission mechanism. We include the trade openness of the economy to capture the importance of foreign developments for domestic monetary policy as well as the exchange rate channel of monetary transmission. Furthermore, as pointed out by Bernanke & Gertler (1995) and Cecchetti (1999), differences in financial structure may explain important portions of heterogeneity in monetary transmission. We include a measure of financial development approximated by the ratio of private credit to GDP.

We add the average level and volatility of inflation, as these may influence price setting behavior as well as monetary transmission (Angeloni *et al.* 2006). We expect that independent central banks are likely to be more credible (Rogoff 1985; Keefer & Stasavage 2003; Perino 2010). In consequence, economic subjects may respond more to monetary policy shocks. We test whether the degree of central bank independence affects the strength of monetary transmission.

Regarding the sources of the data, the trade openness, GDP growth, and GDP level per capita are obtained from Penn World Tables. The consumer price index, used to compute average inflation and inflation volatility, is obtained from the International Monetary Fund's International Financial Statistics. The ratio of domestic credit to GDP is obtained from the World Bank's World Development Indicators, and the index of central bank independence is extracted from Arnone *et al.* (2009).

Data characteristics We control for the frequency of the data used in the VAR model: 63% of specifications use monthly data, the rest rely on quarterly data. To account for possible changes in transmission not explained by the structural variables (for example, changes caused by globalization or financial innovations, see Boivin & Giannoni 2006), we include the average year of the sample period used in the estimation. Finally, we add the total number of observations to assess whether smaller samples yield systematically different outcomes.

Specification characteristics To account for the different measures of the price level we include a dummy which equals one when the GDP deflator is used instead of the usual consumer price index (18% of specifications in primary studies). We add a dummy for the case where the data cover a period of a single monetary policy regime (30%). Next, we include the VAR's lag order normalized by the data frequency. We account for the cases where commodity prices, a money aggregate, foreign variables, a time trend, and seasonal dummies are included in the VAR. We also control for the number of endogenous variables in the model. Since the results might vary depending on the measure of economic activity, we introduce dummies for the cases where industrial production, the output gap, or another measure is used instead of GDP.

Estimation characteristics Most of the studies in our sample estimate VAR models using the standard methods (OLS or Maximum Likelihood); we control for studies using Bayesian methods to address the problem of overparameterization (14% of specifications in primary studies) and for studies using the FAVAR approach to address the problem of omitted variables (5%). As for identification strategies, most of the studies employ recursive identification; we include a dummy for non-recursive identification (30%) and a dummy for identification using sign restrictions (14%).

Publication characteristics To proxy study quality we use the recursive RePEc impact factor of the outlet (because the journal coverage of RePEc is much more comprehensive than in other databases) and the number of Google Scholar citations of the study normalized by the study's age. We add a dummy for authors affiliated with a central bank and a dummy for authors working at policy-oriented institutions such as a Ministry of Finance, the International Monetary Fund, or the Bank for International Settlements. We include a dummy for the case where at least one co-author is "native" to the examined country: such authors may be more familiar with the data at hand, which could contribute positively to the quality of the analysis; on the other hand, such authors may have a vested interest in the results. We consider authors native if they either were born in the country or obtained an academic degree there. Finally, we use the year of publication to account for possible improvements in methodology that are otherwise difficult to codify.

In the first step we estimate a general model containing all explanatory variables; the general model is not reported but is available in the online appendix. All variance inflation factors are lower than 10, indicating that the degree of multicollinearity is not too problematic. In the second step, we drop the variables which are for each horizon jointly insignificant at the 10% level.

For example, GDP per capita, the number of lags used, and most publication characteristics belong to the dropped variables. The insignificance of publication characteristics suggests that the quality of a given study is to a large extent captured by the methods used.

The resulting model is presented in Table 2.4. The specifications reported in this section are based on the mixed-effects multilevel estimator, but the inference would be similar from an OLS with standard errors clustered at the study level; these robustness checks are available in Appendix A. The similarity between the outcomes of these two estimators indicates that the exogeneity assumptions made in the mixed-effects estimation are not seriously violated; in meta-analysis it is difficult to test exogeneity formally because the extreme unbalancedness of the data (some studies report only one impulse response) does not permit the construction of a reasonable fixed-effects model. We prefer mixed effects over OLS because likelihood-ratio tests reject the hypothesis of zero within-study variance, suggesting that the OLS is misspecified.

Table 2.4: Explaining the differences in reported impulse responses

	Mixed-effects multilevel				
Horizon	3 months	6 months	12 months	18 months	36 months
Intercept (publication bias)	-0.112	-0.134	-0.219*	-0.208*	-0.604***
	(0.131)	(0.133)	(0.132)	(0.124)	(0.150)
1/SE	-0.075	-0.125	-0.287	-0.252	-0.154
	(0.117)	(0.147)	(0.181)	(0.169)	(0.202)
Structural heterogeneity					
GDP growth	-0.006	0.009	0.023**	0.023**	0.040***
	(0.008)	(0.010)	(0.011)	(0.011)	(0.012)
Inflation	0.001	-0.001	0.003	0.004	0.009***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Inflation volatility	-0.0004	0.0004	-0.0011	-0.0019	-0.0044***
	(0.0011)	(0.0014)	(0.0014)	(0.0012)	(0.0013)
Financial development	0.101***	0.080*	0.144**	0.072	-0.024
	(0.036)	(0.048)	(0.064)	(0.062)	(0.070)
Openness	-0.028	-0.048	-0.068	-0.090*	-0.283***
	(0.039)	(0.049)	(0.056)	(0.048)	(0.042)
CB independence	0.088	-0.015	-0.040	-0.167*	-0.290***
	(0.070)	(0.089)	(0.097)	(0.085)	(0.079)
Data characteristics					
No. of observations	0.011	0.027	0.049*	0.080***	0.148***
	(0.017)	(0.023)	(0.028)	(0.028)	(0.032)
Average year	0.002	-0.001	0.002	0.005*	0.013***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Specification characteristics					
GDP deflator	0.011	0.039	0.126***	0.157***	0.148
	(0.023)	(0.030)	(0.043)	(0.051)	(0.092)
Single regime	0.028	0.033	0.031	0.026	0.095**
	(0.020)	(0.025)	(0.033)	(0.035)	(0.037)
Commodity prices	-0.045***	-0.066***	-0.127***	-0.151***	-0.226***
	(0.016)	(0.021)	(0.030)	(0.031)	(0.033)
Foreign variables	0.011	0.032	0.062**	0.065*	0.130***

Continued on the next page

Table 2.4: Explaining the differences in reported impulse responses (continued)

	Mixed-effects multilevel				
Horizon	3 months	6 months	12 months	18 months	36 months
	(0.017)	(0.023)	(0.031)	(0.034)	(0.045)
No. of variables	-0.018	-0.024	-0.034	-0.056**	-0.183***
	(0.014)	(0.015)	(0.022)	(0.025)	(0.049)
Industrial production	0.030	0.060**	0.061*	0.064*	-0.011
	(0.023)	(0.027)	(0.035)	(0.038)	(0.039)
Output gap	-0.249	-0.303**	-0.219***	-0.131*	0.015
	(0.162)	(0.136)	(0.084)	(0.070)	(0.036)
Other measures	-0.072**	-0.036	-0.059	-0.041	-0.026
	(0.029)	(0.037)	(0.054)	(0.063)	(0.093)
Estimation characteristics					
BVAR	0.113***	0.085**	0.112**	0.160**	0.153
	(0.033)	(0.036)	(0.055)	(0.070)	(0.132)
FAVAR	-0.135***	-0.182***	-0.105	0.035	0.299**
	(0.036)	(0.059)	(0.082)	(0.085)	(0.122)
SVAR	-0.068***	-0.109***	-0.123***	-0.139***	-0.070***
	(0.016)	(0.018)	(0.023)	(0.022)	(0.026)
Sign restrictions	-0.294***	-0.280***	-0.334***	-0.369***	-0.271*
	(0.036)	(0.051)	(0.069)	(0.083)	(0.141)
Publication characteristics					
Central banker	0.034	0.052*	0.074**	0.076**	0.133***
	(0.022)	(0.027)	(0.033)	(0.035)	(0.038)
Policymaker	-0.057*	-0.029	0.051	0.092**	0.174***
	(0.034)	(0.043)	(0.040)	(0.038)	(0.045)
Within-study correlation	0.32	0.37	0.32	0.37	0.43
Observations	208	215	215	217	205
Studies	69	70	70	70	63

Note: Standard errors in parentheses. Response variable: the approximated t-statistic of the estimate of the percentage response of prices to a one-percentage-point increase in the interest rate.
***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Concerning structural heterogeneity, the results reported in Table 2.4 suggest that GDP growth, the openness of the economy, the level and volatility of inflation, and the degree of central bank independence systematically affect the estimated impulse response of prices to monetary tightening in the medium to long run. The importance of monetary policy shocks weakens in periods of higher GDP growth. This result is consistent with Bernanke & Gertler (1989), who argue that asymmetric information and other credit market frictions could amplify the effects of monetary policy through the so-called financial accelerator. In periods of lower GDP growth and especially during recessions, firms' dependence on external financing increases, and changes in the interest rate become more important.

The expectation channel of monetary transmission can explain why the impact of monetary policy diminishes in periods of higher inflation: high inflation impedes the credibility of the central bank and restricts its ability to control the price level. Furthermore, our results indicate that a higher volatility of inflation strengthens the effect on prices in the long run. This is likely to be a consequence of monetary policy shocks having more lasting effects in more volatile environments (Mohanty & Turner 2008). Next, monetary policy is more effective in open economies, where its impact can be amplified through the exchange rate channel. Following a contractionary monetary policy shock, the real exchange rate appreciates through the uncovered interest parity condition. As a result, imported goods become less expensive, amplifying the drop in the aggregate price level caused by monetary tightening (Dennis *et al.* 2007). As expected, monetary policy is more powerful if the central bank enjoys more independence, which corresponds with the findings of Rogoff (1985) and Perino (2010).

In contrast, the structural variables (that is, those related to fundamentals) are not so effective in explaining the short-run response of prices, with the exception of the financial development indicator. Our results suggest that a more developed financial system weakens the short-run impact of monetary policy. This finding complies with Cecchetti (1999), who reports that the effects of monetary policy are more important in countries with many small banks, less healthy banking systems, and underdeveloped capital markets.

Concerning data characteristics, the results presented in Table 2.4 indicate that the number of observations systematically influences the estimated long-run effect: more data make the reported response of prices weaker. In line with Boivin & Giannoni (2006), who argue that globalization coupled with financial innovations may dampen the effects of monetary policy shocks on the economy, the reported long-run response weakens when newer data are used. We find specification characteristics to be important as well. The GDP deflator reacts less to monetary tightening than does the consumer price index. The inclusion of commodity prices is important for all horizons and amplifies the estimated decrease in prices. When industrial production is used instead of GDP as a measure of economic activity, the reported response is weaker; on the other hand, the reported response strengthens when the output gap is used.

Estimation methods matter especially for the short-run response. For the 3- and 6-month horizons, Bayesian estimation produces a smaller decrease in prices compared with a simple VAR. The use of FAVAR, non-recursive identification, and sign restrictions contributes to reporting more potent monetary policy. It is worth noting that all methodological explanations of the price puzzle that were discussed in Section 2.3 indeed contribute to alleviating the puzzle and therefore to estimating intuitive impulse responses (with the exception of the effect of a single regime of monetary policy, which has the opposite sign, but is statistically insignificant).

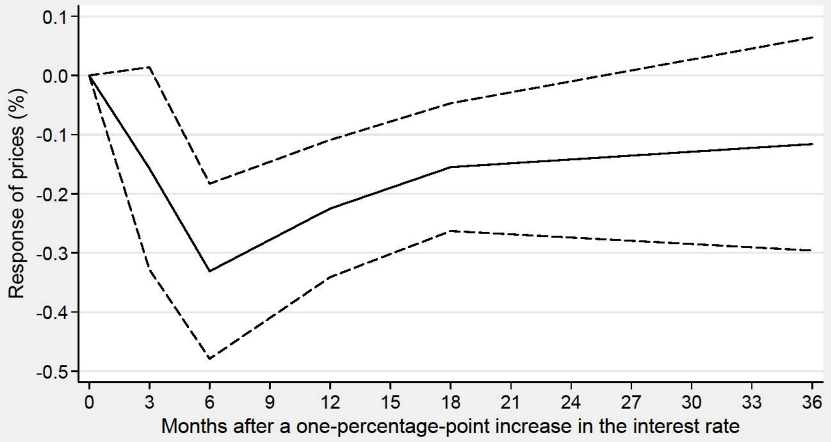
Our results suggest that authors affiliated with central banks report less powerful monetary policy (that is, are more likely to report the price puzzle). This seems counterintuitive since we may expect that central bankers have a vested interest in presenting a well-functioning monetary transmission mechanism. On the other hand, central bank employees may engage less in publication selection—they produce papers needed by their employers and often submit them to academic journals only as a by-product.

The multivariate meta-regression corroborates the evidence for publication selection reported in Section 2.4. The intercept, a measure of publication bias, is statistically significant for the 12-, 18-, and 36-month horizons. The estimate of the true effect in the multivariate model, however, is not simply represented by the regression coefficient for $1/SE$, but is conditional on the variables capturing heterogeneity. In order to estimate the true effect we need to choose the preferred values of the explanatory variables, thus defining some sort of best practice; in this way we create a synthetic study with ideal parameters. A suitably defined best-practice estimation can filter out misspecification bias from the literature, although the approach is subjective since different researchers may have different opinions on what constitutes best practice.

We define best practice by selecting methodology characteristics based on the discussion in Section 2.3: we prefer the output gap over GDP as a measure of economic activity, non-recursive identification over Cholesky decomposition, data covering a single monetary policy regime over mixing more regimes, and the inclusion of commodity prices and foreign variables instead of omitting them. In addition, we prefer Bayesian estimation since overparameterization can be a problem even for systems of modest size (Banbura *et al.* 2010). We insert sample maximums for the number of observations, the year of the data, and the number of endogenous variables. Country-specific variables and dummy variables for central bankers and policymakers are set to their sample means. Similarly to the previous section, the estimate of the impulse response is corrected for funnel plot asymmetry (that is, for publication bias or any other bias contributing to the asymmetry, such as small-sample bias).

The estimated impulse response implied by best practice is depicted in the bottom part of Figure 2.7: after controlling for both publication and misspecification biases, the price puzzle is not present and prices bottom out six months after a one-percentage-point increase in the interest rate. The maximum decrease in the price

Figure 2.6: Impulse response implied by best practice: no price puzzle



Note: Confidence bands are constructed as +/- one standard error.

level reaches 0.33% and is statistically significant at the 5% level. The transmission of monetary policy shocks is quick, which contrasts with the view held at many central banks that there are long lags in the effects of monetary policy on prices (for instance, Bank of England 1999; European Central Bank 2010). The absence of the price puzzle is robust both individually and cumulatively to other possible definitions of best practice: selecting the FAVAR approach instead of the Bayesian approach, selecting the specification using sign restrictions instead of non-recursive identification, or selecting the sample mean of the number of endogenous variables in the VAR system instead of the sample maximum. The price puzzle does not occur even if we set the level of financial development to the sample maximum.

Table 2.5: Consequences of misspecifications

	Implied responses of prices to monetary contraction (in %)				
Horizon	3 months	6 months	12 months	18 months	36 months
Best practice	-0.157	-0.331**	-0.225*	-0.155	-0.116
Without output gap	0.092	-0.028	-0.006	-0.024	-0.131
Without gap and SVAR	0.160**	0.082	0.117	0.115	-0.061
Without gap, SVAR, and commodity prices	0.205**	0.147**	0.244**	0.266**	0.165

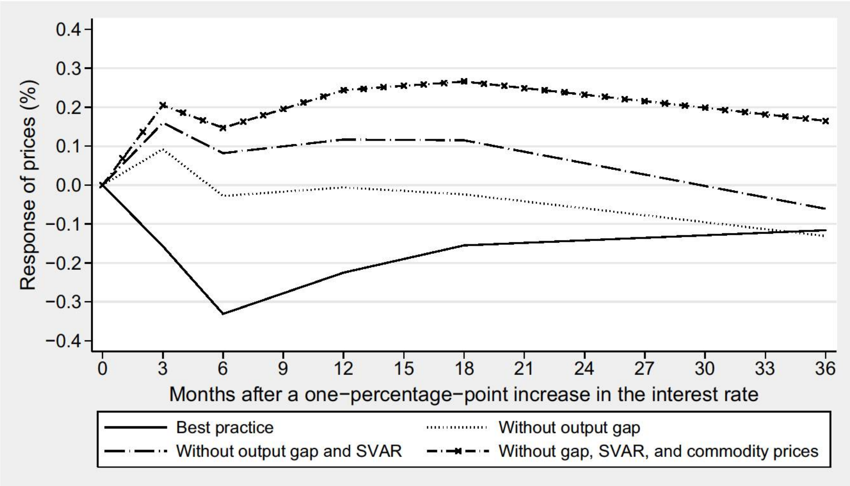
Note: The values represent the percentage response of prices to a one-percentage-point increase in the interest rate.

Without output gap = Best practice omitting output gap. Without gap and SVAR = Best practice omitting output gap and using recursive identification. Without gap, SVAR, and commodity prices = Best practice omitting output gap, using recursive identification, and omitting commodity prices.

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

To illustrate the consequences of misspecifications for the reported impulse responses, Table 2.5 and Figure 2.7 investigate the cases where some aspects of methodology deviate from best practice. When the model does not control for the potential output of the economy, the price puzzle occurs, but prices decline in the medium and long run. When the model combines the omission of the output gap with the use of recursive identification, the puzzle gets stronger, becomes statistically significant, and prices decline below the initial level only after 18 months. When the model additionally omits a measure of commodity prices, the price level is reported never to decline below the initial level during the 3-year horizon after monetary tightening. In sum, our analysis of the VAR studies on monetary transmission indicates that the price puzzle arises systematically from misspecifications of the estimated models.

Figure 2.7: Misspecifications cause the price puzzle



2.6 Conclusion

We examine the impact of monetary policy shocks on the price level by quantitatively reviewing the impulse-response functions from previously published VAR studies on monetary transmission. We collect impulse responses produced by 103 researchers for 31 countries and regress the point estimates on variables capturing study design and country characteristics. To account for within-study dependence in the estimates, we employ mixed-effects multilevel meta-regression. Recently developed meta-analysis methods allow us to estimate the underlying effect of monetary policy implied by the entire literature corrected for potential publication selection and the misspecifications of some VAR models in primary studies.

Our results indicate some evidence of publication selection against the price puzzle, and the selection seems to strengthen for responses with longer horizons after monetary tightening. The finding is in line with Doucouliagos & Stanley (2013), who suggest that publication selection is likely to be stronger for research areas with less theory competition. Macroeconomists agree about the effects of monetary policy on prices in the long run: prices should eventually decrease after a contraction. On the other hand, a smaller consensus arises regarding the exact effects of monetary policy in the short run because of the cost channel, for example. Published results often exhibit the price puzzle for the short run; on the contrary, results showing the price puzzle for the long run would be difficult to publish.

Next, we find that the reported responses of prices to a monetary tightening are systematically affected by study design and country-specific characteristics. Study design is important in particular for the short-run response. When researchers report the price puzzle, they are likely to omit commodity prices, omit potential output, and use recursive identification in their VAR model. When the biases associated with such misspecifications are filtered out, the impulse-response function inferred from the entire literature becomes hump-shaped with no evidence of the price puzzle. The maximum decrease in the price level following a one-percentage-point increase in the interest rate reaches 0.33% and occurs half a year after the tightening.

Finally, our results suggest that the long-run response of prices depends on the characteristics of the examined country; on average, the decrease in prices after a monetary contraction is relatively persistent and does not disappear within three years. The long-run effect of monetary policy is weaker in countries with high average inflation, possibly because high inflation hampers the credibility of the central bank. The effect is stronger in open economies, in countries with a more independent central bank, and during recessions.

For the sake of comparability, in this paper we only include studies using the price level in their VAR models. The robustness of our results could be further examined by conducting a meta-analysis on studies using the inflation rate. In general, the presented method of quantitative synthesis for graphical results can be applied to any other field that uses VARs as a research tool—such as, for example, the literature estimating fiscal multipliers.

References

Abreu, M., H. L. F. de Groot, & R. J. G. M. Florax (2005): "A Meta-Analysis of β-Convergence: the Legendary 2%." *Journal of Economic Surveys* **19(3)**: pp. 389–420.

Andries, M. A. (2008): "Monetary Policy Transmission Mechanism in Romania—A VAR Approach." *Theoretical and Applied Economics* **11(528)(11)**: pp. 250–260.

Angeloni, I., L. Aucremanne, M. Ehrmann, J. Galí, A. Levin, & F. Smets

(2006): "New Evidence on Inflation Persistence and Price Stickiness in the Euro Area: Implications for Macro Modeling." *Journal of the European Economic Association* **4(2-3)**: pp. 562–574.

Anzuini, A. & A. Levy (2007): "Monetary Policy Shocks in the New EU Members: A VAR Approach." *Applied Economics* **39(7-9)**: pp. 1147–1161.

Arin, K. P. & S. P. Jolly (2005): "Trans-Tasman Transmission of Monetary Shocks: Evidence from a VAR Approach." *Atlantic Economic Journal* **33(3)**: pp. 267–283.

Arnone, M., B. J. Laurens, J.-F. Segalotto, & M. Sommer (2009): "Central Bank Autonomy: Lessons from Global Trends." *IMF Staff Papers* **56(2)**: pp. 263–296.

Ashenfelter, O. & M. Greenstone (2004): "Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias." *American Economic Review* **94(2)**: pp. 454–460.

Bagliano, F. C. & C. A. Favero (1998): "Measuring Monetary Policy with VAR Models: An Evaluation." *European Economic Review* **42(6)**: pp. 1069–1112.

Bagliano, F. C. & C. A. Favero (1999): "Information from Financial Markets and VAR Measures of Monetary Policy." *European Economic Review* **43(4-6)**: pp. 825 – 837.

Banbura, M., D. Giannone, & L. Reichlin (2010): "Large Bayesian vector auto regressions." *Journal of Applied Econometrics* **25(1)**: pp. 71–92.

Bank of England (1999): "The Transmission Mechanism of Monetary Policy. A Paper by the Monetary Policy Committee." *Bank of England*, April 1999.

Barakchian, S. M. & C. Crowe (2013): "Monetary policy matters: Evidence from new shocks data." *Journal of Monetary Economics* **60(8)**: pp. 950–966.

Barth, M. J. & V. A. Ramey (2002): "The Cost Channel of Monetary Transmission." In "NBER Macroeconomics Annual 2001, Volume 16," NBER Chapters, pp. 199–256. National Bureau of Economic Research.

Belviso, F. & F. Milani (2006): "Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy." *B.E. Journal of Macroeconomics: Topics in Macroeconomics* **6(3)**: pp. 1–44.

Bernanke, B., J. Boivin, & P. S. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach." *The Quarterly Journal of Economics* **120(1)**: pp. 387–422.

Bernanke, B. & M. Gertler (1989): "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* **79(1)**: pp. 14–31.

Bernanke, B. S. (1986): "Alternative explanations of the money-income correlation." *Carnegie-Rochester Conference Series on Public Policy* **25(1)**: pp. 49–99.

Bernanke, B. S. & A. S. Blinder (1992): "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* **82(4)**: pp. 901–21.

Bernanke, B. S. & M. Gertler (1995): "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* **9(4)**: pp. 27–48.

Bernanke, B. S., M. Gertler, & M. Watson (1997): "Systematic Monetary Policy and the Effects of Oil Price Shocks." *Brookings Papers on Economic Activity* **1(1)**: pp. 91–142.

Berument, H. (2007): "Measuring monetary policy for a small open economy: Turkey." *Journal of Macroeconomics* **29(2)**: pp. 411–430.

Blanchard, O. J. & M. W. Watson (1986): "Are Business Cycles All Alike?" In "The American Business Cycle: Continuity and Change," NBER Chapters, pp. 123–180. National Bureau of Economic Research.

Boivin, J. & M. P. Giannoni (2006): "Has Monetary Policy Become More Effective?" *Review of Economics and Statistics* **88(3)**: pp. 445–462.

Boivin, J. & M. P. Giannoni (2007): "Global Forces and Monetary Policy Effectiveness." In "International Dimensions of Monetary Policy," NBER Chapters, pp. 429–478. National Bureau of Economic Research, Inc.

Borys, M., R. Horvath, & M. Franta (2009): "The effects of monetary policy in the Czech Republic: an empirical study." *Empirica* **36(4)**: pp. 419–443.

Bredin, D. & G. O'Reilly (2004): "An Analysis of the Transmission Mechanism of Monetary Policy in Ireland." *Applied Economics* **36(1)**: pp. 49–58.

Brissimis, S. N. & N. S. Magginas (2006): "Forward-Looking Information in VAR Models and the Price Puzzle." *Journal of Monetary Economics* **53(6)**: pp. 1225–1234.

Brunner, A. D. (2000): "On the Derivation of Monetary Policy Shocks: Should We Throw the VAR out with the Bath Water?" *Journal of Money, Credit, and Banking* **32(2)**: pp. 254–279.

Buckle, R. A., K. Kim, H. Kirkham, N. McLellan, & J. Sharma (2007): "A Structural VAR Business Cycle Model for a Volatile Small Open Economy." *Economic Modelling* **24(6)**: pp. 990–1017.

Canova, F. & G. D. Nicolo (2002): "Monetary disturbances matter for business fluctuations in the G-7." *Journal of Monetary Economics* **49(6)**: pp. 1131–1159.

Card, D., J. Kluve, & A. Weber (2010): "Active Labor Market Policy Evaluations: A Meta-analysis." *The Economic Journal* **120(548)**: pp. F452–F477.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review* **85(2)**: pp. 238–43.

Carlstrom, C. T., T. S. Fuerst, & M. Paustian (2009): "Monetary Policy Shocks, Choleski Identification, and DNK Models." *Journal of Monetary Economics* **56(7)**: pp. 1014 – 1021.

Cecchetti, S. G. (1999): "Legal structure, financial structure, and the monetary policy transmission mechanism." *Federal Reserve Bank of New York Economic Policy Review* **5(2)**: pp. 9–28.

Cespedes, B., E. Lima, & A. Maka (2008): "Monetary Policy, Inflation and the Level of Economic Activity in Brazil after the Real Plan: Stylized Facts from SVAR Models." *Revista Brasileira de Economia* **62(2)**: pp. 123–160.

Chetty, R., A. Guren, D. Manoli, & A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* **101(3)**: pp. 471–75.

Christiano, L. J., M. Eichenbaum, & C. Evans (1996): "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds." *The Review of Economics and Statistics* **78(1)**: pp. 16–34.

Christiano, L. J., M. Eichenbaum, & C. L. Evans (1999): "Monetary policy shocks: What have we learned and to what end?" In J. B. Taylor & M. Woodford (editors), "Handbook of Macroeconomics," volume 1 of *Handbook of Macroeconomics*, chapter 2, pp. 65–148. Elsevier.

Cloyne, J. & P. Hürtgen (2016): "The Macroeconomic Effects of Monetary Policy: A New Measure for the United Kingdom." *American Economic Journal: Macroeconomics* **8(4)**: pp. 75–102.

Coibion, O. (2012): "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics* **4(2)**: pp. 1–32.

Croushore, D. & C. L. Evans (2006): "Data Revisions and the Identification of Monetary Policy Shocks." *Journal of Monetary Economics* **53(6)**: pp. 1135–1160.

Cushman, D. O. & T. Zha (1997): "Identifying monetary policy in a small open economy under flexible exchange rates." *Journal of Monetary Economics* **39(3)**: pp. 433–448.

De Arcangelis, G. & G. Di Giorgio (2001): "Measuring Monetary Policy Shocks in a Small Open Economy." *Economic Notes* **30(1)**: pp. 81–107.

Dedola, L. & F. Lippi (2005): "The monetary transmission mechanism: Evidence from the industries of five OECD countries." *European Economic Review* **49(6)**: pp. 1543–1569.

DeLong, J. B. & K. Lang (1992): "Are all economic hypotheses false?" *Journal of Political Economy* **100(6)**: pp. 1257–72.

Dennis, R., K. Leitemo, & U. Söderström (2007): "Monetary policy in a small open economy with a preference for robustness." *Working Paper Series 2007-04*, Federal Reserve Bank of San Francisco.

Disdier, A.-C. & K. Head (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics* **90(1)**: pp. 37–48.

Doucouliagos, C. & T. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition And Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Doucouliagos, H. & T. D. Stanley (2009): "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* **47(2)**: pp. 406–428.

EFN (2004): "Monetary transmission in acceding countries." In "European Forecasting Network Annex 53," p. 97–142. European University Institute.

Egert, B. & R. MacDonald (2009): "Monetary Transmission Mechanism in Central and Eastern Europe: Surveying the Surveyable." *Journal of Economic Surveys* **23(2)**: pp. 277–327.

Eichenbaum, M. (1992): "Comment on 'Interpreting the macroeconomic time series facts: The effects of monetary policy'." *European Economic Review* **36(5)**: pp. 1001–1011.

Eickmeier, S., B. Hofmann, & A. Worms (2009): "Macroeconomic Fluctuations and Bank Lending: Evidence for Germany and the Euro Area." *German Economic Review* **10**: pp. 193–223.

Elbourne, A. (2008): "The UK Housing Market and the Monetary Policy Transmission Mechanism: An SVAR Approach." *Journal of Housing Economics* **17(1)**: pp. 65–87.

Elbourne, A. & J. de Haan (2006): "Financial structure and monetary policy transmission in transition countries." *Journal of Comparative Economics* **34(1)**: pp. 1–23.

Elbourne, A. & J. de Haan (2009): "Modeling Monetary Policy Transmission in Acceding Countries: Vector Autoregression versus Structural Vector Autoregression." *Emerging Markets Finance and Trade* **45(2)**: pp. 4–20.

European Central Bank (2010): "Monthly Bulletin." *European Central Bank*, May 2010.

Forni, M. & L. Gambetti (2010): "The dynamic effects of monetary policy: A structural factor model approach." *Journal of Monetary Economics* **57(2)**: pp. 203–216.

Fujiwara, I. (2004): "Output Composition of the Monetary Policy Transmission Mechanism in Japan." *Topics in Macroeconomics* **4(1)**: pp. 1–21.

Gan, W. B. & L. Y. Soon (2003): "Characterizing the Monetary Transmission Mechanism in a Small Open Economy: The Case of Malaysia." *Singapore Economic Review* **48(2)**: pp. 113–134.

Gavin, W. T. & D. M. Kemme (2009): "Using extraneous information to analyze monetary policy in transition economies." *Journal of International Money and Finance* **28(5)**: pp. 868–879.

Giordani, P. (2004): "An alternative explanation of the price puzzle." *Journal of Monetary Economics* **51(6)**: pp. 1271–1296.

de Grauwe, P. & C. C. Storti (2004): "The Effects of Monetary Policy: A Meta-Analysis." *CESifo Working Paper Series 1224*, CESifo Group Munich.

Görg, H. & E. Strobl (2001): "Multinational Companies and Productivity Spillovers: A Meta-analysis." *The Economic Journal* **111(475)**: pp. F723–39.

Hanson, M. S. (2004): "The price puzzle reconsidered." *Journal of Monetary Economics* **51(7)**: pp. 1385–1413.

Havranek, T. & Z. Irsova (2011): "Estimating vertical spillovers from FDI: Why results vary and what the true effect is." *Journal of International Economics* **85(2)**: pp. 234–244.

HORVATH, R. & M. RUSNAK (2009): "How Important Are Foreign Shocks in a Small Open Economy? The Case of Slovakia." *Global Economy Journal* **9(1)**.

HULSEWIG, O., E. MAYER, & T. WOLLMERSHAUSER (2006): "Bank Loan Supply and Monetary Policy Transmission in Germany: An Assessment Based on Matching Impulse Responses." *Journal of Banking and Finance* **30(10)**: pp. 2893–2910.

JANG, K. & M. OGAKI (2004): "The Effects of Monetary Policy Shocks on Exchange Rates: A Structural Vector Error Correction Model Approach." *Journal of the Japanese and International Economies* **18(1)**: pp. 99–114.

JAROCINSKI, M. (2009): "Responses to monetary policy shocks in the east and the west of Europe: a comparison." *Journal of Applied Econometrics* **25(5)**: pp. 833–868.

JAROCINSKI, M. & F. R. SMETS (2008): "House Prices and the Stance of Monetary Policy." *Federal Reserve Bank of St. Louis Review* **90(4)**: pp. 339–365.

KEEFER, P. & D. STASAVAGE (2003): "The limits to delegation: Veto players, central bank independence, and the credibility of monetary policy." *American Political Science Review* **97(3)**: p. 407–423.

KIM, S. (2001): "International Transmission of U.S. Monetary Policy Shocks: Evidence from VARs." *Journal of Monetary Economics* **48(2)**: pp. 339–372.

KIM, S. (2002): "Exchange rate stabilization in the ERM: identifying European monetary policy reactions." *Journal of International Money and Finance* **21(3)**: pp. 413–434.

KIM, S. & N. ROUBINI (2000): "Exchange Rate Anomalies in the Industrial Countries: A Solution with a Structural VAR Approach." *Journal of Monetary Economics* **45(3)**: pp. 561–586.

KRUSEC, D. (2010): "The price puzzle in the monetary transmission VARs with long-run restrictions." *Economics Letters* **106(3)**: pp. 147–150.

KUBO, A. (2008): "Macroeconomic Impact of Monetary Policy Shocks: Evidence from Recent Experience in Thailand." *Journal of Asian Economics* **19(1)**: pp. 83–91.

LAGANA, G. & A. MOUNTFORD (2005): "Measuring Monetary Policy in the UK: A Factor-Augmented Vector Autoregression Model Approach." *Manchester School* **73(s1)**: pp. 77–98.

LANGE, R. H. (2010): "Regime-switching monetary policy in Canada." *Journal of Macroeconomics* **32(3)**: pp. 782–796.

Leeper, E. M., C. A. Sims, & T. Zha (1996): "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity* **27(2)**: pp. 1–78.

Li, Y. D., T. B. Iscan, & K. Xu (2010): "The impact of monetary policy shocks on stock prices: Evidence from Canada and the United States." *Journal of International Money and Finance* **29(5)**: pp. 876–896.

McMillin, W. D. (2001): "The Effects of Monetary Policy Shocks: Comparing Contemporaneous versus Long-Run Identifying Restrictions." *Southern Economic Journal* **67(3)**: pp. 618–636.

Mertens, K. (2008): "Deposit rate ceilings and monetary transmission in the US." *Journal of Monetary Economics* **55(7)**: pp. 1290–1302.

Minella, A. (2003): "Monetary Policy and Inflation in Brazil (1975-2000): A VAR Estimation." *Revista Brasileira de Economia* **57(3)**: pp. 605–635.

Mohanty, M. S. & P. Turner (2008): "Monetary policy transmission in emerging market economies: what is new?" In "Transmission mechanisms for monetary policy in emerging market economies," volume 35 of *BIS Papers chapters*, pp. 1–59. Bank for International Settlements.

Mojon, B. (2008): "When Did Unsystematic Monetary Policy Have an Effect on Inflation?" *European Economic Review* **52(3)**: pp. 487–497.

Mojon, B. & G. Peersman (2003): "A VAR description of the effects of monetary policy in the individual countries of the Euro area." In A. K. I Angeloni & B. Mojon (editors), "Monetary Policy Transmission in the Euro Area," chapter 1, pp. 56–74. Cambridge University Press.

Mountford, A. (2005): "Leaning into the Wind: A Structural VAR Investigation of UK Monetary Policy." *Oxford Bulletin of Economics and Statistics* **67(5)**: pp. 597–621.

Nakashima, K. (2006): "The Bank of Japan's Operating Procedures and the Identification of Monetary Policy Shocks: A Reexamination Using the Bernanke-Mihov Approach." *Journal of the Japanese and International Economies* **20(3)**: pp. 406–433.

Nelson, J. & P. Kennedy (2009): "The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment." *Environmental & Resource Economics* **42(3)**: pp. 345–377.

Normandin, M. & L. Phaneuf (2004): "Monetary Policy Shocks: Testing Identification Conditions under Time-Varying Conditional Volatility." *Journal of Monetary Economics* **51(6)**: pp. 1217–1243.

Oros, C. & C. Romocea-Turcu (2009): "The Monetary Transmission Mechanisms In The CEECs: A Structural VAR Approach." *Applied Econometrics and International Development* **9(2)**: pp. 73–86.

Peersman, G. (2004): "The Transmission of Monetary Policy in the Euro Area: Are the Effects Different across Countries?" *Oxford Bulletin of Economics and Statistics* **66(3)**: pp. 285–308.

Peersman, G. (2005): "What Caused the Early Millennium Slowdown? Evidence Based on Vector Autoregressions." *Journal of Applied Econometrics* **20(2)**: pp. 185–207.

Peersman, G. & F. Smets (2003): "The Monetary Transmission Mechanism in the Euro Area: More Evidence from VAR Analysis." In A. K. I.Angeloni & B.Mojon (editors), "Monetary Policy Transmission on the Euro Area," chapter 2, pp. 36–55. Cambridge University Press.

Peersman, G. & R. Straub (2009): "Technology Shocks and Robust Sign Restrictions in a Euro Area SVAR." *International Economic Review* **50(3)**: pp. 727–750.

Perino, G. (2010): "How delegation improves commitment." *Economics Letters* **106(2)**: pp. 137–139.

Pobre, M. L. (2003): "Sources of Shocks and Monetary Policy in the 1997 Asian Crisis: The Case of Korea and Thailand." *Osaka Economic Papers* **53(3)**: pp. 362–373.

Rabanal, P. (2007): "Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model." *Journal of Economic Dynamics and Control* **31(3)**: pp. 906–937.

Rafiq, M. S. & S. K. Mallick (2008): "The Effect of Monetary Policy on Output in EMU3: A Sign Restriction Approach." *Journal of Macroeconomics* **30(4)**: pp. 1756–1791.

Rogoff, K. (1985): "The Optimal Degree of Commitment to an Intermediate Monetary Target." *The Quarterly Journal of Economics* **100(4)**: pp. 1169–89.

Romer, C. D. & D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* **94(4)**: pp. 1055–1084.

Rose, A. K. & T. D. Stanley (2005): "A Meta-Analysis of the Effect of Common Currencies on International Trade." *Journal of Economic Surveys* **19(3)**: pp. 347–365.

Shioji, E. (2000): "Identifying Monetary Policy Shocks in Japan." *Journal of the Japanese and International Economies* **14(1)**: pp. 22–42.

Sims, C. A. (1980): "Macroeconomics and Reality." *Econometrica* **48(1)**: pp. 1–48.

Sims, C. A. (1992): "Interpreting the macroeconomic time series facts: The effects of monetary policy." *European Economic Review* **36(5)**: pp. 975–1000.

Sims, C. A. & T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* **39(4)**: pp. 949–68.

Sims, C. A. & T. Zha (1999): "Error Bands for Impulse Responses." *Econometrica* **67(5)**: pp. 1113–1156.

Sims, C. A. & T. Zha (2006): "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* **96(1)**: pp. 54–81.

Slavin, R. E. (1986): "Best-Evidence Synthesis: An Alternative to Meta-Analytic and Traditional Reviews." *Educational Researcher* **15(9)**: pp. 5–11.

Smets, F. (1997): "Measuring Monetary Policy Shocks in France, Germany and Italy: The Role of the Exchange Rate." *Swiss Journal of Economics and Statistics* **133(3)**: pp. 597–616.

Smith, V. K. & J.-C. Huang (1995): "Can Markets Value Air Quality? A Meta-analysis of Hedonic Property Value Models." *Journal of Political Economy* **103(1)**: pp. 209–27.

Sousa, J. & A. Zaghini (2008): "Monetary Policy Shocks in the Euro Area and Global Liquidity Spillovers." *International Journal of Finance and Economics* **13(3)**: pp. 205–218.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2004): "Does unemployment hysteresis falsify the natural rate hypothesis? a meta-regression analysis." *Journal of Economic Surveys* **18(4)**: pp. 589–612.

Stanley, T. D. (2008): "Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

Stanley, T. D. & H. Doucouliagos (2010): "Picture This: A Simple Graph That Reveals Much Ado About Research." *Journal of Economic Surveys* **24(1)**: pp. 170–191.

Stock, J. H. & M. W. Watson (2001): "Vector Autoregressions." *Journal of Economic Perspectives* **15(4)**: pp. 101–115.

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

Uhlig, H. (2010): "Economics and Reality." *NBER Working Papers 16416*, National Bureau of Economic Research, Inc.

Vargas-Silva, C. (2008): "Monetary Policy and the US Housing Market: A VAR Analysis Imposing Sign Restrictions." *Journal of Macroeconomics* **30(3)**: pp. 977–990.

Voss, G. & L. Willard (2009): "Monetary policy and the exchange rate: Evidence from a two-country model." *Journal of Macroeconomics* **31(4)**: pp. 708–720.

Wolf, C. K. (2016): "Puzzles in Monetary Policy Identification: A Model-Based Perspective." *Technical report*, Princeton University.

Wu, T. (2003): "Stylized Facts on Nominal Term Structure and Business Cycles: An Empirical VAR Study." *Applied Economics* **35(8)**: pp. 901–906.

2.A Robustness Checks

Table 2.6: Test of publication bias and true effect, OLS

	OLS with clustered standard errors				
Horizon	3 months	6 months	12 months	18 months	36 months
Intercept (bias)	-0.277	-0.407**	-0.341**	-0.393***	-0.784***
	(0.176)	(0.186)	(0.156)	(0.147)	(0.122)
$1/SE$ (effect)	0.032**	0.033	-0.007	-0.025*	-0.018**
	(0.014)	(0.021)	(0.016)	(0.014)	(0.008)
R^2	0.05	0.03	0.00	0.02	0.01
Observations	208	215	215	217	205
Studies	69	70	70	70	63

Note: Standard errors, clustered at the study level, in parentheses. Response variable: the approximated t-statistic of the estimate of the percentage response of prices to a one-percentage-point increase in the interest rate.
***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 2.7: Explaining the differences in reported impulse responses, OLS

	OLS with clustered standard errors				
Horizon	3 months	6 months	12 months	18 months	36 months
Intercept (bias)	-0.131	-0.127	-0.240*	-0.221*	-0.538***
	(0.151)	(0.133)	(0.128)	(0.120)	(0.130)
$1/SE$	-0.058	-0.106	-0.237	-0.168	-0.028
	(0.068)	(0.115)	(0.178)	(0.174)	(0.212)
Structural heterogeneity GDP growth	-0.008	0.010	0.024*	0.027*	0.037
	(0.008)	(0.010)	(0.013)	(0.014)	(0.024)
Inflation	-0.000	-0.003	0.003	0.005**	0.008***
	(0.004)	(0.004)	(0.003)	(0.002)	(0.002)
Inflation volatility	-0.000	0.001	-0.001	-0.002**	-0.003***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Financial development	0.093***	0.079	0.174**	0.110	-0.054
	(0.030)	(0.054)	(0.076)	(0.073)	(0.067)
Openness	-0.026	-0.052	-0.089*	-0.130***	-0.258**
	(0.031)	(0.048)	(0.048)	(0.048)	(0.117)
CB independence	0.038	-0.141	-0.135	-0.258**	-0.338***
	(0.068)	(0.106)	(0.133)	(0.123)	(0.061)
Data characteristics					
No. of observations	0.020*	0.043**	0.053**	0.074***	0.127***
	(0.011)	(0.019)	(0.023)	(0.025)	(0.047)
Average year	0.001	-0.001	0.004	0.006**	0.012***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Specification characteristics					
GDP deflator	-0.004	0.023	0.119***	0.141***	0.119*
	(0.013)	(0.021)	(0.039)	(0.046)	(0.060)
Single regime	0.038**	0.034	0.024	0.021	0.109**
	(0.015)	(0.022)	(0.028)	(0.032)	(0.053)
Commodity prices	-0.047***	-0.070***	-0.139***	-0.158***	-0.212***
	(0.008)	(0.018)	(0.023)	(0.027)	(0.059)
Foreign variables	0.009	0.041***	0.068**	0.071*	0.082
	(0.015)	(0.013)	(0.030)	(0.038)	(0.055)
No. of variables	-0.022*	-0.024**	-0.039**	-0.059***	-0.153***
	(0.012)	(0.011)	(0.016)	(0.022)	(0.038)
Industrial production	0.024	0.062***	0.065**	0.069*	-0.026
	(0.016)	(0.018)	(0.032)	(0.040)	(0.041)
Output gap	-0.259***	-0.330***	-0.235***	-0.140***	0.012
	(0.090)	(0.102)	(0.060)	(0.039)	(0.031)
Other measure	-0.094***	-0.066**	-0.065	-0.044	0.018
	(0.022)	(0.030)	(0.058)	(0.077)	(0.079)

Continued on the next page

Table 2.7: Explaining the differences in reported impulse responses, OLS (continued)

	OLS with clustered standard errors				
Horizon	3 months	6 months	12 months	18 months	36 months
Estimation characteristics					
BVAR	0.136***	0.099***	0.105*	0.146	0.131
	(0.026)	(0.027)	(0.055)	(0.089)	(0.164)
FAVAR	-0.084***	-0.118***	-0.073	0.029	0.270***
	(0.025)	(0.037)	(0.054)	(0.063)	(0.068)
SVAR	-0.089***	-0.142***	-0.139***	-0.147***	-0.050
	(0.018)	(0.026)	(0.031)	(0.030)	(0.033)
Sign restrictions	-0.300***	-0.299***	-0.347***	-0.396***	-0.250
	(0.031)	(0.042)	(0.061)	(0.096)	(0.172)
Publication characteristics					
Central banker	0.024*	0.058**	0.089**	0.102**	0.125***
	(0.014)	(0.023)	(0.035)	(0.040)	(0.036)
Policymaker	-0.051**	-0.006	0.070**	0.089***	0.119***
	(0.023)	(0.022)	(0.033)	(0.032)	(0.033)
R^2	0.59	0.58	0.48	0.47	0.45
Observations	208	215	215	217	205
Studies	69	70	70	70	63

Note: Standard errors, clustered at the study level, in parentheses. Response variable: the approximated t-statistic of the estimate of the percentage response of prices to a one-percentage-point increase in the interest rate.

***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

2.B Studies Used in the Meta-Analysis

Table 2.8: List of primary studies

Andries (2008)	Anzuini & Levy (2007)	Arin & Jolly (2005)
Bagliano & Favero (1998)	Bagliano & Favero (1999)	Banbura *et al.* (2010)
Belviso & Milani (2006)	Bernanke *et al.* (2005)	Bernanke *et al.* (1997)
Berument (2007)	Boivin & Giannoni (2007)	Borys *et al.* (2009)
Bredin & O'Reilly (2004)	Brissimis & Magginas (2006)	Brunner (2000)
Buckle *et al.* (2007)	Cespedes *et al.* (2008)	Christiano *et al.* (1996)
Christiano *et al.* (1999)	Croushore & Evans (2006)	Cushman & Zha (1997)
De Arcangelis & Di Giorgio (2001)	Dedola & Lippi (2005)	EFN (2004)
Eichenbaum (1992)	Eickmeier *et al.* (2009)	Elbourne (2008)
Elbourne & de Haan (2006)	Elbourne & de Haan (2009)	Forni & Gambetti (2010)
Fujiwara (2004)	Gan & Soon (2003)	Gavin & Kemme (2009)
Hanson (2004)	Horvath & Rusnak (2009)	Hulsewig *et al.* (2006)
Jang & Ogaki (2004)	Jarocinski (2009)	Jarocinski & Smets (2008)
Kim (2001)	Kim (2002)	Krusec (2010)
Kubo (2008)	Lagana & Mountford (2005)	Lange (2010)
Leeper *et al.* (1996)	Li *et al.* (2010)	McMillin (2001)
Mertens (2008)	Minella (2003)	Mojon (2008)
Mojon & Peersman (2003)	Mountford (2005)	Nakashima (2006)
Normandin & Phaneuf (2004)	Oros & Romocea-Turcu (2009)	Peersman (2004)
Peersman (2005)	Peersman & Smets (2003)	Peersman & Straub (2009)
Pobre (2003)	Rafiq & Mallick (2008)	Romer & Romer (2004)
Shioji (2000)	Sims & Zha (2006)	Smets (1997)
Sousa & Zaghini (2008)	Vargas-Silva (2008)	Voss & Willard (2009)
Wu (2003)		

Chapter 3

Transmission Lags of Monetary Policy: A Meta-Analysis

Abstract

The transmission of monetary policy to the economy is generally thought to have long and variable lags. In this paper we quantitatively review the modern literature on monetary transmission to provide stylized facts on the average lag length and the sources of variability. We collect 67 published studies and examine when prices bottom out after a monetary contraction. The average transmission lag is 29 months, and the maximum decrease in prices reaches 0.9% on average after a one-percentage-point hike in the policy rate. Transmission lags are longer in developed economies (25–50 months) than in post-transition economies (10–20 months). We find that the factor most effective in explaining this heterogeneity is financial development: greater financial development is associated with slower transmission.

The paper was co-authored with Tomas Havranek and published in the *International Journal of Central Banking* [2013, 9(4), pp. 39–76]. The paper received *first prize in the Young Economist Competition* by the Czech Economic Society given to selected papers written by Czech economists younger than 30 years and *Economic Research Award* by the Czech National Bank given to the selected working papers by the Bank. We are grateful to Adam Elbourne, Bill Gavine, and Jakob de Haan for sending us additional data and Oxana Babecka-Kucharcukova, Marek Jarocinski, Jacques Poot, and two anonymous referees of the International Journal of Central Banking for comments on previous versions of the manuscript. Tomas Havranek acknowledges support from the Czech Science Foundation (grant #P402/11/1487). Marek Rusnak acknowledges support from the Grant Agency of Charles University (grant #267011). An online appendix with data, R and Stata code, and a list of excluded studies is available at `meta-analysis.cz/lags`.

3.1 Introduction

Policymakers need to know how long it takes before their actions fully transmit to the economy and what determines the speed of transmission. A common claim about the transmission mechanism of monetary policy is that it has “long and variable” lags (Friedman 1972; Batini & Nelson 2001; Goodhart 2001). This view has been embraced by many central banks and taken into account during their decision making: most inflation-targeting central banks have adopted a value between 12 and 24 months as their policy horizon (see, for example, Bank of England 1999; European Central Bank 2010). Theoretical models usually imply transmission lags of similar length (Taylor & Wieland 2012), but the results of empirical studies vary widely.

In this paper we quantitatively survey studies that employ vector autoregression (VAR) methods to investigate the effects of monetary policy shocks on the price level. We refer to the horizon at which the response of prices becomes the strongest as the transmission lag, and collect 198 estimates from 67 published studies. The estimates of transmission lags in our sample are indeed variable, and we examine the sources of variability. The meta-analysis approach allows us to investigate both how transmission lags differ across countries and how different estimation methodologies within the VAR framework affect the results. Meta-analysis is a set of tools for summarizing the existing empirical evidence; it has been regularly employed in medical research, but its application has only recently spread to the social sciences, including economics (Stanley 2001; Disdier & Head 2008; Card *et al.* 2010; Havranek & Irsova 2011). By bringing together evidence from a large number of studies that use different methods, meta-analysis can extract robust results from a heterogeneous literature.

Several researchers have previously investigated the cross-country differences in monetary transmission. Ehrmann (2000) examines 13 member countries of the European Union and finds relatively fast transmission to prices for most of the countries: between 2 and 8 quarters. Only France, Italy, and the United Kingdom exhibit transmission lags between 12 and 20 quarters. In contrast, Mojon & Peersman (2003) find that the effects of monetary policy shocks in European economies are much more delayed, with the maximum reaction occurring between 16 and 20 quarters after the shock. Concerning cross-country differences, Mojon & Peersman (2003) argue that the confidence intervals are too wide to draw any strong conclusions, but they call for further testing of the heterogeneity of impulse responses. Boivin *et al.* (2008) update the results and conclude that the adoption of the euro contributed to lower heterogeneity in monetary transmission among the member countries.

Cecchetti (1999) finds that for a sample of advanced countries transmission lags vary between 1 and 12 quarters. He links the country-specific strength of monetary policy to a number of indicators of financial structure, but does not attempt to explain the variation in transmission lags. In a similar vein, Elbourne & de Haan (2006) investigate 10 new EU member countries and find that the maximum effects of monetary policy shocks on prices occur between 1 and 10 quarters after the shock. These papers typically look at a small set of countries at a specific point in time; in contrast, we collect estimates of transmission lags from a vast literature that provides evidence for 30 different economies during several decades. Moreover, while some of the previous studies seek to explain the differences in the strength of transmission, they remain silent about the factors driving transmission speed.

In this paper we attempt to fill this gap and associate the differences in transmission lags with a number of country and study characteristics. Our results suggest that the transmission lags reported in the literature really do vary substantially: the average lag, corrected for misspecification in some studies, is 29 months, with a standard deviation of 19 months. Post-transition economies in our sample exhibit significantly faster transmission than advanced economies, and the only robust country-specific determinant of the length of transmission is the degree of financial development. In developed countries financial institutions have more opportunities to hedge against surprises in monetary policy stance, causing greater delays in the transmission of monetary policy shocks. Concerning variables that describe the methods used by primary studies, the frequency of the data employed matters for the reported transmission lags. Our results suggest that researchers who use monthly data instead of quarterly data report systematically faster transmission.

The remainder of the paper is structured as follows. Section 3.2 presents descriptive evidence concerning the differences in transmission lags. Section 3.3 links the variation in transmission lags to 33 country- and study-specific variables. Section 3.4 contains robustness checks. Section 3.5 summarizes the implications of our key results.

3.2 Estimating the Average Lag

We attempt to gather all published studies on monetary transmission that fulfill the following three inclusion criteria. First, the study must present an impulse response of the price level to a shock in the policy rate (that is, we exclude impulse responses of the inflation rate). Second, the impulse response in the study must correspond to a one-percentage-point shock in the interest rate, or the size of the monetary policy shock must be presented so that we can normalize the response. Third, we only include studies that present confidence intervals around the impulse responses—as a simple indicator of quality. The primary studies fulfilling the inclusion criteria are listed in Table 3.1. More details describing the search strategy can be found in a related paper (Rusnak *et al.* 2013), examining which method choices are associated with reporting the "price puzzle" (the short-term increase in the price level following a monetary contraction).

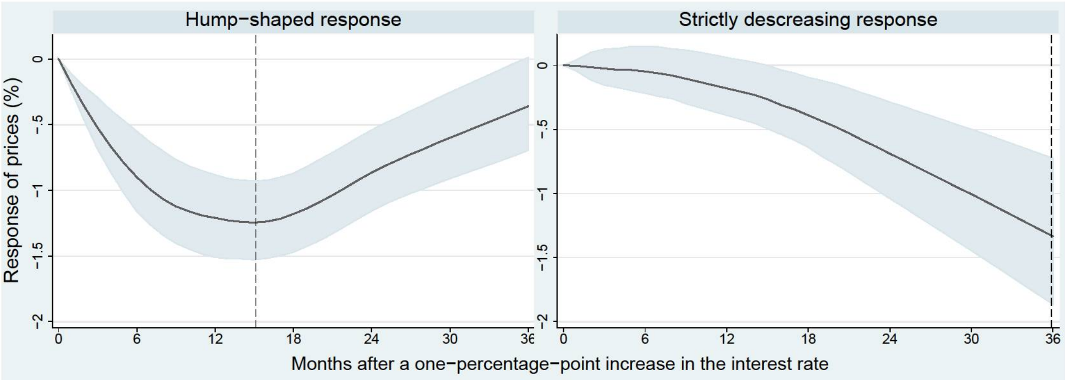
Table 3.1: List of primary studies

Andries (2008)	Jarocinski (2009)
Anzuini & Levy (2007)	Jarocinski & Smets (2008)
Arin & Jolly (2005)	Kim (2001)
Bagliano & Favero (1998)	Kim (2002)
Bagliano & Favero (1999)	Krusec (2010)
Banbura *et al.* (2010)	Kubo (2008)
Belviso & Milani (2006)	Lagana & Mountford (2005)
Bernanke *et al.* (1997)	Lange (2010)
Bernanke *et al.* (2005)	Leeper *et al.* (1996)
Boivin & Giannoni (2007)	Li *et al.* (2010)
Borys *et al.* (2009)	McMillin (2001)
Bredin & O'Reilly (2004)	Mertens (2008)
Brissimis & Magginas (2006)	Minella (2003)
Brunner (2000)	Mojon (2008)
Buckle *et al.* (2007)	Mojon & Peersman (2001)
Cespedes *et al.* (2008)	Mountford (2005)
Christiano *et al.* (1996)	Nakashima (2006)
Christiano *et al.* (1999)	Normandin & Phaneuf (2004)
Cushman & Zha (1997)	Oros & Romocea-Turcu (2009)
De Arcangelis & Di Giorgio (2001)	Peersman (2004)
Dedola & Lippi (2005)	Peersman (2005)
EFN (2004)	Peersman & Smets (2001)
Eichenbaum (1992)	Peersman & Straub (2009)
Eickmeier *et al.* (2009)	Pobre (2003)
Elbourne (2008)	Rafiq & Mallick (2008)
Elbourne & de Haan (2006)	Romer & Romer (2004)
Elbourne & de Haan (2009)	Shioji (2000)
Forni & Gambetti (2010)	Sims & Zha (1998)
Fujiwara (2004)	Smets (1997)
Gan & Soon (2003)	Sousa & Zaghini (2008)
Hanson (2004)	Vargas-Silva (2008)
Horvath & Rusnak (2009)	Voss & Willard (2009)
Hulsewig *et al.* (2006)	Wu (2003)
Jang & Ogaki (2004)	

Notes: The search for primary studies was terminated on September 15, 2010. A list of excluded studies, with reasons for exclusion, is available in the online appendix.

After imposition of the inclusion criteria, our database contains 198 impulse responses taken from 67 previously published studies and provides evidence on the monetary transmission mechanism for 30 countries, mostly developed and post-transition economies. The database is available in the online appendix. For each impulse response we evaluate the horizon at which the decrease in prices following the monetary contraction reaches its maximum. The literature reports two general types of impulse responses, both of which are depicted in Figure 3.1. The left-hand panel shows a hump-shaped (also called U-shaped) impulse response: prices decrease and bounce

Figure 3.1: Stylized impulse responses



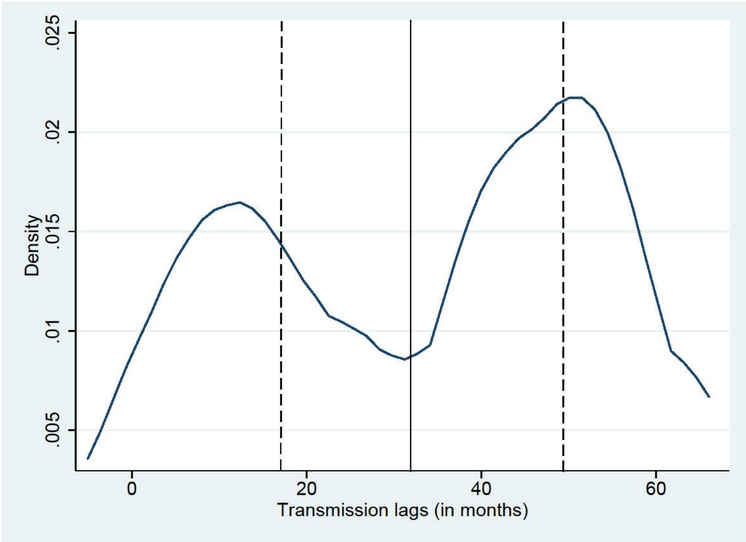
Notes: The figure depicts stylized examples of the price level's response to a one-percentage-point increase in the policy rate. The dashed lines denote the number of months to the maximum decrease in prices.

back after some time following a monetary policy shock; the monetary contraction stabilizes prices at a lower level or the effect gradually dies out. The dashed line denotes the maximum effect, and we label the corresponding number of months passed since the monetary contraction as the transmission lag. In contrast, the right-hand panel shows a strictly decreasing impulse response: prices neither stabilize nor bounce back within the time frame reported by the authors (impulse response functions are usually constructed for a five-year horizon). In this case the response of the price level becomes the strongest in the last reported horizon, so we label the last horizon as the transmission lag.

Researchers often discuss the number of months to the maximum decrease in prices in the case of hump-shaped impulse responses. On the other hand, researchers rarely interpret the timing of the maximum decrease in prices for strictly decreasing impulse responses, as the implied transmission lag often seems implausibly long. Moreover, a strictly decreasing response may indicate nonstationarity of the estimated VAR system (Lütkepohl 2005). Nevertheless we do not limit our analysis to hump-shaped impulse responses since both types are commonly reported: in the data set we have 100 estimates of transmission lags taken from hump-shaped impulse responses and 98 estimates taken from strictly decreasing impulse responses. We do not prefer any particular shape of the impulse response and focus on inference concerning the average transmission lag, but we additionally report results corresponding solely to hump-shaped impulse responses.

Figure 3.2 depicts the kernel density plot of the collected estimates; the figure demonstrates that the transmission lags taken from hump-shaped impulse responses are, on average, substantially shorter than the lags taken from strictly decreasing impulse response functions. Numerical details on summary statistics are reported in Table 3.2. The average of all collected transmission lags is 33.5 months, but

Figure 3.2: Kernel density of the estimated transmission lags



Notes: The figure is constructed using the Epanechnikov kernel function. The solid vertical line denotes the average number of months to the maximum decrease in prices taken from all the impulse responses. The dashed line on the left denotes the average taken from the hump-shaped impulse responses. The dashed line on the right denotes the average taken from the strictly decreasing impulse response functions.

Table 3.2: Summary statistics of the estimated transmission lags

Variable	Obs.	Mean	Median	Std. dev.	Min	Max
Estimates from all impulse responses	198	33.5	37	19.4	1	60
Hump-shaped impulse responses	100	18.2	15	14.1	1	57
Strictly decreasing impulse responses	98	49.1	48	8.6	24	60

the average reaches 49.1 months for transmission lags taken from strictly decreasing impulse responses and 18.2 months for hump-shaped impulse responses. In other words, the decrease in prices following a monetary contraction becomes the strongest, on average, after two years and three quarters. Our data also suggest that the average magnitude of the maximum decrease in prices following a one-percentage-point increase in the policy rate is 0.9% (for a detailed meta-analysis of the strength of monetary transmission at different horizons, see Rusnak *et al.* 2013).

Table 3.3: Transmission lags differ across countries

Developed economies		Post-transition economies	
Economy	Average transmission lag	Economy	Average transmission lag
United States	42.2	Poland	18.7
Euro area	48.4	Czech Republic	14.8
Japan	51.3	Hungary	17.9
Germany	33.4	Slovakia	10.7
United Kingdom	40.4	Slovenia	17.6
France	51.3		
Italy	26.6		

Notes: The table shows the average number of months to the maximum decrease in prices taken from all the impulse responses reported for the corresponding country. We only show results for countries for which the literature has reported at least five impulse responses.

The average of 33.5 is constructed based on data for 30 different countries. To investigate whether transmission lags vary across countries, we report country-specific averages in Table 3.3 (we only show results for countries for which we have collected at least five observations from the literature). We divide the countries into two groups: developed economies and post-transition economies.[1] From the table it is apparent that developed countries display much longer transmission lags than post-transition countries. The developed country with the fastest transmission of monetary policy actions is Italy: the corresponding transmission lag reaches 26.6 months. The slowest transmission is found for Japan and France, with a transmission lag equal to 51.3 months. In general, the transmission lags for developed countries seem to vary between approximately 25 and 50 months. These values sharply contrast with the results for post-transition countries, where all reported transmission lags lie between 10 and 20 months. The result is in line with Jarocinski (2010), who investigates cross-country differences in transmission and finds that post-communist economies exhibit faster transmission than Western European countries. We examine the possible sources of the cross-country heterogeneity in the next section.

[1] The definition of the two groups is somewhat problematic. The Czech Republic, for example, has been considered a developed economy by the World Bank since 2006. We include the country into the second group because pre-2006 time series constitute the bulk of the data used by studies in our sample.

3.3 Explaining the Differences

Two general reasons may explain why the reported transmission lags vary: First, structural differences across countries may cause genuine differences in the speed of transmission. Second, characteristics of the data and other aspects of the methodology employed in the primary studies, such as specification and estimation characteristics, may have a systematic influence on the reported transmission lag.

We collected 33 potential explanatory variables. Several structural characteristics that may account for cross-country differences in the monetary transmission mechanism have been suggested in the literature (Dornbusch *et al.* 1998; Cecchetti 1999; Ehrmann *et al.* 2003). Therefore, to control for these structural differences we include *GDP per capita* to represent the country's overall level of the development, *GDP growth* and *Inflation* to reflect other macroeconomic conditions in the economy, *Financial development* to capture the importance of the financial structure, *Openness* to cover the exchange rate channel of the transmission mechanism, and *Central bank independence* to capture the influence of the institutional setting and credibility on monetary transmission. These variables are computed as averages over the periods that correspond to the estimation periods of the primary studies. The sources of the data for these variables are Penn World Tables, the World Bank's World Development Indicators, and the International Monetary Fund's International Financial Statistics; the central bank independence index is extracted from Arnone *et al.* (2009). We also include variables that control for data, methodology, and publication characteristics of the primary studies. The definitions of the variables are provided in Table 3.4 together with their summary statistics.

Rather than estimating a regression with an ad hoc subset of explanatory variables, we formally address the model uncertainty inherent in meta-analysis (in other words, many method variables may be important for the reported speed of transmission, but no theory helps us select which ones). There are at least two drawbacks to using simple regression in situations where many potential explanatory variables exist. First, if we put all potential variables into one regression, the standard errors get inflated since many redundant variables are included. Second, sequential testing (or the "general-to-specific" approach) brings about the possibility of excluding relevant variables.

To address these issues, Bayesian model averaging (BMA) is employed frequently in the literature on the determinants of economic growth (Fernandez *et al.* 2001; Sala-I-Martin *et al.* 2004; Durlauf *et al.* 2008; Feldkircher & Zeugner 2009; Eicher *et al.* 2011). Recently, BMA has been used to address other questions as well (see Moral-Benito 2011, for a survey). The idea of BMA is to go through all possible combinations of regressors and weight them according to their model fit. BMA thus provides results robust to model uncertainty, which arises when little or nothing is known ex ante about the correct set of explanatory variables. An accessible introduction to BMA can be found in Koop (2003); technical details concerning the implementation of the method are provided by Feldkircher & Zeugner (2009).

Because we consider 33 potential explanatory variables, it is not technically feasible to enumerate all 2^{33} of their possible combinations; on a typical personal computer this would take several months. In such cases, Markov chain Monte Carlo methods are used to go through the most important models. We employ the priors suggested by Eicher *et al.* (2011), who recommend using the uniform model prior and the unit information prior for the parameters, since these priors perform well in forecasting exercises. Following Fernandez *et al.* (2001), we run the estimation with 200 million iterations, ensuring a good degree of convergence. Section 3.A provides diagnostics of our BMA estimation; the online appendix provides R and Stata codes.

Table 3.4: Description and summary statistics of explanatory variables

Variable	Description	Mean	Std. dev.
Country characteristics			
GDP per capita	The logarithm of the country's real GDP per capita.	9.880	0.415
GDP growth	The average growth rate of the country's real GDP.	2.644	1.042
Inflation	The average inflation of the country.	0.078	0.145
Financial dev.	The financial development of the country measured by (domestic credit to private sector)/GDP.	0.835	0.408
Openness	The trade openness of the country measured by (exports + imports)/GDP.	0.452	0.397
CB independence	A measure of central bank independence (Arnone *et al.* 2009).	0.773	0.145
Data characteristics			
Monthly	=1 if monthly data are used.	0.626	0.485
No. of observations	The logarithm of the number of observations used.	4.876	0.661
Average year	The average year of the data used (2000 as a base).	-9.053	7.779
Specification characteristics			
GDP deflator	=1 if the GDP deflator is used instead of the consumer price index as a measure of prices.	0.172	0.378
Single regime	=1 if the VAR is estimated over a period of a single monetary policy regime.	0.293	0.456
No. of lags	The number of lags in the model, normalized by frequency: lags/frequency	0.614	0.373
Commodity prices	=1 if a commodity price index is included.	0.626	0.485
Money	=1 if a monetary aggregate is included.	0.545	0.499
Foreign variables	=1 if at least one foreign variable is included.	0.444	0.498
Time trend	=1 if a time trend is included.	0.131	0.339
Seasonal	=1 if seasonal dummies are included.	0.146	0.354
No. of variables	The logarithm of the number of endogenous variables included in the VAR.	1.748	0.391
Industrial prod.	=1 if industrial production is used as a measure of economic activity.	0.429	0.496
Output gap	=1 if the output gap is used as a measure of economic activity.	0.030	0.172
Other measures	=1 if another measure of economic activity is used (employment, expenditures).	0.121	0.327
Estimation characteristics			
BVAR	=1 if a Bayesian VAR is estimated.	0.121	0.327
FAVAR	=1 if a factor-augmented VAR is estimated.	0.051	0.220
SVAR	=1 if non-recursive identification is employed.	0.313	0.465
Sign restrictions	=1 if sign restrictions are employed.	0.152	0.359
Publication characteristics			
Strictly decreasing	The reported impulse response function is strictly decreasing (that is, it shows the maximum decrease in prices in the last displayed horizon).	0.495	0.501
Price puzzle	The reported impulse response exhibits the price puzzle.	0.530	0.500
Study citations	The logarithm of [(Google Scholar citations of the study)/(age of the study) + 1].	1.875	1.292
Impact	The recursive RePEc impact factor of the outlet.	0.900	2.417

Continued on the next page

Table 3.4: Description and summary statistics of regression variables (continued)

Variable	Description	Mean	Std. dev.
Central banker	=1 if at least one co-author is affiliated with a central bank.	0.424	0.495
Policymaker	=1 if at least one co-author is affiliated with a Ministry of Finance, IMF, OECD, or BIS.	0.061	0.239
Native	=1 if at least one co-author is native to the investigated country.	0.449	0.499
Publication year	The year of publication (2000 as a base).	4.894	3.889

Notes: The sources of data for country characteristics are Penn World Tables, the World Bank's World Development Indicators, and the International Monetary Fund's International Financial Statistics.

The results of the BMA estimation are reported graphically in Figure 3.3. The columns represent individual regression models where the transmission lag is regressed on variables for which the corresponding cell is not blank. For example, the explanatory variables in the first model from the left are *Financial development*, *Strictly decreasing*, *Monthly*, *CB independence*, *Impact*, and *Price puzzle*. The width of the columns is proportional to the so-called posterior model probabilities; that is, it captures the weight each model gets in the BMA exercise. The figure only shows the $5,000$ models with the highest posterior model probabilities. The best models are displayed on the left-hand side and are relatively parsimonious compared to those with low posterior model probabilities. Explanatory variables in the figure are displayed in descending order according to their posterior inclusion probabilities (the sum of the posterior probabilities of the models they are included in). In other words, the variables at the top of the figure are robustly important for the explanation of the variation in transmission lags, whereas the variables at the bottom of the figure do not matter much.

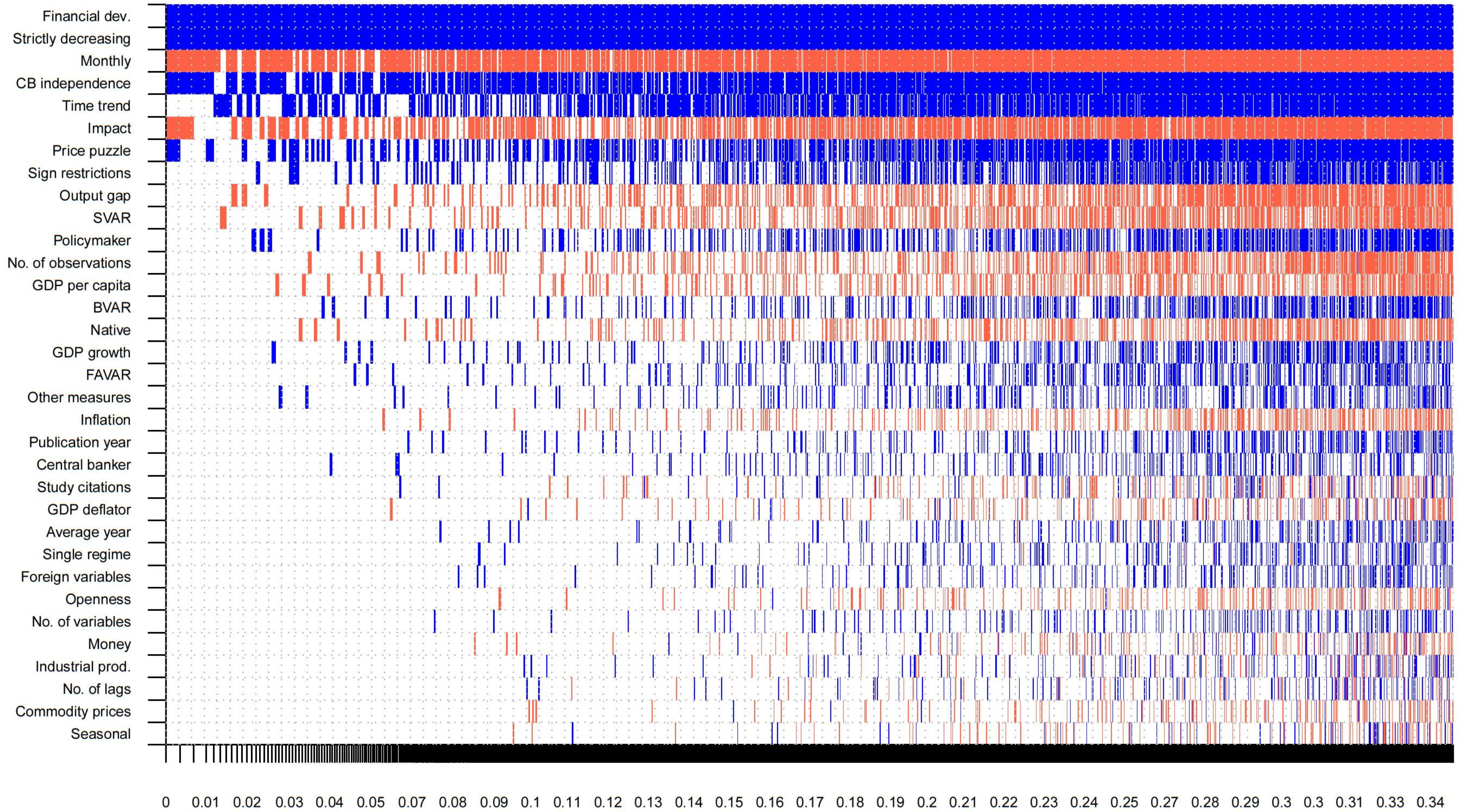
The color of the cell corresponding to each variable included in a model represents the estimated sign of the regression parameter. Blue (darker in grayscale) denotes a positive sign, and red (lighter in grayscale) denotes a negative sign. For example, in the first model from the left the estimated regression sign is positive for *Financial development*, positive for *Strictly decreasing*, negative for *Monthly*, positive for *CB independence*, negative for *Impact*, and positive for *Price puzzle*. As can be seen from the figure, variables with high posterior inclusion probabilities usually exhibit quite stable regression signs. Nevertheless, for a more precise discussion of the importance of individual variables (analogous to statistical significance in the frequentist case), we need to turn to the numerical results of the BMA estimation, reported in Table 3.5.

Table 3.5 shows the posterior means (weighted averages of the models displayed in Figure 3.3) for all regression parameters and the corresponding posterior standard deviations. According to Masanjala & Papageorgiou (2008), variables with the ratio of the posterior mean to the posterior standard deviation larger than 1.3 can be considered effective (or "statistically significant" in the frequentist case). There are only three such variables: *Financial development*, *Monthly*, and *Strictly decreasing*. First, our results suggest that a higher degree of financial development in the country is associated with slower transmission of monetary policy shocks to the price level. Moreover, when researchers use monthly data in the VAR system, they are more likely to report shorter transmission lags. The BMA exercise also corroborates that the transmission lags taken from strictly decreasing impulse responses are much longer than the lags taken from hump-shaped impulse responses; the difference is approximately 26 months.

While many of the method characteristics appear to be relatively unimportant for the explanation of the reported transmission lags, a few (for example, *Sign restrictions* or *Output gap*) have moderate posterior inclusion probabilities. Because some of the method choices are generally considered misspecifications in the literature, we use the results of the BMA estimation to filter out the effects of these misspecifications from the average transmission lag. In other words, we define an ideal study with "best-practice" methodology and maximum publication characteristics (for example the impact factor and the number of citations). Then we plug the chosen values of the explanatory variables into the results of the BMA estimation and evaluate the implied transmission lag.

For the definition of the "ideal" study we prefer the use of more observations in the VAR system (that is, we plug in the sample maximum for variable *No. of observations*), more recent data (*Average year*), the estimation of the VAR system over a period of a single monetary policy regime (*Single regime*), the inclusion of

Figure 3.3: Bayesian model averaging, model inclusion



Notes: Response variable: transmission lag (the number of months to the maximum decrease in prices taken from the impulse responses). Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures cumulative posterior model probabilities. Only the 5,000 models with the highest posterior model probabilities are shown.

Table 3.5: Why do transmission lags vary?

Variable	PIP	Posterior mean	Posterior std. dev.	Standardized coef.
Country characteristics				
GDP per capita	0.099	-0.447	1.647	-0.0096
GDP growth	0.087	0.111	0.444	0.0059
Inflation	0.053	-0.337	1.918	-0.0025
Financial dev.	**1.000**	**12.492**	**3.166**	**0.2630**
Openness	0.029	-0.056	0.631	-0.0011
CB independence	0.705	13.370	10.412	0.1002
Data characteristics				
Monthly	**0.730**	**-4.175**	**3.036**	**-0.1045**
No. of observations	0.127	-0.362	1.136	-0.0123
Average year	0.032	0.003	0.030	0.0012
Specification characteristics				
GDP deflator	0.035	-0.052	0.584	-0.0010
Single regime	0.031	0.039	0.395	0.0009
No. of lags	0.023	0.014	0.436	0.0003
Commodity prices	0.022	-0.009	0.246	-0.0002
Money	0.026	-0.011	0.286	-0.0003
Foreign variables	0.030	0.039	0.385	0.0010
Time trend	0.472	3.681	4.480	0.0643
Seasonal	0.020	-0.004	0.307	-0.0001
No. of variables	0.028	0.036	0.400	0.0007
Industrial prod.	0.025	0.008	0.352	0.0002
Output gap	0.189	-1.464	3.566	-0.0130
Other measures	0.059	0.199	1.038	0.0034
Estimation characteristics				
BVAR	0.096	0.337	1.278	0.0057
FAVAR	0.068	0.304	1.444	0.0034
SVAR	0.153	-0.468	1.303	-0.0112
Sign restrictions	0.200	0.954	2.232	0.0177
Publication characteristics				
Strictly decreasing	**1.000**	**26.122**	**1.798**	**0.6757**
Price puzzle	0.383	1.359	1.999	0.0351
Study citations	0.039	-0.005	0.205	-0.0003
Impact	0.423	-0.305	0.414	-0.0381
Central banker	0.044	0.075	0.497	0.0019
Policymaker	0.149	0.858	2.426	0.0106
Native	0.091	-0.221	0.865	-0.0057
Publication year	0.048	0.011	0.070	0.0022
Constant	1.000	7.271	NA	0.3752

Notes: Estimated by Bayesian model averaging. Response variable: transmission lag (the number of months past to the maximum decrease in prices taken from impulse responses). PIP = posterior inclusion probability. The posterior mean is analogous to the estimate of the regression coefficient in a standard regression; the posterior standard deviation is analogous to the standard error of the regression coefficient in a standard regression. Variables with posterior mean larger than 1.3 posterior standard deviations are typeset in bold; we consider such variables effective (following Masanjala & Papageorgiou 2008).

commodity prices in the VAR system (*Commodity prices*), the inclusion of foreign variables (*Foreign*), the inclusion of seasonal dummies (*Seasonal*), the inclusion of more variables in the VAR (*No. of variables*), the use of the output gap as a measure of economic activity (*Output gap*; *Industrial production* and *Other measures* are set to zero), the use of Bayesian VAR (*BVAR*), the use of sign restrictions (*Sign restrictions*; *FAVAR* and *SVAR* are set to zero), more citations of the study (*Study citations*), and a higher impact factor (*Impact*). All other variables are set to their sample means.

The average transmission lag implied by our definition of the ideal study is 29.2 months, which is less than the simple average by approximately 4 months. The estimated transmission lag hardly changes when FAVAR or SVAR are chosen for the definition of best-practice methodology; the result is also robust to other marginal changes to the definition. On the other hand, the implied transmission lag decreases greatly if one prefers hump-shaped impulse responses: in this case the estimated value is only 16.3 months. Moreover, if one prefers impulse responses that do not exhibit the price puzzle, the implied value diminishes by another month. In sum, when the effect of misspecifications is filtered out and one does not prefer any particular type of impulse response, our results suggest that prices bottom out approximately two and a half years after a monetary contraction.

3.4 Robustness Checks and Additional Results

Our analysis, based on the results of BMA, attributes the differences in transmission lags between (and within) developed and post-transition countries to differences in the level of financial development. The BMA exercise carried out in the previous section controls for methodology and other aspects associated with estimating impulse responses. Nevertheless, it is still useful to illustrate that the differences in results between developed and post-transition countries are not caused by differences in the frequency of reporting strictly decreasing impulse responses or impulse responses showing the price puzzle. To this end, we replicate Table 3.6 but only focus on the subsamples of impulse responses that are hump-shaped (Table 3.6) or that do not exhibit the price puzzle (Table 3.7).

Table 3.6: Transmission lags differ across countries (hump-shaped impulse responses)

Developed economies		Post-transition economies	
Economy	Average transmission lag	Economy	Average transmission lag
United States	23.2	Poland	15.4
Euro area	39.5	Czech Republic	14.8
Japan	40.5	Hungary	14.4
Germany	19.4	Slovakia	5.0
United Kingdom	10.0	Slovenia	13.0
France	24.0		
Italy	9.2		

Notes: The table shows the average number of months to the maximum decrease in prices taken from the impulse responses reported for the corresponding country. Strictly decreasing impulse responses are omitted from this analysis.

Table 3.7: Transmission lags differ across countries (responses not showing the price puzzle)

Developed economies		Post-transition economies	
Economy	Average transmission lag	Economy	Average transmission lag
United States	40.5	Poland	14.0
Euro area	49.2	Czech Republic	8.8
Japan	57.0	Hungary	15.4
Germany	34.5	Slovakia	10.7
United Kingdom	10.0	Slovenia	17.8
France	52.8		
Italy	30.0		

Notes: The table shows the average number of months to the maximum decrease in prices taken from the impulse responses reported for the corresponding country. Impulse responses exhibiting the price puzzle are omitted from this analysis.

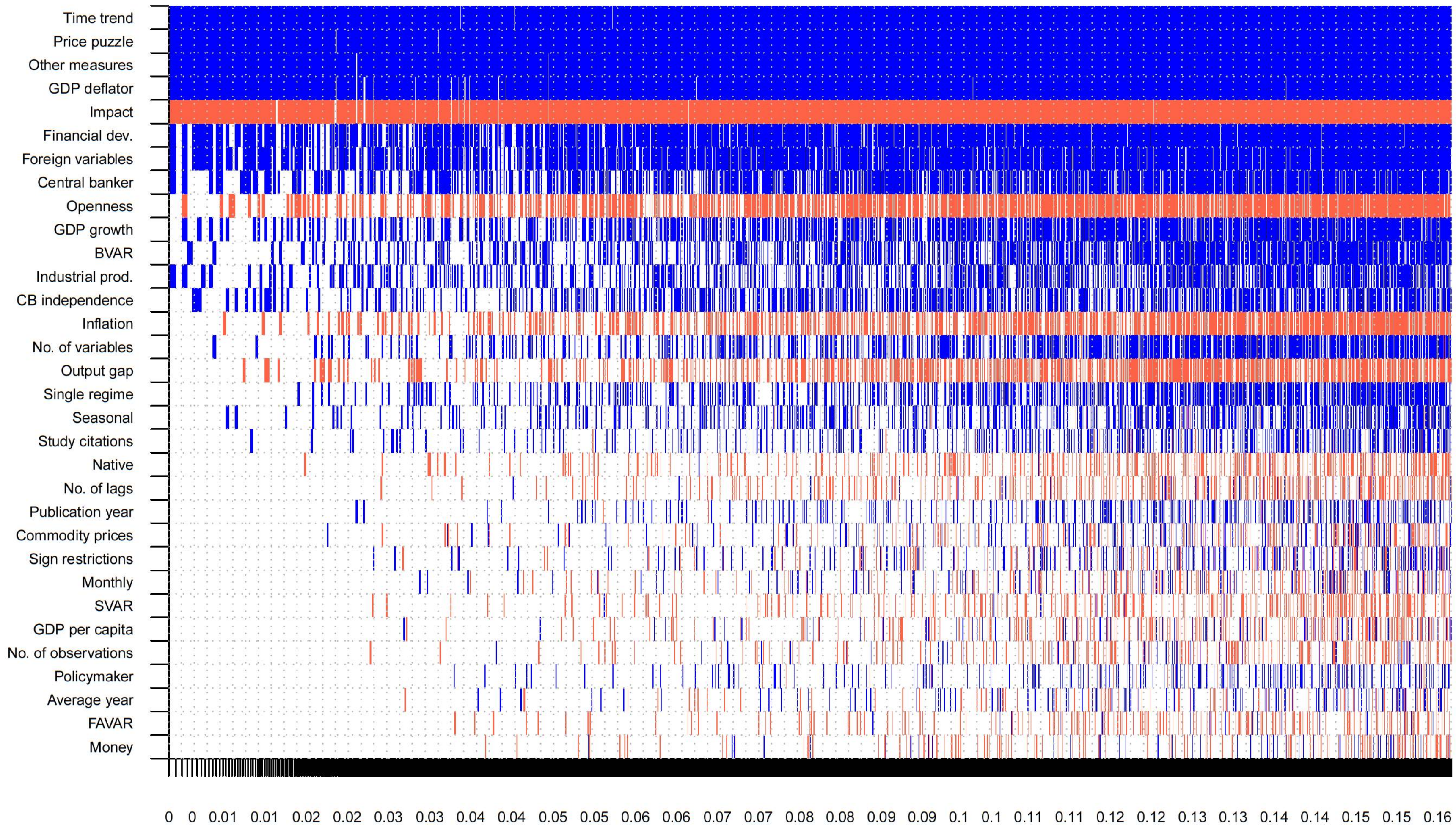
The tables show that developed countries exhibit longer transmission lags even if strictly decreasing impulse responses or impulse responses showing the price puzzle are disregarded. But the difference is smaller for the subsample of hump-shaped impulse responses, where some developed countries (for example, Italy) exhibit shorter transmission lags than some post-transition countries (for example, Poland). There are two potential explanations of this result. First, compared with Table 3.3, now we only have approximately half the number of observations, and for some countries we are even left with less than five impulse responses, which makes the average number imprecise. Second, strictly decreasing impulse responses, which are associated with longer transmission lags, are more often reported for developed economies than for post-transition economies. The reason is that shorter data spans are available for post-transition countries, which makes researchers often choose monthly data. Since monthly data are associated with shorter reported lags, researchers investigating monetary transmission in post-transition countries are less likely to report strictly decreasing impulse responses. Nevertheless, in the BMA estimation we control for data frequency as well as for the shape of the impulse response, and financial development still emerges as the most important factor causing cross-country differences in transmission lags.

In our baseline model from the previous section we combine data from hump-shaped and strictly decreasing impulse response functions. For strictly decreasing impulse responses, however, our definition of the transmission lag (the maximum effect of a monetary contraction on prices) is influenced by the reporting window chosen by researchers. To see whether the result concerning financial development is robust to omitting data from strictly decreasing impulse response functions, we repeat the BMA estimation from the previous section using a subsample of hump-shaped impulse responses.

The results are presented graphically in Figure 3.4. The variable corresponding to financial development retain its estimated sign from the baseline model and still represents the most important country-level factor explaining the differences in monetary transmission lags. Compared to the baseline model, in this specification additional method variables seem to be important. The use of other measures than GDP, the output gap, or industrial production as a proxy for economic activity is associated with slower reported transmission. The choice to represent prices by the GDP deflator instead of the consumer price index on average translates into longer transmission lags. Also the inclusion of foreign variables in the VAR system makes researchers report slower transmission.

By excluding all strictly decreasing impulse responses, however, we lose half of the information contained in our data set. For this reason we consider a second way of taking into account the effect of the reporting window: censored regression. The

Figure 3.4: Bayesian model averaging, model inclusion (hump-shaped impulse responses)



Notes: Response variable: transmission lag (the number of months to the maximum decrease in prices taken from the impulse responses). Only transmission lags from hump-shaped impulse responses are included in the estimation. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures cumulative posterior model probabilities. Only the 5,000 models with the highest posterior model probabilities are shown.

reporting window of primary studies is often set to five years, so we use 60 months as the upper limit and estimate the regression using the Tobit model. (Changing the upper limit to three or four years, which are sometimes used as the reporting window, does not qualitatively affect the results). Unfortunately, it is cumbersome to estimate Tobit using BMA. Thus, we estimate a general model with all potential explanatory variables and then employ the general-to-specific approach. The general model is reported in Table B1 in Section 3.B. The inclusion of all potential explanatory variables, many of which may not be important for explanation of the differences in transmission lags, inflates the standard errors of the relevant variables. Hence, in the next step we eliminate the insignificant variables one by one, starting from the least significant variable. As mentioned before, the general-to-specific approach is far from perfect—but in this case it represents an easy alternative to BMA.

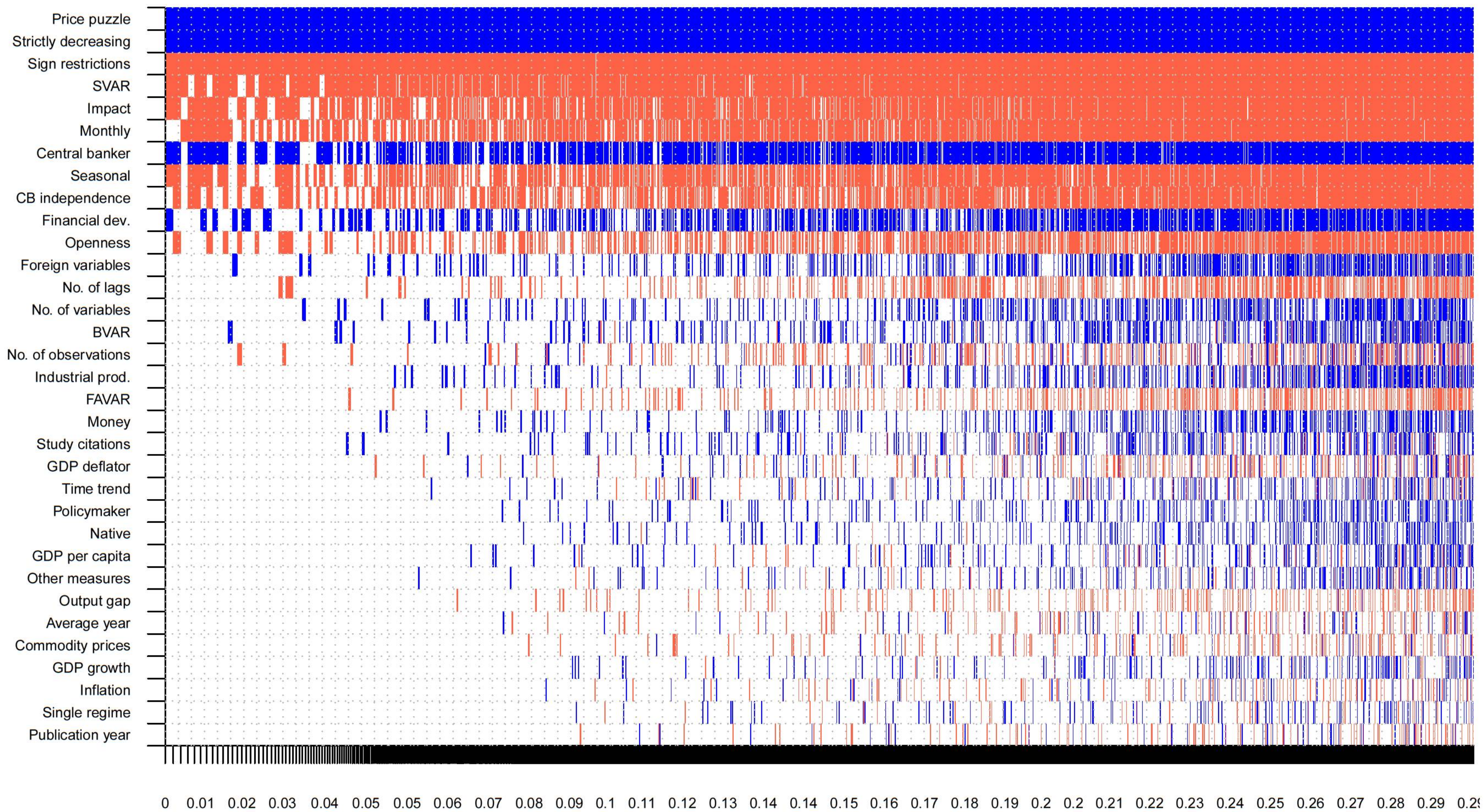
Table 3.8: Censored regression, specific model

Response variable: transmission lag		
GDP per capita	-11.48**	(4.793)
Price puzzle	4.667**	(2.343)
Inflation	-17.25**	(8.739)
Financial dev.	21.61***	(5.375)
Openness	-12.67***	(4.670)
CB independence	29.38***	(10.64)
Monthly	-12.04***	(3.821)
No. of observations	6.526**	(2.951)
Policymaker	12.37**	(5.012)
Constant	86.58**	(43.69)
Observations	198	

Notes: Standard errors in parentheses. Estimated by Tobit with the upper limit for transmission lags equal to 60 months. The specific model is a result of the backward stepwise regression procedure applied to the general model, which is reported in Section 3.B (the cut-off level for p-values was 0.1). ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The results presented in Table 3.8 and Table B1 corroborate that, even using this methodology, financial development is highly important for the explanation of transmission lags; in both specifications it is significant at the 1% level. The use of monthly data is associated with faster reported transmission, which is also consistent with the baseline model. In line with our results from the previous sections, Table 3.8

Figure 3.5: Bayesian model averaging, model inclusion (time to -0.1% decrease in prices)



Notes: Response variable: the number of months to a -0.1% decrease in prices following a one-percentage-point increase in the policy rate. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures cumulative posterior model probabilities. Only the 5,000 models with the highest posterior model probabilities are shown.

suggests that impulse responses exhibiting the price puzzle are likely to show longer transmission lags. In contrast to the baseline model, some other variables seem to be important as well: *GDP per capita*, *Inflation*, and *Openness*, among others. Because, however, the results concerning these variables are not confirmed by other specifications, we do not want to put much emphasis on these variables. The variable *Strictly decreasing*, which was crucial for the baseline BMA estimation, is omitted from the present analysis because it defines the censoring process.

So far we have analyzed the time it takes before a monetary contraction translates into the maximum effect on the price level. The extent of the maximum effect, however, varies a lot across different impulse responses. Therefore, as a complement to the previous analysis, we collect data on how long it takes before a one-percentage-point increase in the policy rate leads to a decrease in the price level of 0.1%. This number was chosen because most of the impulse response functions in our sample (173 out of 198) reach this level at some point. In contrast, if we chose a value of 0.5%, for example, we would have to disregard almost two thirds of all the impulse responses.

The results of the BMA estimation using the new response variable are reported in Figure 3.5. Again, the shape of the impulse response and the frequency of the data used in the VAR system seem to be associated with the reported transmission lag. Financial development still belongs among the most important country-level variables, together with central bank independence and trade openness. According to this specification, monetary transmission is faster in countries that are more open to international trade and that have a more independent central bank; these results may point at the importance of the exchange rate and expectation channels of monetary transmission. Additionally, some method variables matter for the estimated transmission lag: for example, the use of sign restrictions, structural VAR, and seasonal adjustment. Our results also suggest that articles published in journals with a high impact factor tend to present faster monetary transmission.

3.5 Concluding Remarks

Building on a sample of 67 previous empirical studies, we examine why the reported transmission lags of monetary policy vary. Our results suggest that the cross-country variation in transmission is robustly associated with differences in financial development. To explain the variation of results between different studies for the same country, the frequency of the data used is important: the use of monthly data makes researchers report transmission faster by 4 months, holding other things constant. This is in line with Ghysels (2012), who shows that responses from low- and high-frequency VARs may indeed differ due to mixed-frequency sampling or temporal aggregation of shocks. The shape of the impulse response matters as well. Strictly decreasing impulse responses, which may suggest that the underlying VAR system is not stationary, exhibit much longer transmission lags.

The key result of our meta-analysis is that a higher degree of financial development translates into slower transmission of monetary policy. The finding can be interpreted in the following way. If financial institutions lack opportunities to protect themselves against unexpected monetary policy actions (due to either low levels of capitalization or low sophistication of financial instruments provided by the undeveloped financial system), they need to react immediately to monetary policy shocks, thus speeding up the transmission. In financially developed countries, in contrast, financial institutions have more opportunities to hedge against surprises in monetary policy stance, causing greater delays in the transmission of monetary policy shocks.

More generally, our results imply that monetary transmission may slow down as the financial system of emerging countries develops, since financial innovations allow banks to protect better against surprise shocks in monetary policy.

References

Andries, M. A. (2008): "Monetary Policy Transmission Mechanism in Romania—A VAR Approach." *Theoretical and Applied Economics* **11**: pp. 250–260.

Anzuini, A. & A. Levy (2007): "Monetary Policy Shocks in the New EU Members: A VAR Approach." *Applied Economics* **39(7-9)**: pp. 1147–1161.

Arin, K. P. & S. P. Jolly (2005): "Trans-Tasman Transmission of Monetary Shocks: Evidence from a VAR Approach." *Atlantic Economic Journal* **33(3)**: pp. 267–283.

Arnone, M., B. J. Laurens, J.-F. Segalotto, & M. Sommer (2009): "Central Bank Autonomy: Lessons from Global Trends." *IMF Staff Papers* **56(2)**: pp. 263–296.

Bagliano, F. C. & C. A. Favero (1998): "Measuring Monetary Policy with VAR Models: An Evaluation." *European Economic Review* **42(6)**: pp. 1069–1112.

Bagliano, F. C. & C. A. Favero (1999): "Information from Financial Markets and VAR Measures of Monetary Policy." *European Economic Review* **43(4-6)**: pp. 825 – 837.

Banbura, M., D. Giannone, & L. Reichlin (2010): "Large Bayesian vector auto regressions." *Journal of Applied Econometrics* **25(1)**: pp. 71–92.

Bank of England (1999): "The Transmission Mechanism of Monetary Policy. A Paper by the Monetary Policy Committee." *Bank of England*, April 1999.

Batini, N. & E. Nelson (2001): "The Lag from Monetary Policy Actions to Inflation: Friedman Revisited." *International Finance* **4(3)**: pp. 381–400.

Belviso, F. & F. Milani (2006): "Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy." *B.E. Journal of Macroeconomics: Topics in Macroeconomics* **6(3)**: pp. 1–44.

Bernanke, B., J. Boivin, & P. S. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach." *The Quarterly Journal of Economics* **120(1)**: pp. 387–422.

Bernanke, B. S., M. Gertler, & M. Watson (1997): "Systematic Monetary Policy and the Effects of Oil Price Shocks." *Brookings Papers on Economic Activity* **1(1)**: pp. 91–142.

Boivin, J. & M. P. Giannoni (2007): "Global Forces and Monetary Policy Effectiveness." In "International Dimensions of Monetary Policy," NBER Chapters, pp. 429–478. National Bureau of Economic Research, Inc.

Boivin, J., M. P. Giannoni, & B. Mojon (2008): "How Has the Euro Changed the Monetary Transmission?" In "NBER Macroeconomics Annual 2008," NBER Chapters. National Bureau of Economic Research, Inc.

Borys, M., R. Horvath, & M. Franta (2009): "The effects of monetary policy in the Czech Republic: an empirical study." *Empirica* **36(4)**: pp. 419–443.

Bredin, D. & G. O'Reilly (2004): "An Analysis of the Transmission Mechanism of Monetary Policy in Ireland." *Applied Economics* **36(1)**: pp. 49–58.

Brissimis, S. N. & N. S. Magginas (2006): "Forward-Looking Information in VAR Models and the Price Puzzle." *Journal of Monetary Economics* **53(6)**: pp. 1225–1234.

Brunner, A. D. (2000): "On the Derivation of Monetary Policy Shocks: Should We Throw the VAR out with the Bath Water?" *Journal of Money, Credit, and Banking* **32(2)**: pp. 254–279.

Buckle, R. A., K. Kim, H. Kirkham, N. McLellan, & J. Sharma (2007): "A Structural VAR Business Cycle Model for a Volatile Small Open Economy." *Economic Modelling* **24(6)**: pp. 990–1017.

Card, D., J. Kluve, & A. Weber (2010): "Active Labor Market Policy Evaluations: A Meta-analysis." *The Economic Journal* **120(548)**: pp. F452–F477.

Cecchetti, S. G. (1999): "Legal structure, financial structure, and the monetary policy transmission mechanism." *Federal Reserve Bank of New York Economic Policy Review* **5(2)**: pp. 9–28.

Cespedes, B., E. Lima, & A. Maka (2008): "Monetary Policy, Inflation and the Level of Economic Activity in Brazil after the Real Plan: Stylized Facts from SVAR Models." *Revista Brasileira de Economia* **62(2)**: pp. 123–160.

Christiano, L. J., M. Eichenbaum, & C. Evans (1996): "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds." *The Review of Economics and Statistics* **78(1)**: pp. 16–34.

Christiano, L. J., M. Eichenbaum, & C. L. Evans (1999): "Monetary policy shocks: What have we learned and to what end?" In J. B. Taylor & M. Woodford (editors), "Handbook of Macroeconomics," volume 1 of *Handbook of Macroeconomics*, chapter 2, pp. 65–148. Elsevier.

Cushman, D. O. & T. Zha (1997): "Identifying monetary policy in a small open economy under flexible exchange rates." *Journal of Monetary Economics* **39(3)**: pp. 433–448.

De Arcangelis, G. & G. Di Giorgio (2001): "Measuring Monetary Policy Shocks in a Small Open Economy." *Economic Notes* **30(1)**: pp. 81–107.

Dedola, L. & F. Lippi (2005): "The monetary transmission mechanism: Evidence from the industries of five OECD countries." *European Economic Review* **49(6)**: pp. 1543–1569.

Disdier, A.-C. & K. Head (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics* **90(1)**: pp. 37–48.

Dornbusch, R., C. Favero, F. Giavazzi, H. Genberg, & A. K. Rose (1998): "Immediate Challenges for the European Central Bank." *Economic Policy* **13(26)**: pp. 17–64.

Durlauf, S., A. Kourtellos, & C. Tan (2008): "Are Any Growth Theories Robust?" *Economic Journal* **118(527)**: pp. 329–346.

EFN (2004): "Monetary transmission in acceding countries." In "European Forecasting Network Annex 53," p. 97–142. European University Institute.

Ehrmann, M. (2000): "Comparing Monetary Policy Transmission across European Countries." *Review of World Economics* **136(1)**: pp. 58 – 83.

Ehrmann, M., L. Gambacorta, J. Martinez-Pages, P. Sevestre, & A. Worms (2003): "The Effects of Monetary Policy in the Euro Area." *Oxford Review of Economic Policy* **19(1)**: pp. 58–72.

Eichenbaum, M. (1992): "Comment on 'Interpreting the macroeconomic time series facts: The effects of monetary policy'." *European Economic Review* **36(5)**: pp. 1001–1011.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Eickmeier, S., B. Hofmann, & A. Worms (2009): "Macroeconomic Fluctuations and Bank Lending: Evidence for Germany and the Euro Area." *German Economic Review* **10**: pp. 193–223.

Elbourne, A. (2008): "The UK Housing Market and the Monetary Policy Transmission Mechanism: An SVAR Approach." *Journal of Housing Economics* **17(1)**: pp. 65–87.

Elbourne, A. & J. de Haan (2006): "Financial structure and monetary policy transmission in transition countries." *Journal of Comparative Economics* **34(1)**: pp. 1–23.

Elbourne, A. & J. de Haan (2009): "Modeling Monetary Policy Transmission in Acceding Countries: Vector Autoregression versus Structural Vector Autoregression." *Emerging Markets Finance and Trade* **45(2)**: pp. 4–20.

European Central Bank (2010): "Monthly Bulletin." *European Central Bank*, May 2010.

Feldkircher, M. & S. Zeugner (2009): "Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging." *IMF Working Papers 09/202*, International Monetary Fund.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Model uncertainty in cross-country growth regressions." *Journal of Applied Econometrics* **16(5)**: pp. 563–576.

Forni, M. & L. Gambetti (2010): "The dynamic effects of monetary policy: A structural factor model approach." *Journal of Monetary Economics* **57(2)**: pp. 203–216.

Friedman, M. (1972): "Have Monetary Policies Failed?" *American Economic Review* **62(2)**: pp. 11–18.

Fujiwara, I. (2004): "Output Composition of the Monetary Policy Transmission Mechanism in Japan." *Topics in Macroeconomics* **4(1)**: pp. 1–21.

Gan, W. B. & L. Y. Soon (2003): "Characterizing the Monetary Transmission Mechanism in a Small Open Economy: The Case of Malaysia." *Singapore Economic Review* **48(2)**: pp. 113–134.

Ghysels, E. (2012): "Macroeconomics and the Reality of Mixed Frequency Data." *University of North Carolina (UNC) at Chapel Hill - Department of Economics.* mimeo.

Goodhart, C. A. (2001): "Monetary transmission lags and the formulation of the policy decision on interest rates." *Review, Federal Reserve Bank of St. Louis* **(Jul)**: pp. 165–186.

Hanson, M. S. (2004): "The price puzzle reconsidered." *Journal of Monetary Economics* **51(7)**: pp. 1385–1413.

Havranek, T. & Z. Irsova (2011): "Estimating vertical spillovers from FDI: Why results vary and what the true effect is." *Journal of International Economics* **85(2)**: pp. 234–244.

Horvath, R. & M. Rusnak (2009): "How Important Are Foreign Shocks in a Small Open Economy? The Case of Slovakia." *Global Economy Journal* **9(1)**.

HULSEWIG, O., E. MAYER, & T. WOLLMERSHAUSER (2006): "Bank Loan Supply and Monetary Policy Transmission in Germany: An Assessment Based on Matching Impulse Responses." *Journal of Banking and Finance* **30(10)**: pp. 2893–2910.

JANG, K. & M. OGAKI (2004): "The Effects of Monetary Policy Shocks on Exchange Rates: A Structural Vector Error Correction Model Approach." *Journal of the Japanese and International Economies* **18(1)**: pp. 99–114.

JAROCINSKI, M. (2009): "Responses to monetary policy shocks in the east and the west of Europe: a comparison." *Journal of Applied Econometrics* **25(5)**: pp. 833–868.

JAROCINSKI, M. (2010): "Responses to monetary policy shocks in the east and the west of Europe: a comparison." *Journal of Applied Econometrics* **25(5)**: pp. 833–868.

JAROCINSKI, M. & F. R. SMETS (2008): "House Prices and the Stance of Monetary Policy." *Federal Reserve Bank of St. Louis Review* **90(4)**: pp. 339–365.

KIM, S. (2001): "International Transmission of U.S. Monetary Policy Shocks: Evidence from VARs." *Journal of Monetary Economics* **48(2)**: pp. 339–372.

KIM, S. (2002): "Exchange rate stabilization in the ERM: identifying European monetary policy reactions." *Journal of International Money and Finance* **21(3)**: pp. 413–434.

KOOP, G. (2003): *Bayesian Econometrics.* John Wiley & Sons.

KRUSEC, D. (2010): "The price puzzle in the monetary transmission VARs with long-run restrictions." *Economics Letters* **106(3)**: pp. 147–150.

KUBO, A. (2008): "Macroeconomic Impact of Monetary Policy Shocks: Evidence from Recent Experience in Thailand." *Journal of Asian Economics* **19(1)**: pp. 83–91.

Lagana, G. & A. Mountford (2005): "Measuring Monetary Policy in the UK: A Factor-Augmented Vector Autoregression Model Approach." *Manchester School* **73(s1)**: pp. 77–98.

Lange, R. H. (2010): "Regime-switching monetary policy in Canada." *Journal of Macroeconomics* **32(3)**: pp. 782–796.

Leeper, E. M., C. A. Sims, & T. Zha (1996): "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity* **27(2)**: pp. 1–78.

Li, Y. D., T. B. Iscan, & K. Xu (2010): "The impact of monetary policy shocks on stock prices: Evidence from Canada and the United States." *Journal of International Money and Finance* **29(5)**: pp. 876–896.

Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis.* Springer-Verlag.

Masanjala, W. H. & C. Papageorgiou (2008): "Rough and lonely road to prosperity: a reexamination of the sources of growth in Africa using Bayesian model averaging." *Journal of Applied Econometrics* **23(5)**: pp. 671–682.

McMillin, W. D. (2001): "The Effects of Monetary Policy Shocks: Comparing Contemporaneous versus Long-Run Identifying Restrictions." *Southern Economic Journal* **67(3)**: pp. 618–636.

Mertens, K. (2008): "Deposit rate ceilings and monetary transmission in the US." *Journal of Monetary Economics* **55(7)**: pp. 1290–1302.

Minella, A. (2003): "Monetary Policy and Inflation in Brazil (1975-2000): A VAR Estimation." *Revista Brasileira de Economia* **57(3)**: pp. 605–635.

Mojon, B. (2008): "When Did Unsystematic Monetary Policy Have an Effect on Inflation?" *European Economic Review* **52(3)**: pp. 487–497.

Mojon, B. & G. Peersman (2001): "A VAR description of the effects of monetary policy in the individual countries of the Euro area." *Working Paper Series 092*, European Central Bank.

Mojon, B. & G. Peersman (2003): "A VAR description of the effects of monetary policy in the individual countries of the Euro area." In A. K. I Angeloni & B. Mojon (editors), "Monetary Policy Transmission in the Euro Area," chapter 1, pp. 56–74. Cambridge University Press.

Moral-Benito, E. (2011): "Model averaging in economics." *Banco de Espana Working Papers 1123*, Banco de Espana.

Mountford, A. (2005): "Leaning into the Wind: A Structural VAR Investigation of UK Monetary Policy." *Oxford Bulletin of Economics and Statistics* **67(5)**: pp. 597–621.

Nakashima, K. (2006): "The Bank of Japan's Operating Procedures and the Identification of Monetary Policy Shocks: A Reexamination Using the Bernanke-Mihov Approach." *Journal of the Japanese and International Economies* **20(3)**: pp. 406–433.

Normandin, M. & L. Phaneuf (2004): "Monetary Policy Shocks: Testing Identification Conditions under Time-Varying Conditional Volatility." *Journal of Monetary Economics* **51(6)**: pp. 1217–1243.

Oros, C. & C. Romocea-Turcu (2009): "The Monetary Transmission Mechanisms In The CEECs: A Structural VAR Approach." *Applied Econometrics and International Development* **9(2)**: pp. 73–86.

Peersman, G. (2004): "The Transmission of Monetary Policy in the Euro Area: Are the Effects Different across Countries?" *Oxford Bulletin of Economics and Statistics* **66(3)**: pp. 285–308.

Peersman, G. (2005): "What Caused the Early Millennium Slowdown? Evidence Based on Vector Autoregressions." *Journal of Applied Econometrics* **20(2)**: pp. 185–207.

Peersman, G. & F. Smets (2001): "The monetary transmission mechanism in the

Euro area: more evidence from VAR analysis." *Working Paper Series 091*, European Central Bank.

Peersman, G. & R. Straub (2009): "Technology Shocks and Robust Sign Restrictions in a Euro Area SVAR." *International Economic Review* **50(3)**: pp. 727–750.

Pobre, M. L. (2003): "Sources of Shocks and Monetary Policy in the 1997 Asian Crisis: The Case of Korea and Thailand." *Osaka Economic Papers* **53(3)**: pp. 362–373.

Rafiq, M. S. & S. K. Mallick (2008): "The Effect of Monetary Policy on Output in EMU3: A Sign Restriction Approach." *Journal of Macroeconomics* **30(4)**: pp. 1756–1791.

Romer, C. D. & D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* **94(4)**: pp. 1055–1084.

Rusnak, M., T. Havranek, & R. Horvath (2013): "How to Solve the Price Puzzle? A Meta-Analysis." *Journal of Money, Credit and Banking* **45(1)**: pp. 37–70.

Sala-I-Martin, X., G. Doppelhofer, & R. I. Miller (2004): "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review* **94(4)**: pp. 813–835.

Shioji, E. (2000): "Identifying Monetary Policy Shocks in Japan." *Journal of the Japanese and International Economies* **14(1)**: pp. 22–42.

Sims, C. A. & T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* **39(4)**: pp. 949–68.

Smets, F. (1997): "Measuring Monetary Policy Shocks in France, Germany and Italy: The Role of the Exchange Rate." *Swiss Journal of Economics and Statistics* **133(3)**: pp. 597–616.

Sousa, J. & A. Zaghini (2008): "Monetary Policy Shocks in the Euro Area and Global Liquidity Spillovers." *International Journal of Finance and Economics* **13(3)**: pp. 205–218.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Taylor, J. B. & V. Wieland (2012): "Surprising comparative properties of monetary models: Results from a new model database." *The Review of Economics and Statistics* **94(3)**: pp. 800–816.

Vargas-Silva, C. (2008): "Monetary Policy and the US Housing Market: A VAR Analysis Imposing Sign Restrictions." *Journal of Macroeconomics* **30(3)**: pp. 977–990.

Voss, G. & L. Willard (2009): "Monetary policy and the exchange rate: Evidence from a two-country model." *Journal of Macroeconomics* **31(4)**: pp. 708–720.

Wu, T. (2003): "Stylized Facts on Nominal Term Structure and Business Cycles: An Empirical VAR Study." *Applied Economics* **35(8)**: pp. 901–906.

3.A Diagnostics of Bayesian Model Averaging

Table 3.9: Summary of BMA estimation (baseline model)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
8.1261	$2 \cdot 10^8$	$1 \cdot 10^8$	11.88852 hours
No. models visited	*Modelspace*	*Visited*	*Topmodels*
83,511,152	$8.6 \cdot 10^9$	0.97%	34%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9999	198	uniform / 16.5	UIP
Shrinkage-Stats			
Av= 0.995			

Notes: UIP = unit information prior, PMP = posterior model probability.

Figure 3.6: Model size and convergence (baseline model)

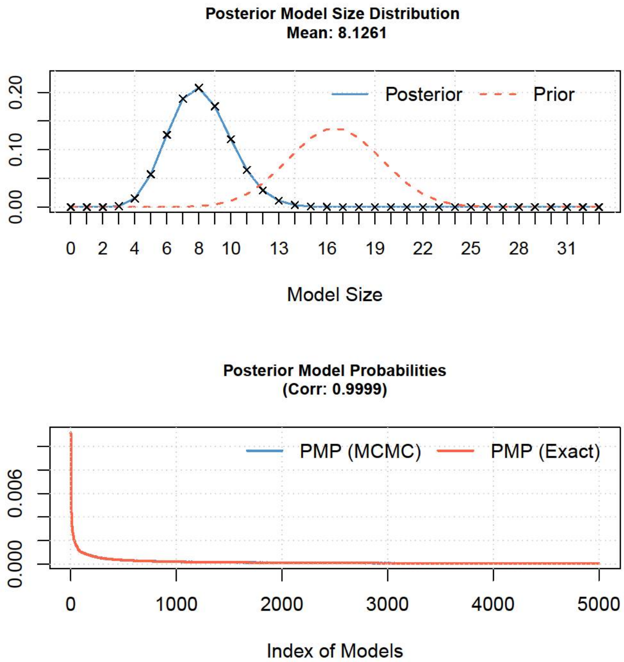


Table 3.10: Summary of BMA estimation (hump-shaped impulse responses)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
10.7143	$2 \cdot 10^8$	$1 \cdot 10^8$	12.15215 hours
No. models visited	*Modelspace*	*Visited*	*Topmodels*
104, 093, 439	$4.3 \cdot 10^9$	2.4%	16%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9997	100	uniform / 16	UIP
Shrinkage-Stats			
Av= 0.9901			

Notes: UIP = unit information prior, PMP = posterior model probability.

Figure 3.7: Model size and convergence (hump-shaped impulse responses)

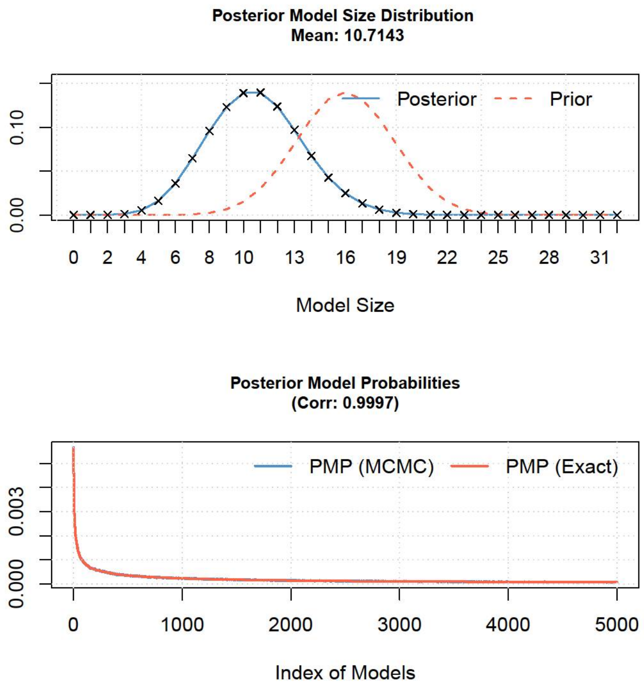
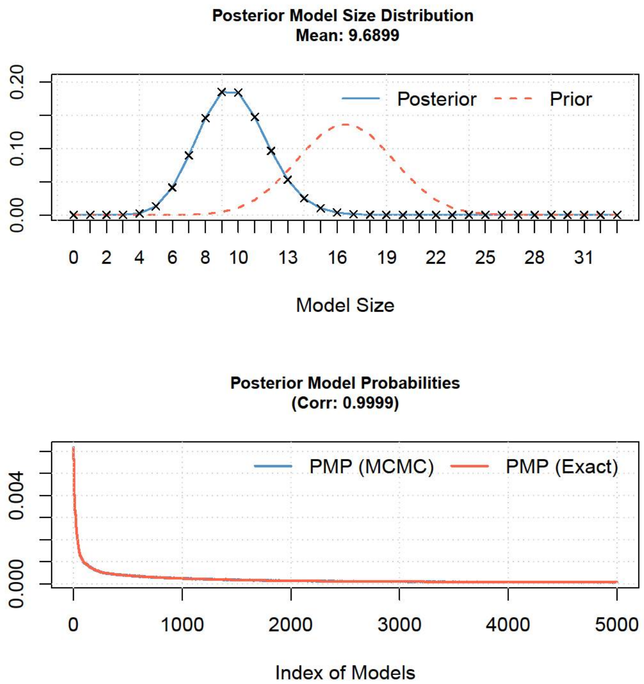


Table 3.11: Summary of BMA estimation (time to −0.1% decrease in prices)

Mean no. regressors 9.6899	*Draws* $2 \cdot 10^8$	*Burn-ins* $1 \cdot 10^8$	*Time* 12.0976 hours
No. models visited 87,125,827	*Modelspace* $8.6 \cdot 10^9$	*Visited* 1%	*Topmodels* 30%
Corr PMP 0.9999	*No. Obs.* 173	*Model Prior* uniform / 16.5	*g-Prior* UIP
Shrinkage-Stats Av= 0.9943			

Notes: UIP = unit information prior, PMP = posterior model probability.

Figure 3.8: Model size and convergence (time to −0.1% decrease in prices)



3.B Results of Censored Regression

Table B1: Censored regression, general model (all variables are included)

Response variable: transmission lag		
Country characteristics		
GDP per capita	-9.792*	(5.192)
GDP growth	1.512	(1.346)
Inflation	-17.41**	(8.695)
Financial dev.	22.17***	(6.084)
Openness	-11.16**	(5.595)
CB independence	30.20**	(12.27)
Data characteristics		
Monthly	-4.402	(6.920)
No. of observations	4.287	(5.186)
Average year	-0.168	(0.367)
Specification characteristics		
GDP deflator	5.102	(4.281)
Single regime	4.143	(3.497)
No. of lags	8.132*	(4.744)
Commodity prices	-1.284	(2.861)
Money	1.768	(2.949)
Foreign variables	4.102	(3.400)
Time trend	2.700	(5.791)
Seasonal	7.231*	(4.057)
No. of variables	1.352	(3.536)
Industrial prod.	-6.785*	(3.904)
Output gap	-10.41	(7.681)
Other measures	-6.246	(5.017)
Estimation characteristics		
BVAR	-1.147	(5.094)
FAVAR	14.53**	(6.525)
SVAR	-4.243	(3.008)
Sign restrictions	-3.270	(5.163)
Publication characteristics		
Price puzzle	3.651	(2.537)
Study citations	-0.717	(1.734)
Impact	-0.742	(0.699)
Central banker	5.313	(3.633)
Policymaker	9.024	(6.137)
Native	-1.996	(3.043)
Publication year	0.0475	(0.453)
Constant	62.32	(50.10)
Observations	198	

Notes: Standard errors in parentheses. Estimated by Tobit with the upper limit for transmission lags equal to 60 months. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Chapter 4

Habit Formation in Consumption: A Meta-Analysis

Abstract

We examine 597 estimates of habit formation reported in 81 published studies. The mean reported strength of habit formation equals 0.4, but the estimates vary widely both within and across studies. We use Bayesian and frequentist model averaging to assign a pattern to this variance while taking into account model uncertainty. Studies employing macro data report consistently larger estimates than micro studies: 0.6 vs. 0.1 on average. The difference remains 0.5 when we control for 30 factors that reflect the context in which researchers obtain their estimates, such as data frequency, geographical coverage, variable definition, estimation approach, and publication characteristics. We also find that evidence for habits strengthens when researchers use lower data frequencies, employ log-linear approximation of the Euler equation, and utilize open-economy DSGE models. Moreover, estimates of habits differ systematically across countries.

The paper was co-authored with Tomas Havranek and Anna Sokolova and published in the *European Economic Review* [2017, 95, pp. 142–167]. We gratefully acknowledge helpful suggestions by the editor Eric Leeper, an anonymous associate editor, and two anonymous referees. We thank Jan Babecky, Alexei Boulatov, Kamil Galuscak, Zuzana Irsova, Simona Malovana, Udara Peiris, Sergey Pekarski, Tom Stanley, Federica Teppa, Dimitrios Tsomocos, and Stefan Zimmermann for their comments on earlier versions of the manuscript. Havranek and Rusnak acknowledge support from the Czech Science Foundation (grant #P402/12/G097). Sokolova acknowledges support from the Czech Science Foundation (grant #15-02411S). An online appendix with data, code, and additional results is available at `meta-analysis.cz/habits`.

4.1 Introduction

The concept of habit formation in consumption is crucial for the explanation of various stylized facts in macroeconomics and finance. For example, in the asset pricing literature consumption habit helps reconcile the theory with the observed moments of asset returns. Mehra & Prescott (1985) show that the standard Lucas (1978) tree model fails to replicate the high equity premium and low risk-free rate at reasonable model parameters. Constantinides (1990) argues that habit formation solves this problem, as it can generate large variability in the marginal rate of substitution in consumption alongside smooth consumption growth—a feature that allows one to replicate high risk premium without having to rely on large risk aversion. Further refinements of the model suggested by Campbell & Cochrane (1999), Abel (1999), and Allais (2004) make it possible also to generate plausible variability in equity returns and the risk-free rate, while adding habits to a real business cycle framework helps explain the joint behavior of asset prices and consumption (Boldrin *et al.* 2001).

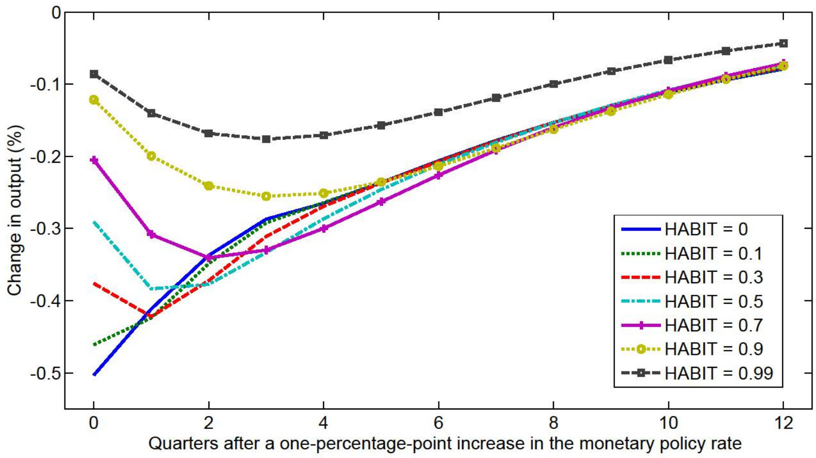
The presence of habit formation implies that past consumption choices affect current preferences. This notion violates the independence axiom used by Koopmans (1960) to derive the classic discounted utility model. Due to the growing popularity of models with habits, researchers have made efforts to develop theoretical underpinning for utility that is non-separable over time and features habit formation. Rozen (2010) lays out axiomatic foundation for a utility function displaying linear internal habits, describing a decision maker whose preferences depend on the history of past consumption choices. Rustichini & Siconolfi (2014) present a general axiomatic approach that allows for time-separable and non-separable utility as special cases, while He *et al.* (2013) put forward a model that incorporates habits as well as satiation in utility.[1]

Studies that feature general equilibrium models have come to rely on consumption habit as a means of replicating a delayed hump-shaped response of macro variables

[1]Habit formation can also be thought as a case in which an agent's consumption exhibits a form of hysteresis, in that his current consumption depend on his past consumption (Becker & Murphy 1988; Obstfeld 1992).

to policy shocks (Fuhrer 2000; Del Negro *et al.* 2007). This is because habit formation makes abrupt changes in consumption costly, thereby inducing smoothness in consumption dynamics (for a detailed discussion see Kano & Nason 2014). But the quantitative predictions of such models largely depend on the size of the parameter specifying the strength of habit formation. Figure 4.1 shows how the impulse response of output to a nominal interest rate shock changes in the popular model by Smets & Wouters (2007) when we assume different values of habit formation: the modeled behavior of the economy within one year after the shock depends heavily on the assumed strength of habits.[2]

Figure 4.1: The importance of habit formation for DSGE models



Notes: The figure shows simulated impulse responses of GDP to a one-percentage-point increase in the monetary policy rate. We use a calibrated version of the model developed by Smets & Wouters (2007) and vary the value of the habit formation parameter while leaving all other parameters calibrated at the posterior values from Smets & Wouters (2007). For the simulations we use Matlab code from the Macroeconomic Model Data Base (Wieland *et al.* 2012).

Dozens of papers have estimated the habit formation parameter, but their results vary widely. The variance can be partially attributed to differences in the data used in the estimation: some studies analyze Euler equations for aggregate consumption

[2]Figure 4.1 closely resembles Kano & Nason (2014, Figure 1) depicting the impulse-response function of consumption growth rate to a real interest rate shock for different values of the habits parameter. Remarkably, the result of Kano & Nason (2014) does not rely on the full New Keynesian DSGE model: they derive it using a log-linear approximation of the Euler equation and an AR(1) process for the real interest rate. Absent habit formation, the Euler equation sets consumption growth equal to the real interest rate, an AR(1); therefore, the impulse-response function peaks at 0 and decays afterward. With habit formation, current changes in consumption are associated with a utility loss in the future—in consequence, changes in the interest rate lead to a gradual adjustment of consumption growth.

(Fuhrer 2000; Carroll *et al.* 2011; Everaert & Pozzi 2014), some employ micro panel data sets (Dynan 2000; Collado & Browning 2007; Alessie & Teppa 2010), and others use DSGE models (Christiano *et al.* 2005; Smets & Wouters 2007), often employing prior values for the habit parameter. A brief look at the results of the seminal studies in each category suggests that the estimates are all over the place: Fuhrer (2000) shows that habit formation is crucial for his model to fit the data and obtains estimates that lie within the range 0.8–0.9. In contrast, Dynan (2000) uses panel household data and finds no evidence of habit formation. Christiano *et al.* (2005) estimate the same parameter using a DSGE model and obtain a value in the range 0.5–0.7.[3]

In this paper we investigate whether this diversity in the estimates of the habit parameter can be explained through differences in study designs used by researchers. We present what to our knowledge is the first quantitative synthesis—or a meta-analysis—of the evidence from the literature estimating habit formation. Meta-analyses attempt to trace variation in the estimates reported in the literature to differences in how the studies are conducted; it is the quantitative method of research review frequently used in medical research, which has recently become used by economists as well (Stanley 2001). In economics the method has been applied to a wide range of topics: the effect of the minimum wage on unemployment (Card & Krueger 1995), returns from education (Ashenfelter *et al.* 1999), the effect of distance on trade (Disdier & Head 2008), the intertemporal elasticity of substitution in labor supply (Chetty *et al.* 2011), and the impact of FDI on domestic firms' productivity (Havranek & Irsova 2011), among others.

We gather 81 published studies presenting estimates of habit formation and collect 31 aspects related to study design, such as the estimation techniques used, variable definition, data characteristics, geographical coverage, and model specification. We attempt to establish whether these aspects systematically affect the reported estimates of the habit parameter.

[3]Sampling uncertainty would seem to suggest that habit estimates of 0.5 and 0.9, for example, are not that far apart, but Figure 4.1 shows that the economic implications of consumption habit differ greatly across this range of estimates.

We cannot claim that our method allows us to explain variation in the true degree of habit formation; instead, we attempt to explain differences in its *estimates* reported in previous studies—a task that meta-analysis can accomplish. One obstacle that we face is the uncertainty over which of the 31 study characteristics should be included in the model approximating the process that generates habit estimates. To address this problem we employ Bayesian model averaging (BMA; Raftery *et al.* 1997; Moral-Benito 2015)—a method that estimates many regressions consisting of subsets of the potential explanatory variables and weights them by model fit and model complexity. As a robustness check we use frequentist model averaging, which does not rely on Bayesian methods.

Our results show that the difference between micro estimates (think Dynan 2000) and macro estimates (think Fuhrer 2000) remains large even after controlling for other aspects of study design. This finding resonates with Chetty *et al.* (2011), who report similar divergence between micro and macro estimates in the literature estimating the intertemporal elasticity of labor supply. Furthermore, the frequency of the data used in the estimation matters: estimates from studies employing monthly data tend to be substantially smaller than those obtained with lower frequencies, with the largest estimates being associated with the use of annual data. We also find that the use of second-order approximation of the Euler equation yields smaller estimates, which indicates that it is important to account for the precautionary saving motive when evaluating habit formation. Estimates obtained using US data tend to be larger than those reported for Japan, Europe, and other regions. Additionally, our results suggest that among the DSGE studies the ones that rely on the open-economy framework typically require higher degrees of habit formation to match the dynamics of the observables.

By contrast, we find that studies using the moments of asset returns do not report estimates that differ systematically from those obtained without the use of stock market data. In a similar vein, given the features of the data employed by the particular study, the use of the DSGE methodology itself does not result in estimates that are systematically different from those obtained by other methods. This finding suggests that reproducing empirical moments of the data within structural models requires roughly the same degree of habit formation as what would typically arise from reduced-form estimation with similar data sets.

We do not find evidence of systematic differences between the estimated magnitude of external habits ("keeping up with the Joneses") and internal habits (past own consumption decreases present utility) when other data and method characteristics are controlled for. The result is in line with Dennis (2009), who shows that the distinction between internal and external habits has a limited effect on the business cycle characteristics of New Keynesian models, and Kano & Nason (2014), who show in their online appendix that for log-linear approximation of the Euler equation under additive habits there is observational equivalence between external and internal specifications.[4] We also find that estimates of habits formed at the level of individual goods do not systematically differ from those of habits formed over the whole consumption bundle. Furthermore, studies using total non-durable consumption, food expenditures, or measures that include durable consumption come up with estimates that are roughly the same. However, we find that the use of simple panel techniques that do not rely on instrumental variables systematically affects the results. We also observe a correlation between the reported estimates and the characteristics of the journal where the study is published.

The remainder of the paper is structured as follows. Section 4.2 describes the approach we use to collect estimates of habit formation and presents the summary statistics for our data set. Section 4.3 tests for publication bias in the literature. Section 4.4 investigates the sources of heterogeneity in the estimated habit formation parameters. Section 4.5 concludes. Appendix A provides the correlation matrix of the variables used, shows diagnostics of the Bayesian model averaging exercise, and a provides a robustness check using an alternative set of priors. Appendix B discusses issues related to model uncertainty. Appendix C shows the list of studies included in the data set. An online appendix with data, code, and additional results is available

[4]The online appendix to Kano & Nason (2014) is available at `http://hermes-ir.lib.hit-u.ac.jp/rs/bitstream/10086/23297/1/070econDP12-08.pdf`.

at meta-analysis.cz/habits.

4.2 The Data Set of Habit Formation Estimates

4.2.1 Estimating the Degree of Habit Formation

Modeling habit formation usually involves the following utility function:

$$\sum_t \beta^t u(c_{i,t} - \gamma h_{i,t}), \tag{4.1}$$

where β is a discount factor, $u(\cdot)$ denotes the instantaneous utility function, $c_{i,t}$ is the consumption of individual i in period t, $h_{i,t}$ is the reference habit stock, and $\gamma \in [0,1)$ captures the strength of habit formation ($\gamma = 0$ gives the standard time-separable utility function). Papers that explore internal habits assume $h_{i,t} = c_{i,t-1}$: lagged own consumption decreases current utility. Under internal habits, therefore, utility is determined by consumption growth, not just the level of current consumption. Papers studying external habits ("catching up with the Joneses," Abel 1990) assume that utility is determined by the difference between the current consumption of an individual and the consumption of the corresponding reference group (for instance, the city where the consumer lives). External habits can be modeled by defining $h_{i,t} = \tilde{c}_{t-1}$, where $\tilde{c}_{t-1}$ denotes aggregate consumption in the preceding period. Several studies investigate "deep" habits formed at the level of individual goods rather than the whole consumption bundle (e.g., Ravn *et al.* 2006; Lubik & Teo 2014). Additionally, instead of using consumption directly, some papers use the variable "habit stock" defined by an autoregressive process (for example, Fuhrer 2000). Finally, a few studies model habits using a multiplicative rather than an additive specification; for example, Andrés *et al.* (2009) and Bjornland *et al.* (2011).

A common approach to estimating γ is to evaluate an approximation of the consumption Euler equation that incorporates habits. For example, with internal habit formation instantaneous utility depends on the household's past consumption; therefore, from the households' perspective, an increase in consumption today affects not only utility of the current period, but also future utility—by affecting future habit stock. The marginal effect on welfare of an increase in current consumption c_t is then given by

$$\lambda_t = u'(c_t - \gamma c_{t-1}) - \gamma\beta E_t u'(c_{t+1} - \gamma c_t). \tag{4.2}$$

Habit-forming households will internalize this effect when making consumption decisions; this is reflected in the first-order condition of the households' problem:

$$\lambda_t = \beta E_t \left\{ \frac{\lambda_{t+1} R_{t+1}}{1 + \pi_{t+1}} \right\}, \tag{4.3}$$

which relates expected marginal effects of changes in c_t and c_{t+1} to the nominal interest rate R_{t+1} and inflation π_{t+1}.

As shown in Kano & Nason (2014), assuming that utility is logarithmic and total factor productivity is driven by a random walk, Euler equation (4.3) can be approximated by

$$\Delta c_t = \phi_1 \Delta c_{t-1} + \frac{(\alpha^* - \beta\gamma)(\alpha^* - \gamma)}{\alpha^{*2}\phi_2} \sum_{j=0}^{\infty} \phi_2^{-j} E_t q_{t+j}, \tag{4.4}$$

where q_t is the demeaned real interest rate, α^* is a steady-state growth, $\phi_1 = \gamma\alpha^{*-1}$ and $\phi_2 = \alpha^*(\beta\gamma)^{-1}$ (see Kano & Nason 2014, Equation 1). Therefore, Euler equation (4.3) implies a relationship between current and past consumption growth, and a forward-looking component related to the expected discounted sum of future interest rates. Several studies derive their estimates of the habit parameter from a simplified version of approximation (4.4) that assumes a constant interest rate (e.g., Dynan 2000; Carroll *et al.* 2011; Sommer 2007). Furthermore, some studies employ higher-order approximations and account for the precautionary saving motive by including a measure of consumption risk in the estimated specification (e.g., Guariglia 2002).

Approximation (4.4) is derived from the problem faced by an individual household, and a number of studies obtain estimates of habit formation by using individual household data (for example, Dynan 2000; Guariglia 2002; Alessie & Teppa 2010). But micro studies often have data covering only short periods of time, and only on a fraction of consumption (such as food expenditures), and micro data are also often noisy, yielding imprecise estimates. Therefore, in practice similar specifications are often estimated on aggregate data. However, such treatment of the Euler equation may bias estimation results.

Attanasio & Weber (1993) point out that a correct aggregation of log-linear representation of the Euler equation necessitates examining a sum of the logarithms of the individual expenditures. Nevertheless, aggregate national accounts data only provide information on the sum of expenditures; researchers relying on aggregate data then take logarithms of the sum. Attanasio & Weber (1993) argue that if the cross-sectional distribution of expenditures varies over time, then the difference between the two measures will not be constant, and in fact is likely to be serially correlated. This induces serial correlation in the error term of log-specifications estimated on the aggregate data, a problem that could potentially lead to spurious results when estimating the habit parameter.

Another potential problem associated with the use of aggregate data stems from the fact that macro studies typically cannot account for households' taste shifters such as age, number of children, and employment status, all of which are likely to influence consumption decisions. Attanasio & Weber (1993) point out that when changes in these taste shifters do not cancel out after aggregating across the population, then the Euler equation cannot be consistently estimated on aggregate data. For example, it is well known that the consumption profile varies within the life cycle and depends on the individual's age (e.g., Attanasio & Weber 1995). If population composition with respect to young and old changes over time, then the aggregate specification that does not account for this effect will be prone to omitted variable bias. Attanasio & Weber (1993) also argue that aggregation bias may arise if households do not have full information about aggregate events: that would make aggregate instruments invalid for identifying household preference parameters. All these issues may bias estimates of the habit parameter obtained using aggregate data.

The voluminous macro literature that estimates habits is diverse, employing various data sets and approaches to estimation, as we discuss below. These papers estimate consumption habit while studying issues like sticky consumption growth (Carroll *et al.* 2011), habit persistence in current account data (Gruber 2004; Kano 2009), predictability of aggregate consumption growth (Everaert & Pozzi 2014), inflation dynamics (Fuhrer 2000), and moments of asset returns (Heaton 1995). Many estimates of the habit formation parameter come from dynamic stochastic general equilibrium models. Those estimates can be obtained by minimizing the distance between the model predictions and the empirical impulse response function (Christiano *et al.* 2005), by maximizing the likelihood of the state space representation of the model (Bouakez *et al.* 2005), or by using Bayesian methods (Smets & Wouters 2007).

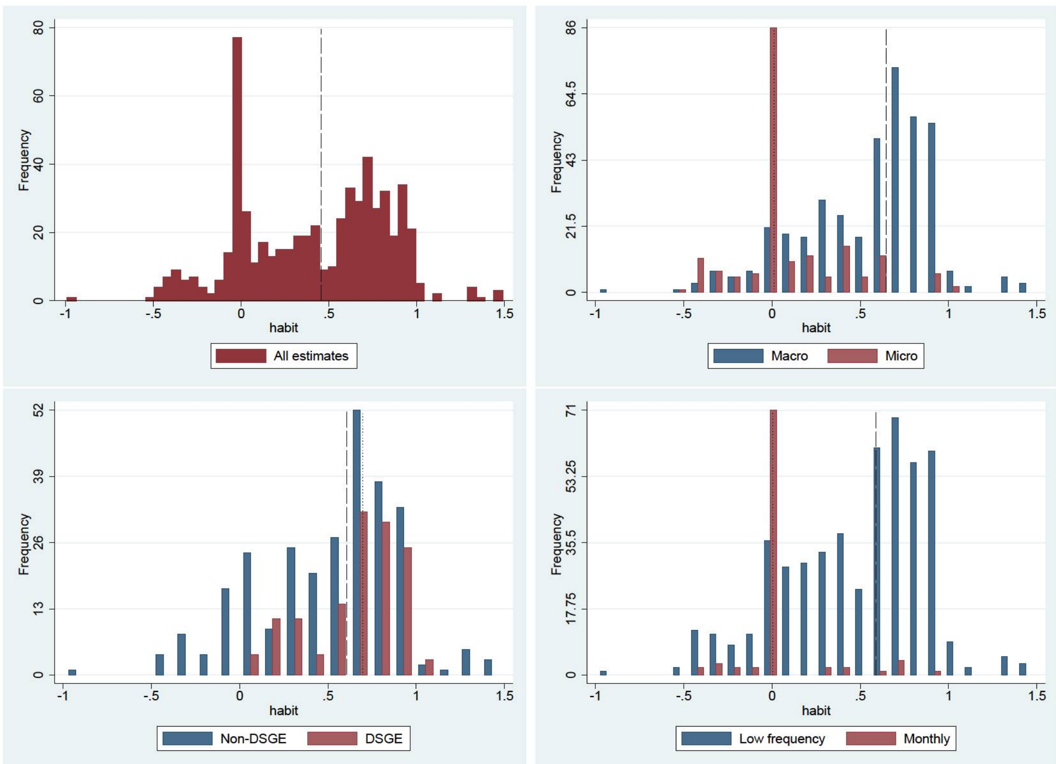
4.2.2 Collecting Estimates of γ

The first step of any meta-analysis is to gather the empirical studies on the topic, usually referred to as "primary studies." To collect primary studies, meta-analyses in economics often employ the RePEc or EconLit databases. We use Google Scholar because it provides powerful full-text search, whereas RePEc and EconLit only allow searching through abstracts and keywords related to the studies, thereby making it harder to devise an exhaustive search query. We first collect papers that contain the exact phrases "habit formation" or "habit persistence" and, at the same time, feature occurrences or synonyms of the following words: consumption, estimate, regression, and empirical. After reading the abstracts of the studies returned by our search query we download those that show any promise of containing empirical estimates of the habit formation parameter. In the next step we extend our search to the references of these studies and add the last study on March 1, 2016.

We apply the following three inclusion criteria. First, the study must provide an empirical estimate of the habit formation parameter. Second, the study must include an estimate of the standard error (or a statistic from which the standard error can be computed). Finally, the third inclusion criterion is that the study must be published in a peer-reviewed journal. Meta-analyses differ in their treatment of unpublished results—sometimes they include unpublished papers as well, especially when the resulting data set would otherwise be small. Since there are many published studies estimating the habit formation parameter, we prefer to focus on studies that have been subjected to a peer-review process. We find 81 studies that comply with our selection criteria, and we list them in Appendix C.

Each primary study typically reports several estimates, and the median number of estimates per study is four. It is hard to pin down each study's representative estimate, because the authors themselves rarely say explicitly which one they prefer. Therefore, we collect all estimates reported in each study. This approach results in an unbalanced data set, as some studies report many more estimates than others—nevertheless, it allows us to exploit the differences in data and method choices both within and across individual studies. Wherever possible, we include study fixed effects to filter out the effects of study-level characteristics that are otherwise unobservable. All studies combined provide us with 597 estimates of the habit formation parameter, and for each of them we collect 31 variables reflecting the context in which researchers obtain the estimates.

Figure 4.2: Estimates of habit formation vary widely



Notes: The figure shows histograms of the estimates of the habit formation parameter reported in individual studies. The top-left panel shows all estimates, the top-right panel splits the sample based on the level of aggregation in the data, the bottom-left panel examines a subsample of macro studies, splitting it based on whether the DSGE methodology is used, and the bottom-right panel splits the overall sample based on the frequency of the data. Dotted and dashed lines correspond to the respective medians.

The top-left panel of Figure 4.2 present a histogram of the estimated parameters, providing additional insights. First, the distribution of the estimates is far from normal, and both the lower and upper boundaries of the range 0–1, consistent with habit formation, seem to affect the probability of an estimate being reported. This result, however, may also reflect the constraints that researchers use in the process of estimation. Second, while not normal, the distribution of estimates is relatively symmetric, as both the lower and the upper tails are cut off, and the mean estimate virtually equals the median. Third, the histogram has multiple peaks, suggesting heterogeneity generated by different estimation methods.

To shed some light on the sources of heterogeneity, we split the sample of all estimates into subsamples, depending on whether the study uses household-level or aggregate data (the top-right panel, Figure 4.2). The histogram of micro estimates peaks at a much lower level of the habit parameter than that of macro estimates. Furthermore, neither distribution seems to be symmetrical: for micro studies, the right tail is heavier, while for macro studies the opposite holds. We further split the sample of macro studies, distinguishing between studies that estimate the habit parameter within DSGE models and those that use other techniques. The shapes of the histograms displayed on the bottom-left panel of Figure 4.2 are similar, suggesting that two groups of estimates may come from the same distribution. Finally, we investigate the role played by the frequency of the data by splitting the overall sample into estimates obtained using monthly data versus data using lower frequencies (the bottom-right panel). Data frequency seems to affect the estimates' distribution, and the use of monthly data is likely to result in lower estimates of the habit parameter.

We compute average and median values for different groups of estimates and display them in Table 4.1. The overall mean of the reported estimates is approximately 0.4. Studies using micro data deliver much smaller estimates on average—about 0.1. By contrast, macro studies tend to generate larger estimates: around 0.6. Among the macro approaches to assessing habit formation, DSGE studies tend to yield slightly larger estimates. The nature of the habit formation process seems to matter, too. Estimates of internal habit formation average 0.3, while estimates of external habits

Table 4.1: Habit formation estimates for different data and methods

	Unweighted				Weighted				
	Mean	Med.	5%	95%	Mean	Med.	5%	95%	N
All estimates	0.43	0.47	-0.32	0.97	0.55	0.62	-0.21	0.98	597
Micro studies	0.10	0.00	-0.39	0.62	0.12	0.08	-0.41	0.62	183
Macro studies	0.57	0.66	-0.11	0.98	0.62	0.69	0.00	0.99	414
Internal	0.28	0.15	-0.38	0.95	0.41	0.44	-0.34	0.96	369
External	0.66	0.67	0.16	1.00	0.72	0.71	0.16	1.48	228
Asset returns	0.43	0.62	-0.44	0.96	0.47	0.57	-0.29	0.96	87
Micro - internal	0.03	0.00	-0.40	0.60	0.09	0.01	-0.41	0.62	147
Micro - external	0.40	0.37	0.06	0.96	0.40	0.37	0.06	0.96	36
Macro - internal	0.45	0.61	-0.33	0.97	0.51	0.63	-0.22	0.98	222
Macro - external	0.70	0.71	0.21	1.00	0.73	0.73	0.19	1.48	192
Macro - non DSGE	0.52	0.62	-0.28	1.10	0.52	0.55	-0.28	1.10	279
Macro - DSGE	0.67	0.71	0.16	0.97	0.68	0.71	0.18	0.98	135

Notes: 5% and 95% denote the corresponding percentiles. Weighted = summary statistics based on the observations weighted by the inverse of the number of estimates reported per individual study. In such case each study receives the same weight in the computation of the summary statistics. Med. = Median, N = number of estimates

tend to be more than twice as large at around 0.7. The difference between estimates of external and internal habits remains substantial, albeit smaller, even when we calculate the means separately for macro and micro studies. For macro data, estimates of external habits are still larger—this finding seems to contradict the argument of Carroll *et al.* (1997), who suggest that estimates of external and internal habits are empirically indistinguishable when using macro data. We will revisit the difference between internal and external habits in Section 4.4, where we will control for other aspects of study design. The conclusions outlined above remain intact even when we weight the estimates by the inverse of the number of estimates reported in each study, thereby giving each study the same weight regardless of the number of estimates the study produces.

Most estimates in our data set are obtained using US data (63%). All studies combined provide results for 17 countries, arguably contributing to the heterogeneity we observe, but the number of countries is not large enough to connect the differences in estimates to the structural differences among the economies. Nevertheless, in Table 4.2 we compare group averages for the US, Japan, countries belonging to the EU, and the rest of the countries (other OECD economies, such as Australia,

Table 4.2: Habit formation differs across countries

	Unweighted				Weighted				
	Mean	Median	5%	95%	Mean	Median	5%	95%	N
All estimates									
US	0.42	0.40	-0.08	0.96	0.60	0.67	-0.04	1.00	377
EU countries	0.51	0.63	-0.27	1.00	0.48	0.61	-0.27	0.91	151
Japan	0.07	-0.23	-0.46	0.94	0.32	0.39	-0.41	0.96	27
Other countries	0.34	0.30	-0.03	0.78	0.36	0.31	-0.03	0.98	42
Micro estimates									
US	0.13	0.00	-0.09	0.63	0.15	0.08	-0.06	0.49	126
EU countries	0.10	0.07	-0.46	0.99	0.08	0.03	-0.46	0.62	36
Japan	-0.37	-0.39	-0.50	-0.23	-0.37	-0.39	-0.50	-0.23	14
Other countries	0.59	0.58	0.56	0.62	0.59	0.58	0.56	0.62	7
Macro estimates									
US	0.58	0.67	-0.26	0.98	0.66	0.72	0.00	1.00	251
EU countries	0.64	0.70	-0.08	1.12	0.60	0.69	0.07	0.91	115
Japan	0.55	0.64	0.02	0.96	0.50	0.39	0.09	0.96	13
Other countries	0.29	0.24	-0.04	0.93	0.30	0.21	-0.04	0.98	35

Notes: 5% and 95% denote the corresponding percentiles. Weighted = summary statistics based on the observations weighted by the inverse of the number of estimates reported per individual study. In such case each study receives the same weight in the computation of the summary statistics. N = number of estimates

Canada, New Zealand, and Korea) and notice several regularities. The estimates of habit formation for the US and EU tend to be larger on average than the estimates for Japan and other countries. At the same time, for macro studies the difference between Japan, the US, and the EU is smaller, while for micro data the highest estimates correspond to the group "other countries," which, however, only includes seven observations. It is not clear how to interpret these differences, as seeming cross-country diversity may be driven by differences among other features of the data sets, such as their length or frequency. Cross-country papers focusing on habit formation are rare, and the prominent study of this category, Carroll *et al.* (2011), finds homogeneous coefficients for a number of countries in our sample. Thus, we refrain from making any conclusions at this point, but will return to this issue in the more detailed analysis in the Section 4.4.

4.3 Publication Bias

The mean and median reported estimates may represent a biased reflection of the underlying research results if some estimates are more likely than others to be selected for publication. For this reason, most meta-analyses test—and, if necessary, correct—for so-called publication bias. Brodeur *et al.* (2016) collect 50,000 p-values reported in economics and document widespread publication bias. A recent survey among the members of the European Economic Association, Necker (2014), reveals that a third of economists in Europe admit that they have engaged in presenting empirical findings selectively so they confirm their arguments and in searching for control variables until they get a desired result. Ioannidis *et al.* (2017) survey meta-analyses conducted in economics and find that most fields suffer from the bias, as editors, referees, or authors themselves prefer statistically significant results that have an intuitive sign.

For example, Havranek (2015) finds strong publication bias in the literature that uses consumption Euler equations to estimate the elasticity of intertemporal substitution (often the same specification used to estimate habit formation).[5] Most economists believe that the elasticity of substitution should be positive because negative elasticity implies a convex utility function. Therefore, negative estimates of the elasticity are rarely reported in the literature, as are statistically insignificant estimates. The under-reporting of negative estimates and estimates that are positive but small and imprecise biases the means upward because it is not matched by corresponding under-reporting of large imprecise estimates.

The empirical literature on habit formation differs from studies estimating the elasticity of intertemporal substitution in two major aspects. First, negative estimates of the habit formation parameter allow for intuitive interpretation: although inconsistent with habit formation, they may result from durability of the consumption measure used in the estimation—and may thus be more publishable than negative estimates of the elasticity of intertemporal substitution. Second, unlike large esti-

[5] In a similar vein, Havranek & Sokolova (2016) identify publication bias in the literature on the excess sensitivity of consumption growth to anticipated income changes.

mates of the elasticity, estimates of the habit formation parameter that exceed 1 are implausible because they imply non-stationary consumption growth. Figure 4.2, discussed in Section 4.2, suggests that the most common estimates lie close to the midpoint between the lower and upper boundaries of the 0–1 interval (consistent with habit formation), and that when an estimate surpasses either limit, its probability of being reported drops—in other words, both very small and very large estimates can sometimes be discarded by the researchers. This relative symmetry in decision rules on discarding implausible estimates implies that even if there is publication selection in the literature on habit formation, it does not necessarily lead to publication bias.

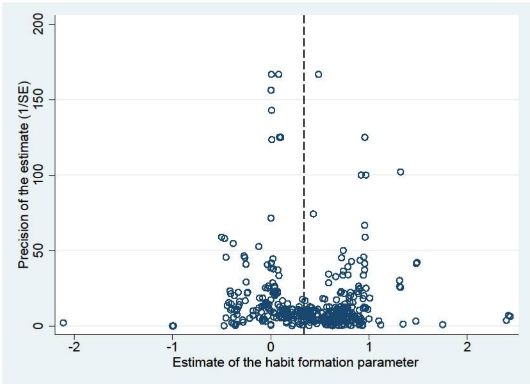
To test for potential publication bias researchers often evaluate so-called funnel plots (Egger *et al.* 1997). A funnel plot is a scatter plot of the estimates (on the horizontal axis) against the inverse of their standard errors, the estimates' precision (on the vertical axis). In the absence of publication bias the scatter plot forms an inverted funnel: the most precise estimates lie close to each other, while the less precise ones are more dispersed. The funnel plot should be symmetric because a feature of most estimation methods is that the ratio of the estimate to its standard error exhibits a symmetric distribution. Therefore all imprecise estimates, small and large, should have the same probability of being reported. If some estimates are reported less often because of their magnitude, the funnel will become asymmetric; if statistically insignificant estimates get under-reported, the funnel will become hollow.

The majority of the estimates in our sample are obtained via estimation methods presupposing that the ratio of the estimate to the corresponding standard error has a t-distribution. These methods do not place explicit constraints on the estimates that force them to lie between 0 and 1; therefore, the estimates can lie outside the $(0, 1)$ interval even if the underlying habit parameter lies within, given sufficient imprecision in estimation. Yet our sample also contains estimates from DSGE and other structural models that are often obtained under *a priori* restrictions. For example, some studies estimate DSGE models using maximum likelihood or minimum distance techniques, explicitly restricting the set of admissible values of the habit formation parameter to lie within the $[0, 1]$ range (see Bouakez *et al.* 2005). Other studies use

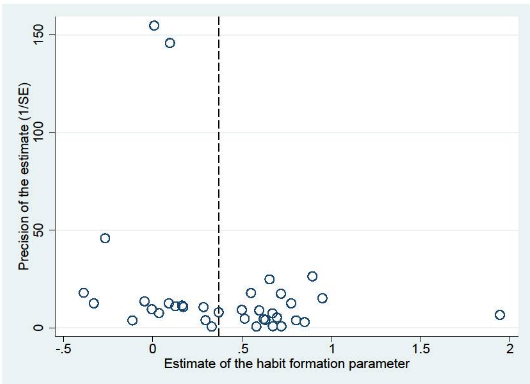
Bayesian techniques, but employ prior distributions for the habit parameter that are bounded to the $[0, 1]$ interval (e.g., Smets & Wouters 2003; Levine *et al.* 2012, who assume Beta distribution). Such structural estimates could generate spurious evidence for publication bias, so we exclude them from the analysis in this section.

Figure 4.3: Funnel plots suggest slight publication bias

(a) All non-restricted estimates



(b) Median non-restricted estimates from studies



Notes: In the absence of publication bias the funnel should be symmetrical around the most precise estimates of the habit formation parameter. The dashed vertical lines denote the mean of all the estimates in panel (a) and the mean of the median estimates reported in the studies in panel (b). Multiple peaks of the funnel suggest heterogeneity. For ease of exposition we exclude estimates with extreme precision values from the figure, but we use all the estimates in the statistical tests.

Figure 4.3 presents funnel plots for non-restricted estimates of the habit formation parameter. The left-hand panel depicts all estimates, while the right-hand panel plots median estimates reported in the studies against their precision. The plots show signs of asymmetry, and both 0 and 1 seem to be the boundaries that affect the probability of estimates being reported. The upper limit seems to be slightly more important than the lower one. An explanation of this result is that while negative estimates can be interpreted as evidence of durability, estimates larger than 1 are inconsistent with theory and are thus harder to justify. Researchers may consider these large estimates as evidence of model misspecification and adjust their models accordingly to produce more intuitive results.

Compared with funnel plots reported in other economic meta-analyses, however, the plot for habit estimates seems to be less skewed—thus the potential downward bias in this literature might be offset by the discarding of negative estimates. In what follows we test funnel asymmetry formally. We assess the extent of the bias and uncover the underlying mean estimate of habit formation. Our specification is based on Card & Krueger (1995) and Stanley (2008):

$$\widehat{HABIT}_{ij} = \alpha_0 + \delta \cdot SE(\widehat{HABIT}_{ij}) + \varepsilon_{ij}, \tag{4.5}$$

where $\widehat{HABIT}_{ij}$ is the i-th estimate from j-th study, $SE(\widehat{HABIT}_{ij})$ is the reported standard error of this estimate, and ε_{ij} is the disturbance term. As we have mentioned, most empirical methods estimating habit formation are based on the assumption that the ratio of the estimate to the standard error is t-distributed. This property implies that the numerator and the denominator of the ratio should be statistically independent quantities. Correlation between the two variables arises because of publication bias: suppose that researchers would only like to report estimates that are positive and statistically significant. Given the particular data and estimation technique (and thus given the standard error), they would need to search for a specification that delivers a point estimate of habit formation large enough to offset the standard error and show significance. Therefore, coefficient δ in regression (4.5), capturing the relation between estimates and their standard errors, indicates the magnitude of publication bias. α_0 is the mean estimate of the habit formation parameter conditional on the standard error approaching zero: it shows the mean reported habit formation parameter corrected for publication bias.

While several studies report very small standard errors, other studies report standard errors that are many orders of magnitude greater. To account for these outliers

Table 4.3: Funnel asymmetry tests indicate no publication bias

	Baseline	Instrument	Study	Precision	Median
SE (publication bias)	-0.222	-0.133	-0.214	0.174***	0.276
	(0.211)	(0.854)	(0.165)	(0.0315)	(0.207)
Constant (effect beyond bias)	0.397***	0.380**	0.444***	0.000679***	0.345***
	(0.0397)	(0.161)	(0.0405)	(0.0000417)	(0.0858)
Observations	462	462	462	462	38

Notes: The table presents the results of regression $\widehat{HABIT}_{ij} = \alpha_0 + \delta \cdot SE(\widehat{HABIT}_{ij}) + \varepsilon_{ij}$. $\widehat{HABIT}_{ij}$ and $SE(\widehat{HABIT}_{ij})$ are the i-th estimates of the habit formation parameter and their standard errors reported in the j-th studies. As in Figure 4.3, we only use non-restricted estimates. The standard errors of the regression parameters are clustered at study level. All estimations except for the last include study fixed effects. *Instruments*: we use the inverse of the square root of the number of observations in the individual study as an instrument for the standard error of the estimate of the habit formation parameter. *Study*: we weight the estimates by the inverse of the number of estimates reported in the study. *Precision*: we weight the estimates by the inverse of the reported estimate's standard error. *Median*: we estimate the equation by including the median estimate of the habit formation parameter and the median standard error of the estimated habit formation parameter reported in the individual studies.

we winsorize the data on standard errors at 5% on both sides of the distribution. Our main results are not sensitive to the choice of the fraction of data to be winsorized at each tail (as long as the largest outliers are discounted: winsorizing at 0.5% delivers largely similar results). The results are also robust to dropping the observations from the 5% tails on each side of the distribution.

Table 4.3 presents the results of regression (4.5) for non-restricted estimates; these results can also be interpreted as a test of funnel plot asymmetry. We consider several versions of the test. First, we estimate an OLS regression with study fixed effects and standard errors clustered at the study level. We include fixed effects to filter out unobservable study-specific factors that influence the reported estimates. Second, to address the potential endogeneity problem in meta-analysis we estimate the regression using the instrumental variable technique, while also including study fixed effects. Some method choices are likely to affect both the estimate and its standard error in the same direction, thus creating correlation between the disturbance term ε_{ij} and $SE(\widehat{HABIT}_{ij})$ and resulting in an inconsistent estimate of δ. As an instrument, we use the inverse of the square root of the number of observations used in each primary study: this variable is roughly proportional to the standard error, but not likely to be correlated with the method choice. Third, we estimate the regression by weighting each estimate by the inverse of the number of estimates reported in the corresponding study, thereby giving each study an equal weight in the regression. Fourth, we weight the estimates by their precision to remove heteroskedasticity. Finally, we exploit between- (instead of within-) study variation in the data using the median estimates and median standard errors reported in the primary studies.

The results can be summarized as follows. Four methods out of five yield insignificant estimates of δ (the magnitude of publication bias) and significant and large estimates of α_0 (the underlying mean habit formation parameter corrected for publication bias). We estimate the mean corrected habit formation to be around 0.4, close to the sample mean and median reported in the previous section. These results suggest that publication selection does not create a substantial bias in the reported habit formation parameters.

In contrast, the precision-weighted specification delivers a statistically significant estimate of publication bias and a much smaller underlying mean for habit formation. While precision-weighting removes heteroskedasticity, it is highly sensitive to small values of the standard error. Moreover, this specification yields a positive estimate of δ, suggesting an upward publication bias, which is at odds with the intuition suggested by Figure 4.3. According to the guidelines by Doucouliagos & Stanley (2013), the estimate of δ around 0.174 can be classified as "little to modest" publication bias, and would have to be more than five times larger to be classified differently. Finally, the results of the precision-weighted specification do not hold if we employ instrumental variable estimation, using the inverse of the square root of the number of observations as an instrument for the standard error (this specification is not reported). Therefore, we argue that precision-weighted estimation overstates the effect of publication bias.

To sum up, while we find some indications of publication selection related to the 0 and 1 thresholds that define the range consistent with habit formation, we find little evidence of any systematic bias resulting from this selection. Our findings suggest that the effects of potential selection against negative estimates and potential selection against estimates larger than 1 cancel each other out, rendering the mean estimate reported in the habit formation literature unbiased.

4.4 Why Do Estimates of Habit Formation Vary?

4.4.1 Explanatory Variables

We have noted that the estimates of habit formation differ substantially both within and between studies. In this section we attempt to relate the differences in the estimates to differences in the design of primary studies. To this end we collect 31 variables that reflect each study's data characteristics, geographical coverage, variable definitions, estimation technique, specification features (for studies estimating DSGE models), and publication characteristics (for example the number of citations). We cannot hope that these 31 variables will explain all differences across estimates—the set of potential explanatory variables is close to unlimited—but we believe that our selection reflects the most common choices faced by researchers who estimate habit formation.

Data characteristics For each study we collect the number of observations and the average year of the data used. We specify whether the study employs micro or aggregate data, as the discussion in Subsection 4.2.1 and the statistics in Subsection 4.2.2 suggest that this dimension may have crucial effect on the estimates. We also control for the frequency of the data: Bansal *et al.* (2012) argue that studies estimating consumption Euler equations should account for the difference between the econometrician's sampling frequency and consumers' decision frequency; the authors estimate the latter to be approximately monthly. Habit formation estimates are likely to be affected by the data frequency because at sufficiently high frequencies every consumption good displays durability, rendering the habit formation parameter negative: a full meal makes people saturated for the next few hours. Most studies employ quarterly data; for those using monthly and annual data we include controls.

Countries examined Although habit formation is supposed to be a so-called deep parameter, differences in structural characteristics of economies (such as culture) might cause the parameter to vary across countries. Havranek *et al.* (2015) find substantial cross-country heterogeneity in the elasticity of intertemporal substitution in consumption associated with cross-country differences in income and stock market participation. Since the number of countries investigated by the studies in our sample is small, we only use regional dummy variables instead of the underlying characteristics of the countries. We include dummies for US data, data on Japan, and data from countries that are members of the European Union. The remaining studies estimate the habit formation parameter for other non-European OECD countries.

Variable definitions In Section 4.2 we show that the mean reported estimates of internal and external habit formation differ. To see whether the difference holds after we control for other aspects of data and methodology, we create a dummy variable attributed to the type of habits under investigation. We also create a control for studies that investigate "deep" habits, for which the habit is formed over individual goods rather than the whole consumption basket. Such formulation has an important effect on the dynamics of DSGE models: it implies that habits affect not only the demand side of the economy, but also the supply side, as demand for individual goods becomes dependent on current sales, altering the optimal pricing behavior of firms and yielding countercyclical mark-ups of prices over marginal costs.

Estimates may also differ depending on the consumption good used in the estimation. Studies that include durable goods should obtain lower estimates of the habit formation parameter, while estimates based on food consumption may be biased if food is non-separable from other consumption goods (Attanasio & Weber 1995). We distinguish three categories of consumption proxies: food consumption, total non-durable consumption, and measures that include durable consumption; non-durable consumption represents our reference category. Finally, a prominent group of studies obtain habit formation parameters by scrutinizing asset pricing moments—we create a control signifying whether the study uses financial data other than capturing returns on government bonds or the economy-wide risk-free interest rate.

Estimation approach A common wisdom in empirical economics is that different estimation approaches often yield different results. We want to find out whether the use of a particular method is associated with systematic differences in the reported habit formation parameter. Most studies estimate habit formation by using reduced-form regression models. For such studies the most common method choice is GMM, although some assume homoskedasticity and employ TSLS. A few panel studies use fixed effects estimation that does not account for the Nickell (1981) bias, or random effects estimation, the assumptions of which are unlikely to hold in consumption Euler equations. Finally, a small fraction of studies estimate habit formation with OLS—we use this estimation approach as the reference group (using, for example, Bayesian estimation as the baseline does not affect the results).

Most of the regression-based estimates are obtained by first-order approximation of the Euler equation. This approach is criticized by Carroll (2001), who argues that terms of higher order are correlated with structural parameters and thus cannot be ignored. Some studies in our sample use second-order approximation, which allows the researchers to account for the precautionary saving motive, relating consumption growth to the degree of income uncertainty. Finally, some studies obtain the habit formation parameter by estimating dynamic stochastic general equilibrium models. These studies use maximum-likelihood-based methods, minimum distance estimators, or Bayesian techniques.

DSGE specification While most DSGE studies in our sample model closed economies, some extend their analysis to the open-economy framework, introducing exchange rate and current account fluctuations. All such models in our sample can be classified as small-open-economy models in the sense that they do not allow for feedback from domestic variables to foreign output, inflation, and the interest rate. An exogenous world interest rate in particular may play an important role in how the model identifies consumption habit because it pins down domestic interest rate movements. (Unfortunately it is not feasible to control for the different version of open-economy DSGE models; additional dummy variables focusing on these aspects would have very limited variation.) Another feature of the DSGE approach we control for is the set of observables used for estimation. Guerron-Quintana (2010) argues that using too few observables may lead to identification problems and biased impulse responses. Specifically, excluding consumption or the real wage from the set of observables may cause bimodality in the model's posterior and strongly affect the estimate of the habit parameter. The author compares model forecasting properties and impulse responses for different sets of observables and finds evidence in favor of the set that includes the seven observables used by Levin *et al.* (2005) and Smets & Wouters (2007).[6] Moreover, Adolfson *et al.* (2008) suggest that models estimated to match impulse responses to monetary policy shocks (e.g., Christiano *et al.* 2005) tend to deliver lower real friction parameters than models estimated to match all variation in the observables. We introduce a corresponding dummy variable to account for these specification characteristics.

Publication characteristics Finally, we control for the publication characteristics of individual studies. We include the year of publication to capture methodological advances that are otherwise hard to codify or that have not been employed by a sufficient number of studies yet. To account for approximate study quality beyond the observed differences in data and methodology, we include the number of citations, the recursive impact factor of the journal that published the study, and a dummy variable for studies published in top journals. We collect the data on the impact factor from RePEc: unlike other databases, RePEc covers virtually all economics journals and provides a discounted recursive impact factor well-suited for comparison of outlets in economics.

Table 4.4 describes the 31 explanatory variables mentioned above, listing their means, standard deviations, and means weighted by the inverse of the number of estimates reported in individual studies. The correlation matrix of all the collected explanatory variables is presented in Figure 4.6 in Appendix A; it shows that the variables reflect different aspects of the studies. A large correlation appears between micro data and the number of observations: micro-level studies tend to have more

[6] This set includes output, consumption, investment, real wages, total labor, interest rates, and inflation.

observations available than macro studies. Bayesian techniques are often employed within the framework of DSGE models, which renders the variable *Bayes* correlated with controls describing the set of observables. Furthermore, the DSGE models that do not include consumption in the set of observables tend to also omit wages, while studies that replicate responses to monetary policy shocks typically rely on minimum distance estimators. Finally, the positive correlation we observe between the year of publication of the study and the average year of data used in the study is intuitive.

Table 4.4: Description and summary statistics of regression variables

Variable	Description	Mean	Std. dev.	WM
Habit	The estimate of the habit formation parameter (response variable).	0.43	0.45	0.55
SE	The standard error of the estimate of the habit formation parameter.	0.16	0.26	0.15
Data characteristics				
No. of obs.	The logarithm of the number of observations.	6.14	1.81	5.48
Average year	The midpoint of the sample used for the estimation of habit formation (the base is the sample minimum: 1932).	53.30	11.73	52.52
Micro	= 1 if micro data are used for the estimation.	0.31	0.46	0.15
Monthly	= 1 if the frequency of the data used for the estimation is monthly.	0.15	0.36	0.06
Annual	= 1 if the frequency of the data used for the estimation is annual.	0.32	0.47	0.20
Countries examined				
US	= 1 if habit formation is estimated for the US.	0.63	0.48	0.68
EU	= 1 if habit formation is estimated for a country belonging to the EU.	0.25	0.44	0.21
Japan	= 1 if habit formation is estimated for Japan.	0.05	0.21	0.06
Variable definition				
External	= 1 if external habit formation is estimated.	0.38	0.49	0.45
Deep	= 1 if habits apply to individual goods.	0.05	0.21	0.04
Durable	= 1 if durable consumption goods are included in the measure of consumption.	0.74	0.44	0.76
Food	= 1 if food expenditures are used as a proxy for consumption.	0.12	0.32	0.07
Asset returns	= 1 if data on risky financial assets (e.g., stocks, house prices) are used.	0.15	0.35	0.16
Estimation approach				
GMM	= 1 if the general method of moments is employed for the estimation.	0.46	0.50	0.27
TSLS	= 1 if the two-step-least-squares method is employed for the estimation.	0.14	0.35	0.06
Panel	= 1 if a panel technique (fixed effects, random effects) is employed for the estimation.	0.05	0.23	0.02
Second-order approx.	= 1 if second-order approximation is employed.	0.05	0.21	0.08

Continued on next page

Table 4.4: Description and summary statistics of regression variables (continued)

Variable	Description	Mean	Std. dev.	WM
DSGE	= 1 if the estimation uses a dynamic stochastic general equilibrium model.	0.23	0.42	0.53
Bayes	= 1 if the estimation uses Bayesian inference.	0.20	0.40	0.42
Minimum distance	= 1 if the minimum distance method is employed for the estimation.	0.06	0.24	0.10
ML	= 1 if the maximum likelihood method is employed.	0.03	0.16	0.09
DSGE specification				
Open-economy DSGE	= 1 if the open-economy DSGE framework is employed.	0.02	0.13	0.05
Matching IR to mon. policy	= 1 if the study matches theoretical and empirical impulse responses to monetary policy shocks.	0.05	0.21	0.09
No. of observables in DSGE	The number of observables the study matches.	1.47	2.89	3.69
Seven observables from SW	= 1 if the list of observables includes proxies for output, consumption, investment, the wage, labor, the interest rate, and inflation.	0.06	0.23	0.16
No consumption	= 1 if the list of observables does not include consumption.	0.13	0.34	0.25
No wage	= 1 if the list of observables does not include real wages.	0.15	0.36	0.31
Publication characteristics				
Publication year	The year in which the study was published (base = 1991).	14.50	6.76	14.74
Citations	The logarithm of the mean number of Google Scholar citations received per year since the study was published (collected in May 2016).	0.54	0.33	0.62
Top journal	= 1 if the study was published in one of the top five journals in economics.	0.08	0.26	0.12
Impact	The recursive discounted RePEc impact factor of the outlet (collected in May 2016).	0.88	0.80	1.00

Notes: The variables are collected from published studies estimating the habit formation parameter. The following journals are considered top journals in economics: American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies. WM = mean weighted by the inverse of the number of estimates reported in a study.

4.4.2 Estimation and Results

Estimates of the habit parameter may vary both because of variation in the underlying degree of habit formation for different data sets (e.g., due to cultural differences across countries) and because of differences in estimation methods (e.g., due to differences in how the study approximates the Euler equation). In the previous subsection we pointed out 31 factors that in our view can contribute to explaining the heterogeneity among the estimates. A number of studies in our sample already explore the effects of some of these elements by conducting a series of experiments with data sets and methodologies. Researcher wishing to study variation across data sets can estimate the habit parameter on different data and compare results (e.g., Carroll *et al.* 2011 who make cross-country comparisons, or Ferson & Constantinides 1991 who compare results for data of different frequencies and definitions of consumption). To examine the consequences of using certain methodology, studies can compare results obtained by applying different methods to the same data (e.g., Guerron-Quintana 2010 who studies the effects of varying the set of observables in DSGE, or Korniotis 2010 who uses a log-linear approximation of the Euler equation that incorporates both internal and external habits to compare the two specifications; he also adds a term capturing consumption risk to the log-linear specification to check whether accounting for the precautionary saving motive alters habit estimates).

While the methodology outlined above could potentially shed light on some of the sources of heterogeneity, it also has major disadvantages. First, with this strategy it is impossible to address all 31 aspects of study design within the same framework. Habit parameter estimates can be obtained using data and methods that differ along many dimensions, some of which impact the estimates' distribution (e.g., see Figure 4.2, in which we compare the distribution of micro and macro estimates). This means that we would not be able to draw meaningful quantitative comparisons of the associated effects, unless we explicitly assumed that some of the 31 factors could be excluded from consideration without loss of generality. Second, this method would not address the variation *observed* in the literature, as doing so requires factoring in both quantitative effects associated with each aspect of study design as well as data describing the literature itself.

For example, studies that apply DSGE methodology seem to come up with estimates that are larger than the average. This may be because matching dynamics of the observables in DSGE models requires a degree of habit formation that is higher than that estimated from individual Euler equations—an observation that would make the two methods seem inconsistent with each other. An alternative explanation, however, would say that this is because studies that use macro data suffer from aggregation bias, and that includes studies estimating DSGE models. We cannot construct a DSGE study that would use micro data and account for households' taste shifters. We could potentially compare estimates obtained on the same macro data set by using DSGE and non-DSGE approaches, but it is not straightforward which DSGE specification should be used (e.g., open or closed economy), which observables should be matched, etc. Furthermore, it is not clear whether quantitative comparisons drawn from such exercise can be used to infer something about similar differences for different data.

In this paper we would like to make quantitative comparisons of the effects of all 31 explanatory variables on the estimates of the habit parameter, which is why instead of taking the approach outlined above we perform a meta-analysis. Rather than evaluating the degree of habit formation from consumption data while trying to fit all the different approaches and methodologies into a unified framework—a task that we deem impossible to accomplish—we focus on the *estimates* that were previously obtained within the literature and investigate their variation. We consider the following regression:

$$\hat{\gamma}_{ij} = \alpha_0 + \sum_{k=1}^{31} \theta_k Z_{k,ij} + \varepsilon_{ij}, \tag{4.6}$$

where $\hat{\gamma}_{ij}$ is an i-th estimate from a j-th study, and $Z_{k,ij}$ is a corresponding value of the k-th explanatory variable (introduced in the previous subsection). Model (4.6) is meant to approximate the process generating estimates of the habit parameter. Estimating (4.6) would not allow us to comment on the sources of variation in the habit parameter itself, but it would capture some of the variation in the habit parameter *estimates*, and allow for meaningful quantitative comparisons of the effects of choosing different study designs.

We believe that each variable in our set can contribute to explaining the heterogeneity among the estimates. But including all 31 variables in the regression would inflate the standard errors and yield inefficient estimates, because some of the variables are likely to prove redundant. The theory does not give us enough guidance to determine the exact subset of the 31 variables that should be included in the final regression. Sequential t-testing (sometimes called the "general-to-specific approach"), which is often used to decide which variables belong to the underlying model, is not statistically valid and gives rise to the possibility of excluding relevant variables. The large number of potential variables thus brings about problems related to model uncertainty that could result in severely erroneous inference. To address these issues, we employ the Bayesian model averaging technique (BMA)—a method that does not require selecting one individual specification.

Inference in BMA is based on a weighted average of individual regressions that include different combinations of explanatory variables; the weights reflect the posterior model probabilities (PMPs) of the corresponding individual specifications. PMPs can be thought of as a Bayesian analogy of information criteria used in frequentist econometrics (at least under certain assumptions, such as that model shocks are Gaussian). Researchers typically want to check the robustness of their results by estimating several regressions that include different combinations of explanatory variables; BMA generalizes this approach. Our intention here is to explain the basics of the BMA method and the terms needed for inference, not to give an exhaustive introduction to the BMA procedure; readers interested in such information should consult Koop (2003) for an introduction and Moral-Benito (2015) for a survey of BMA applications in economics. BMA have been used in meta-analysis, among others, by Havranek & Irsova (2017) and Irsova & Havranek (2013).

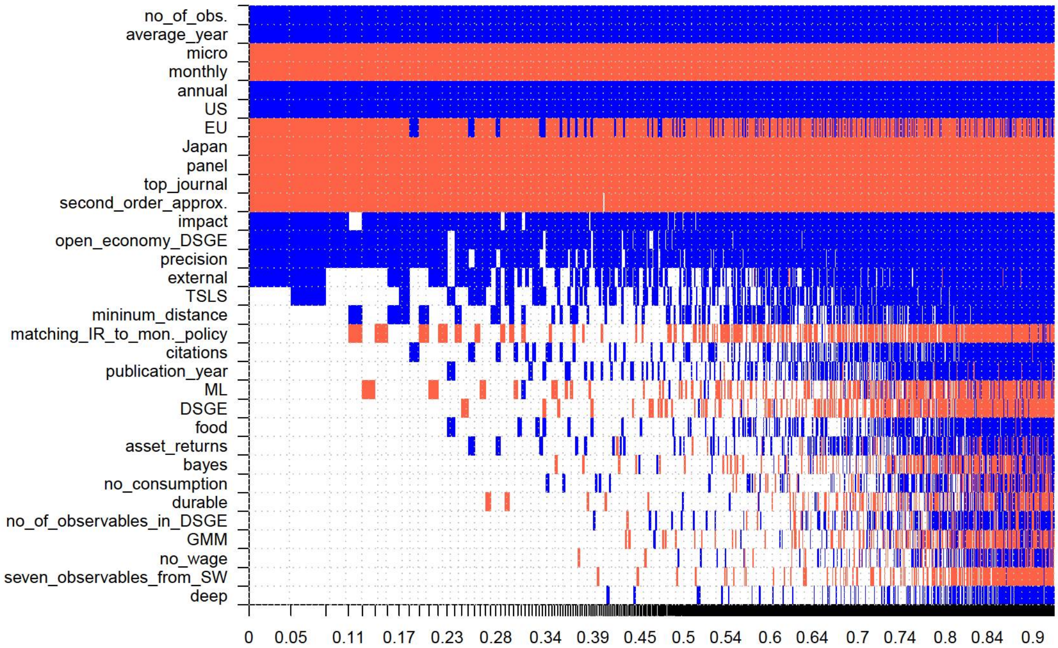
Estimating regression (4.6) means treating the estimates of habit as if they were observed data points. Nevertheless, each estimate is specific to the data set used in the estimation process and has a degree of uncertainty attached to it. Because of this feature our application of BMA departs from the standard approach: we explore uncertainty over which of the 31 elements should belong in auxiliary regression (4.6) describing the estimates, while leaving out uncertainty related to the structural models (e.g., the specification of log-linear Euler equation 4.4) that our primary studies choose to estimate (see Appendix B for further discussion).

We partially address some of the problems arising from treating estimates as data with the following strategy. First, we fix a subset of eight variables pertaining to data characteristics and geographical coverage, so that all eight variables appear in every regression estimated in the BMA exercise. In other words, we condition the estimates on the use of data of similar extent, age, aggregation, frequency, and regional coverage—any regression that fails to control for these factors is likely to suffer from omitted variable bias. Second, following the literature on estimated dependent variable models, we weight each observation by the precision of the estimates $[1/SE(\hat{\gamma}_{ij})]$, effectively giving more weight to estimates that are more precise.

All of the computations are performed using the R package `BMS`. Estimating all 2^{31-8} possible specifications is computationally too demanding—therefore, we approximate the whole model space by using the Model Composition Markov Chain Monte Carlo algorithm (Madigan & York 1995), which only traverses the most important part of the model space: that is, the models with high posterior model probabilities. Such a simplification is commonly applied in applications of BMA (see, for example, Feldkircher & Zeugner 2009). For the BMA estimation we also have to choose priors for the parameters and model space. We follow Eicher *et al.* (2011), who recommend using the unit information prior for the parameters and the uniform model prior for the model space because these priors perform well in predictive exercises. Our prior setting can be interpreted as follows: the unit information prior provides the same amount of information as one observation of data, while the uniform model prior means that each model has the same prior probability (thereby giving higher prior probabilities to medium model *sizes*).

Figure 4.4 presents the results of the BMA exercise. The variables are sorted from top to bottom by posterior inclusion probability (which can be thought of as a Bayesian analogy of statistical significance); the columns denote individual models. The color of the cell reflects the sign of the corresponding regression coefficient: negative signs are depicted in red (lighter in greyscale) and positive ones in blue (darker in greyscale); a white cell means that the variable is not included in the given model. The width of the columns is proportional to the posterior model probability (that is, how well the model fits the data relative to its size). Apart from the eight

Figure 4.4: Model inclusion in Bayesian model averaging



Notes: Response variable: the estimate of the habit formation parameter. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in greyscale) = the variable is included and the estimated sign is positive. Red color (lighter in greyscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures the cumulative posterior model probabilities; only the 5,000 models with the highest posterior probabilities are shown. Numerical results of the BMA exercise are reported in Table 4.5. A detailed description of all variables is available in Table 4.4.

variables we fix, the most important variables in explaining the heterogeneity among the estimates are *Panel*, *Top*, *Second-order approx.*, *Impact*, *Open*, and *External*. The regression signs for all of these variables are stable regardless of whether or not other control variables are included.

Table 4.5 presents the numerical results of Bayesian model averaging. In BMA the key statistic is the posterior inclusion probability (PIP), which reflects the importance of each variable. For a given variable the PIP is calculated by summing the posterior model probabilities of all models in which the variable is included. According to the rule of thumb proposed by Jeffreys (1961) and refined by Kass & Raftery (1995), the significance of each regressor is weak, positive, strong, or decisive if the PIP lies between 0.5–0.75, 0.75–0.95, 0.95–0.99, or 0.99–1, respectively. Additionally, we plot the posterior distribution of the estimated parameters corresponding to the first eight variables we fix, because for these variables the PIP is not informative: in the

Table 4.5: Explaining the differences in the estimates of habit formation

Response variable:	Bayesian model averaging		
Estimate of habit formation	Post. mean	Post. std. dev.	PIP
Precision	0.258	0.155	0.836
Data characteristics			
No. of obs.	0.000092	0.0002	1.000
Average year	0.006	0.002	1.000
Micro	-0.565	0.080	1.000
Monthly	-0.343	0.127	1.000
Annual	0.228	0.073	1.000
Countries examined			
US	0.199	0.048	1.000
EU	-0.002	0.022	1.000
Japan	-0.107	0.074	1.000
Variable definition			
External	0.093	0.101	0.538
Deep	0.001	0.007	0.041
Durable	-0.002	0.016	0.066
Food	0.016	0.053	0.120
Asset returns	0.013	0.054	0.119
Estimation approach			
GMM	-0.003	0.024	0.058
TSLS	0.080	0.124	0.353
Panel	-0.525	0.069	1.000
Second-order approx.	-0.385	0.118	0.982
DSGE	-0.018	0.060	0.136
Bayes	-0.004	0.025	0.076
Minimum distance	0.085	0.157	0.286
ML	-0.005	0.040	0.161
DSGE specification			
Open	0.226	0.110	0.893
Matching IR to mon. policy	-0.024	0.057	0.249
No. of observables	0.001	0.006	0.062
Seven observables from SW	-0.003	0.026	0.042
No consumption	0.002	0.017	0.067
No wage	0.001	0.013	0.044
Publication characteristics			
Publication year	0.002	0.005	0.164
Citations	0.018	0.047	0.174
Top journal	-0.571	0.119	1.000
Impact	0.125	0.061	0.923
Constant	-2.802	NA	1.000
Studies	81		
Observations	597		

Notes: PIP = posterior inclusion probability. More details on the BMA estimation are available in Table 4.7 and Figure 4.7.

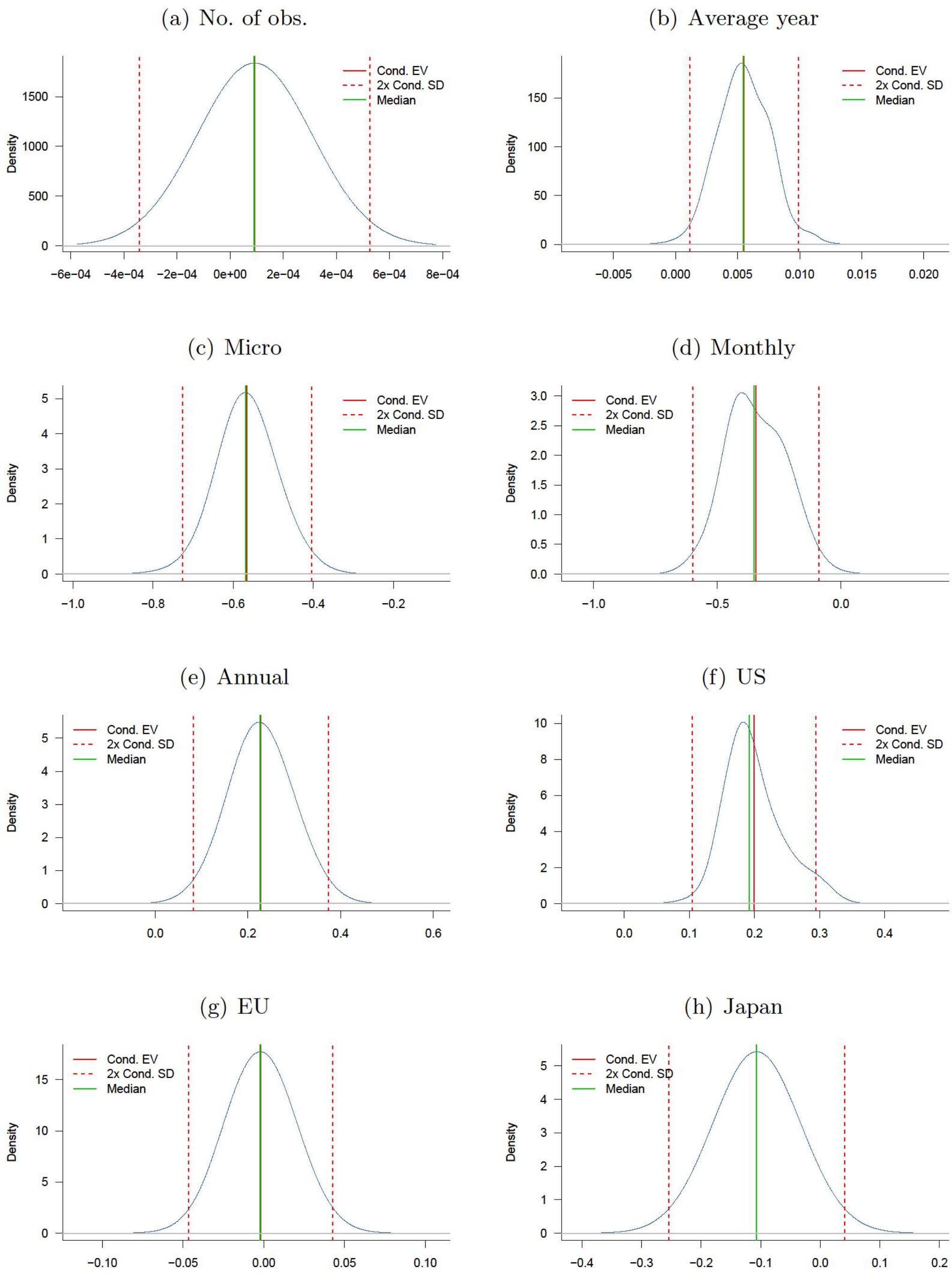
BMA exercise we force it to equal 1 by design (see Figure 4.5). We can see that the posterior means for the parameter estimates for all the variables except *Japan* and *EU* are more than two posterior standard deviations away from zero.

The level of data aggregation seems to be crucial for explaining the differences among estimates: micro data dramatically reduce estimates of habits (by more than 0.5), which corroborates the conclusion drawn from the histograms and summary statistics in Section 4.2. This resonates with the findings of Attanasio & Weber (1993) who argue that substituting national accounts data into log-linear Euler equations means incorrectly aggregating Euler equations of individual households, and that not accounting for taste shifters of individual households or household cohorts may make Euler equation estimation inconsistent (see discussion in Section 4.2).

Attanasio & Weber (1993) show that specifications that do not factor in these effects fail to pass the excess sensitivity test, delivering significant correlation between changes in consumption and predictable changes in income. Our results point toward a similar problem. If changes in demographic and labor market characteristics do not even out across the population (e.g., due to population aging, or because labor market participation follows the business cycle), then past changes in consumption may partially proxy for these omitted effects, resulting in a biased estimate of the habits parameter.

Another key factor is data frequency: the higher the frequency, the lower the estimate of habit formation, with the lowest estimates corresponding to monthly data. At high frequency substitution effects in consumption get more important, as some consumer goods become durable. For example, clothing expenditure will probably show durability at monthly frequency, but not at annual frequency. This notion is in line with the findings of Eichenbaum & Hansen (1990) and Dunn & Singleton (1986), who report evidence of such substitution at monthly frequencies, and Heaton (1995) and Allais (2004), who show that adding consumption substitution at nearby dates to an asset pricing model featuring habit formation improves model fit for moments of asset returns. Alternatively, the higher estimates reported for low frequencies may result from a bias introduced by time aggregation: Heaton (1995) points out that if

Figure 4.5: Posterior density of parameter estimates for fixed variables



Notes: The figure depicts the densities of the regression parameters corresponding to the set of data characteristics that we include in every evaluated model. The posterior means for the parameter estimates for all the variables except *Japan* and *EU* are more than two posterior standard deviations away from zero, which can be interpreted as analogous to statistical significance at the 5% level.

the decision frequency is higher than that of the data, time aggregation can induce positive autocorrelation in changes in consumption.

We find some evidence of country heterogeneity in the estimates of habit formation. The parameters estimated for the US tend to be 0.2 larger than those reported for other countries (and Japan in particular). To our knowledge, the only study that discusses cross-country differences in habit formation is Carroll *et al.* (2011), who find little heterogeneity across countries, but do not consider Japan. The cross-country differences in habit formation might reflect cultural differences—nevertheless, the specifics of the data may play a role, too. For instance, Carroll *et al.* (2011) mention several problems with Japanese data on consumption related to adjustments in the Japanese national accounts methodology.

Furthermore, we find that some estimation techniques deliver results systematically different from those obtained via other methods. The use of simple panel data techniques such as fixed effect method results in estimates that are substantially smaller. On the one hand, such methods can take into account heterogeneity between individuals beyond that captured by observed taste shifters. On the other hand, they may be prone to Nickell (1981) bias resulting from not taking into account the endogeneity created by including a lagged value of the dependent variable among the explanatory variables. Our result corroborates observations made by Naik & Moore (1996) who document that the use of fixed effects reduces estimates of the habits parameter. As noted before, studies that employ first-order approximation of the Euler equation cannot account for the precautionary saving motive, in the presence of which growth in consumption depends positively on the degree of consumption risk, as households postpone consumption when faced with uncertainty. This feature may be important for estimating habits: if consumption uncertainty is correlated with lagged consumption growth, then first-order approximation will bias the estimate of the habit formation parameter because lagged consumption growth would partially proxy for precautionary saving. We find support for this conjecture, as the use of second-order approximation tends to reduce the estimate of habit formation by about 0.4.

The BMA exercise suggests that the specification of habit as external slightly increases the estimated parameter (by about 0.1), even though the PIP is weak according to the classification by Kass & Raftery (1995). This contradicts our observation based on Table 4.1 that estimates of external habits are 0.4 higher on average. Nevertheless, the contradiction can be explained by three observations. First, micro studies use internal habits about four times more often than macro studies, as shown in Table 4.1. This feature is likely to increase the average difference between external and internal specifications, as micro studies deliver lower estimates regardless of the method used. Second, all 100 estimates obtained from monthly data pertain to internal habits, which also plays a role, as high-frequency data deliver lower estimates. Third, 26 out of the 28 estimates obtained via second-order approximation employ internal habits, which has a similar downward effect on the average internal habits parameter.

It is well known that to replicate certain empirical facts (i.e., the response of consumption and output to a monetary policy shock) DSGE models require a high degree of habit persistence. Therefore, it is reasonable to expect the DSGE methodology to deliver higher estimates of the habit formation parameter. At the same time, DSGE studies use exclusively macro data, which are prone to aggregation bias. Furthermore, none of the DSGE studies in our sample employ data at monthly frequency. We find that the reason for the higher average DSGE estimates is most likely the fact that DSGE studies use aggregate data of low frequencies, not the DSGE methodology itself. This result is supported by histograms in Figure 4.2 indicating that estimates obtained within DSGE models seem to belong to the same distribution as other macro estimates. Our finding echoes that of Kano & Nason (2014), who point out the resemblance between the impulse response functions obtained within DSGE models that include consumption habits and those generated using the log-linear approximation of the Euler equation (4.4) on its own. At the same time, among the DSGE models, those featuring open economies tend to deliver estimates of habit formation that are about 0.2 higher. This result corroborates the observation made by Adolfson *et al.* (2008), who compare open- and closed-economy estimates of habit formation and find that habits tend to show stronger in open-economy models. Moreover, our results suggest that when other aspects of the data are controlled for, studies scrutinizing moments of asset returns report estimates that are close to those found in the rest of the literature.

We perform a robustness check using an alternative prior setup, employing the benchmark g-priors for the parameters suggested by Fernandez *et al.* (2001) along with the beta-binomial model prior for the model space, which gives each model size equal prior probability (Ley & Steel 2009). The results, reported in Table 4.8, are very similar to the baseline specification, with one notable exception: the posterior inclusion probability pertaining to *External* drops below 0.5, rendering this variable ineffective in explaining any variation among the reported estimates of habit formation.

4.4.3 Frequentist Model Averaging

We have stressed earlier that our dependent variable (habit parameters reported in previous studies) is estimated, which gives rise to conceptual problems for the BMA technique most commonly used in model averaging exercises. We have tried to address this issue in three ways: by including the data characteristics to all models estimated by BMA, by using precision of the estimates as weights, and by discussing the potential implications of this problem for our results (see Appendix B). An alternative approach is to employ a frequentist method of model averaging and for individual regressions utilize the standard technique of the literature on estimated dependent variable models.

The intuition of frequentist model averaging is analogous to that of BMA discussed in detail earlier: many models featuring different combinations of explanatory variables are estimated and weighted according to their goodness of fit and parsimony. The dominance of BMA in model averaging applications is given by the computational ease of Bayesian relative to frequentist methods in this field. For example, we are not aware of any previous meta-analysis that would employ frequentist model averaging. Many studies, especially in the literature on growth determinants, use combinations of Bayesian and frequentist approaches (for example, Sala-I-Martin *et al.* 2004). The few studies that rely on purely frequentist techniques typically use information criteria as weights. Nevertheless, Hansen (2007) shows that weights selected by minimizing the Mallows criterion (an estimate of the average squared error from the model average fit) are asymptotically optimal. Another problem is how to simplify the model space: it would take us several months to estimate all the 2^{32} models, and we cannot use the Model Composition Markov Chain Monte Carlo algorithm that helped us in the case of BMA. We therefore follow the approach suggested by Amini & Parmeter (2012), who build on the pioneering insight of Magnus *et al.* (2010) and use orthogonalization of the covariate space, thus reducing the number of models that need to be estimated from 2^{32} to 32. In individual regressions we use inverse-variance weights to account for the estimated dependent variable issue.

The results of frequentist model averaging are shown in Table 4.6 and are broadly similar to that of BMA. It is worth noting at this point that the standard errors displayed in the table are approximate and probably conservative (Amini & Parmeter 2012), since a formal asymptotic theory for Mallows model averaging is still to be developed. We can see from the table that, even using the frequentist approach, micro estimates are found to be substantially smaller than macro estimates on average (the difference is even larger than what BMA suggests). Next, the frequency of the data matters, as studies with annual data tend to find substantially more evidence for consumption habit. Habit formation is stronger for the US than for other countries, which is also consistent with the BMA evidence. Once again we find no significant difference between the estimates of internal and external habit once other aspects of data and methodology are controlled for. Simple panel data techniques bring systematically smaller estimates of consumption habit, which might be caused by the Nickell (1981) bias. Open-economy DSGE models are associated with larger habit estimates, and the top journals in economics tend to report, ceteris paribus, weaker evidence for habits compared to other outlets.

4.5 Concluding Remarks

In this paper we collect and examine estimates of the habit formation parameter previously reported in the literature. We document that the mean value of the parameter is 0.4 overall, but that it differs greatly between micro studies (0.1) and macro studies (0.6). None of these values is large enough to explain some of the best-known empirical puzzles in macroeconomics and finance: for example, Constantinides (1990) shows that to account for the equity premium puzzle the habit formation parameter must exceed 0.8, while Fuhrer (2000) reproduces a humped-shaped response of consumption to various shocks with values of habit formation in the range 0.8–0.9. We find that the mean habit formation parameter produced by studies that estimate DSGE models is close to 0.7, which seems to corroborate the notion that structural estimation requires a high degree of habit formation. Nevertheless, when we turn to a more detailed investigation and control for the context in which researchers obtain their estimates, we get alternative explanations for the large habit formation reported by DSGE studies.

We show that the specifics of the data have a crucial impact on the estimated consumption habit. The difference between the results of micro and macro studies remains large when 30 other aspects of study design are controlled for. The distinction arises because micro and macro studies focus on different sources of variation in consumption: micro studies exploit variation at the level of individual households, but often lack information on consumption patterns over longer time horizons (and typically only use a fraction of consumption, such as food expenditures). By contrast, macro studies make use of consumption variation over time, while neglecting demographic characteristics and taste shifters. Our results also suggest that the frequency of the data matters—estimates obtained employing monthly frequency tend to be substantially smaller than when quarterly and annual frequencies are used. This finding may be due to the fact that at higher frequencies more consumption goods are likely to display durability, or may arise because of the time aggregation problem widely recognized in the asset pricing literature (e.g., Heaton 1995).

We also find evidence indicating the importance of the order of approximation of the Euler equation: the use of second-order approximation tends to reduce the estimate of consumption habit. This result may signify that the precautionary saving motive plays an important role in the behavior of consumers. By contrast, we find that the use of the DSGE methodology *per se* (when other aspects of study design are controlled for) does not necessarily yield higher estimates of habits. Thus, a part of the explanation of why many estimated structural models require high degrees of habit formation may lie in their use of aggregate and low-frequency data. Additionally, because such studies typically rely on log-linearized specifications, they might be subject to the omitted variable bias, as high estimates of habits may partially capture the precautionary saving motive we have mentioned. Similarly, we find that, everything else being held equal, studies focusing on moments of asset returns deliver habit parameters that are roughly the same as those reported by other studies. We also show that estimates reported in DSGE studies are affected by model specification: in line with Adolfson *et al.* (2008), our results indicate that open-economy models tend to report higher estimates of habit formation than closed-economy models. Finally, unlike Carroll *et al.* (2011), we find cross-country heterogeneity in habit formation, with the US displaying stronger habit formation than other countries.

References

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching Up with the Joneses." *American Economic Review* **80(2)**: pp. 38–42.

Abel, A. B. (1999): "Risk premia and term premia in general equilibrium." *Journal of Monetary Economics* **43(1)**: pp. 3–33.

Adolfson, M., S. Laseen, J. Linde, & M. Villani (2007): "Bayesian estimation of an open economy DSGE model with incomplete pass-through." *Journal of International Economics* **72(2)**: pp. 481–511.

Adolfson, M., S. Laséen, J. Lindé, & M. Villani (2008): "Empirical Properties

Of Closed- And Open-Economy Dsge Models Of The Euro Area." *Macroeconomic Dynamics* **12(S1)**: pp. 2–19.

Alessie, R. & A. Kapteyn (1991): "Habit Formation, Interdependent References and Demographic Effects in the Almost Ideal Demand System." *Economic Journal* **101(406)**: pp. 404–19.

Alessie, R. & F. Teppa (2010): "Saving and habit formation: evidence from Dutch panel data." *Empirical Economics* **38(2)**: pp. 385–407.

Allais, O. (2004): "Local Substitution and Habit Persistence: Matching the Moments of the Equity Premium and the Risk-Free Rate." *Review of Economic Dynamics* **7(2)**: pp. 265–296.

Altig, D., L. Christiano, M. Eichenbaum, & J. Linde (2011): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle." *Review of Economic Dynamics* **14(2)**: pp. 225–247.

Amini, S. M. & C. F. Parmeter (2012): "Comparison Of Model Averaging Techniques: Assessing Growth Determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

Andreasen, M. M. (2012): "An estimated DSGE model: Explaining variation in nominal term premia, real term premia, and inflation risk premia." *European Economic Review* **56(8)**: pp. 1656–1674.

Andrés, J., J. David López-Salido, & E. Nelson (2009): "Money and the natural rate of interest: Structural estimates for the United States and the euro area." *Journal of Economic Dynamics and Control* **33(3)**: pp. 758–776.

Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* **6(4)**: pp. 453–470.

Attanasio, O. P. & G. Weber (1993): "Consumption Growth, the Interest Rate and Aggregation." *Review of Economic Studies* **60(3)**: pp. 631–49.

Attanasio, O. P. & G. Weber (1995): "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* **103(6)**: pp. 1121–57.

Auray, S. & P. Feve (2008): "On Sunspots, Habits and Monetary Facts." *Macroeconomic Dynamics* **12(01)**: pp. 72–96.

Baltagi, B. H., G. Bresson, J. M. Griffin, & A. Pirotte (2003): "Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from French regional gasoline consumption." *Empirical Economics* **28(4)**: pp. 795–811.

Bansal, R., D. Kiku, & A. Yaron (2012): "Risks For the Long Run: Estimation with Time Aggregation." *NBER Working Papers 18305.*

Bartolomeo, G. D., L. Rossi, & M. Tancioni (2010): "Monetary policy, rule-of-thumb consumers and external habits: a G7 comparison." *Applied Economics* **43(21)**: pp. 2721–2738.

Batini, N., A. Justiniano, P. Levine, & J. Pearlman (2006): "Robust inflation-forecast-based rules to shield against indeterminacy." *Journal of Economic Dynamics and Control* **30(9-10)**: pp. 1491–1526.

Becker, G. S. & K. M. Murphy (1988): "A Theory of Rational Addiction." *Journal of Political Economy* **96(4)**: pp. 675–700.

Bekaert, G., S. Cho, & A. Moreno (2010): "New Keynesian Macroeconomics and the Term Structure." *Journal of Money, Credit and Banking* **42(1)**: pp. 33–62.

Bjornland, H., K. Leitemo, & J. Maih (2011): "Estimating the natural rates in a simple New Keynesian framework." *Empirical Economics* **40(3)**: pp. 755–777.

Boivin, J. & M. P. Giannoni (2006): "Has Monetary Policy Become More Effective?" *The Review of Economics and Statistics* **88(3)**: pp. 445–462.

Boldrin, M., L. J. Christiano, & J. D. M. Fisher (2001): "Habit Persistence, Asset Returns, and the Business Cycle." *American Economic Review* **91(1)**: pp. 149–166.

BOUAKEZ, H., E. CARDIA, & F. J. RUGE-MURCIA (2005): "Habit formation and the persistence of monetary shocks." *Journal of Monetary Economics* **52(6)**: pp. 1073–1088.

BOVER, O. (1991): "Relaxing Intertemporal Separability: A Rational Habits Model of Labor Supply Estimated from Panel Data." *Journal of Labor Economics* **9(1)**: pp. 85–100.

BRAUN, P. A., G. M. CONSTANTINIDES, & W. E. FERSON (1993): "Time nonseparability in aggregate consumption: International evidence." *European Economic Review* **37(5)**: pp. 897–920.

BRODEUR, A., M. LE, M. SANGNIER, & Y. ZYLBERBERG (2016): "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.

CAMPBELL, J. Y. & J. COCHRANE (1999): "Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* **107(2)**: pp. 205–251.

CAMPBELL, J. Y. & N. G. MANKIW (1991): "The response of consumption to income: A cross-country investigation." *European Economic Review* **35(4)**: pp. 723–756.

CARD, D. & A. B. KRUEGER (1995): "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review* **85(2)**: pp. 238–43.

CARROLL, C. D. (2001): "Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation)." *Advances in Macroeconomics* **1(1)**: pp. 1–38.

CARROLL, C. D., J. OVERLAND, & D. N. WEIL (1997): "Comparison Utility in a Growth Model." *Journal of Economic Growth* **2(4)**: pp. 339–67.

CARROLL, C. D., J. SLACALEK, & M. SOMMER (2011): "International Evidence on Sticky Consumption Growth." *The Review of Economics and Statistics* **93(4)**: pp. 1135–1145.

Castelnuovo, E. & S. Nistico (2010): "Stock market conditions and monetary policy in a DSGE model for the U.S." *Journal of Economic Dynamics and Control* **34(9)**: pp. 1700–1731.

Chetty, R., A. Guren, D. Manoli, & A. Weber (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* **101(3)**: pp. 471–75.

Chib, S. & S. Ramamurthy (2010): "Tailored randomized block MCMC methods with application to DSGE models." *Journal of Econometrics* **155(1)**: pp. 19–38.

Christiano, L. J., M. Eichenbaum, & C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* **113(1)**: pp. 1–45.

Christoffel, K., K. Kuester, & T. Linzert (2009): "The role of labor markets for euro area monetary policy." *European Economic Review* **53(8)**: pp. 908–936.

Collado, M. D. & M. Browning (2007): "Habits and heterogeneity in demands: a panel data analysis." *Journal of Applied Econometrics* **22(3)**: pp. 625–640.

Constantinides, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy* **98(3)**: pp. 519–43.

Cooley, T. F. & M. Ogaki (1996): "A Time Series Analysis of Real Wages, Consumption, and Asset Returns." *Journal of Applied Econometrics* **11(2)**: pp. 119–34.

De Graeve, F. (2008): "The external finance premium and the macroeconomy: US post-WWII evidence." *Journal of Economic Dynamics and Control* **32(11)**: pp. 3415–3440.

Del Negro, M. & F. Schorfheide (2008): "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)." *Journal of Monetary Economics* **55(7)**: pp. 1191–1208.

Del Negro, M., F. Schorfheide, F. Smets, & R. Wouters (2007): "On the Fit of New Keynesian Models." *Journal of Business & Economic Statistics* **25**: pp. 123–143.

Dennis, R. (2009): "Consumption Habits in a New Keynesian Business Cycle Model." *Journal of Money, Credit and Banking* **41(5)**: pp. 1015–1030.

Disdier, A.-C. & K. Head (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics* **90(1)**: pp. 37–48.

Doucouliagos, C. & T. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition And Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Dunn, K. B. & K. J. Singleton (1986): "Modeling the term structure of interest rates under non-separable utility and durability of goods." *Journal of Financial Economics* **17(1)**: pp. 27–55.

Dynan, K. E. (2000): "Habit Formation in Consumer Preferences: Evidence from Panel Data." *American Economic Review* **90(3)**: pp. 391–406.

Edge, R. M., M. T. Kiley, & J.-P. Laforte (2008): "Natural rate measures in an estimated DSGE model of the U.S. economy." *Journal of Economic Dynamics and Control* **32(8)**: pp. 2512–2535.

Egger, M., G. Smith, M. Schneider, & C. Minder (1997): "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* **316(7129)**: pp. 469–471.

Eichenbaum, M. & L. P. Hansen (1990): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data." *Journal of Business & Economic Statistics* **8(1)**: pp. 53–69.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Everaert, G. & L. Pozzi (2014): "The Predictability Of Aggregate Consumption Growth In OECD Countries: A Panel Data Analysis." *Journal of Applied Econometrics* **29(3)**: pp. 431–453.

Feldkircher, M. & S. Zeugner (2009): "Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging." *IMF Working Papers 09/202*.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

Fernández-Villaverde, J. (2010): "The econometrics of DSGE models." *SERIEs - Journal of the Spanish Economic Association* **1(1)**: pp. 3–49.

Ferson, W. E. & G. M. Constantinides (1991): "Habit persistence and durability in aggregate consumption: Empirical tests." *Journal of Financial Economics* **29(2)**: pp. 199–240.

Ferson, W. E. & C. R. Harvey (1992): " Seasonality and Consumption-Based Asset Pricing." *Journal of Finance* **47(2)**: pp. 511–52.

Flavin, M. & S. Nakagawa (2008): "A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence." *American Economic Review* **98(1)**: pp. 474–95.

Fuhrer, J. C. (2000): "Habit Formation in Consumption and Its Implications for Monetary-Policy Models." *American Economic Review* **90(3)**: pp. 367–390.

Fusaro, M. A. & D. H. Dutkowsky (2011): "What explains consumption in the very short-run? Evidence from checking account data." *Journal of Macroeconomics* **33(4)**: pp. 542–552.

Gerali, A., S. Neri, L. Sessa, & F. M. Signoretti (2010): "Credit and Banking in a DSGE Model of the Euro Area." *Journal of Money, Credit and Banking* **42(s1)**: pp. 107–141.

Gruber, J. W. (2004): "A present value test of habits and the current account." *Journal of Monetary Economics* **51(7)**: pp. 1495–1507.

Guariglia, A. (2002): "Consumption, habit formation, and precautionary saving: evidence from the British Household Panel Survey." *Oxford Economic Papers* **54(1)**: pp. 1–19.

Guerron-Quintana, P. A. (2008): "Refinements on macroeconomic modeling: The role of non-separability and heterogeneous labor supply." *Journal of Economic Dynamics and Control* **32(11)**: pp. 3613–3630.

Guerron-Quintana, P. A. (2010): "What you match does matter: the effects of data on DSGE estimation." *Journal of Applied Econometrics* **25(5)**: pp. 774–804.

Hansen, B. E. (2007): "Least Squares Model Averaging." *Econometrica* **75(4)**: pp. 1175–1189.

Havranek, T. (2015): "Measuring Intertemporal Substitution: The Importance Of Method Choices And Selective Reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

Havranek, T., R. Horvath, Z. Irsova, & M. Rusnak (2015): "Cross-Country Heterogeneity in Intertemporal Substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

Havranek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85(2)**: pp. 234–244.

Havranek, T. & Z. Irsova (2017): "Do Borders Really Slash Trade? A Meta-Analysis." *IMF Economic Review* **65(2)**: pp. 365–396.

Havranek, T. & A. Sokolova (2016): "Do Consumers Really Follow a Rule of Thumb? Three Thousand Estimates from 130 Studies Say 'Probably Not'." *Working paper 8/2016*, Czech National Bank.

He, Y., J. S. Dyer, & J. C. Butler (2013): "On the axiomatization of the satiation and habit formation utility models." *Operations Research* **61(6)**: pp. 1399–1410.

Heaton, J. (1993): "The Interaction between Time-Nonseparable Preferences and Time Aggregation." *Econometrica* **61(2)**: pp. 353–385.

Heaton, J. (1995): "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications." *Econometrica* **63(3)**: pp. 681–717.

Hirose, Y. (2008): "Equilibrium Indeterminacy and Asset Price Fluctuation in Japan: A Bayesian Investigation." *Journal of Money, Credit and Banking* **40(5)**: pp. 967–999.

Hirose, Y. & S. Naganuma (2010): "Structural Estimation Of The Output Gap: A Bayesian Dsge Approach." *Economic Inquiry* **48(4)**: pp. 864–879.

Iacoviello, M. (2004): "Consumption, house prices, and collateral constraints: a structural econometric analysis." *Journal of Housing Economics* **13(4)**: pp. 304–320.

Iacoviello, M. & S. Neri (2010): "Housing Market Spillovers: Evidence from an Estimated DSGE Model." *American Economic Journal: Macroeconomics* **2(2)**: pp. 125–64.

Ioannidis, J. P. A., T. D. Stanley, & H. Doucouliagos (2017): "The power of bias in economics research." *The Economic Journal* **127(605)**: pp. F236–F265.

Irsova, Z. & T. Havranek (2013): "Determinants of Horizontal Spillovers from FDI: Evidence from a Large Meta-Analysis." *World Development* **42(C)**: pp. 1–15.

Iwamoto, K. (2013): "Habit formation in household consumption: evidence from Japanese panel data." *Economics Bulletin* **33(1)**: pp. 323–333.

Jeffreys, H. (1961): *Theory of Probability.* Oxford: Clarendon Press.

Justiniano, A. & B. Preston (2010): "Can structural small open-economy models account for the influence of foreign disturbances?" *Journal of International Economics* **81(1)**: pp. 61–74.

Kano, T. (2009): "Habit formation and the present-value model of the current account: Yet another suspect." *Journal of International Economics* **78(1)**: pp. 72–85.

Kano, T. & J. M. Nason (2014): "Business Cycle Implications of Internal Consumption Habit for New Keynesian Models." *Journal of Money, Credit and Banking* **46(2-3)**: pp. 519–544.

Kass, R. & A. Raftery (1995): "Bayes Factors." *Journal of the American Statistical Association* **90(430)**: pp. 773–795.

Kiley, M. T. (2010): "Habit Persistence, Nonseparability between Consumption and Leisure, or Rule-of-Thumb Consumers: Which Accounts for the Predictability of Consumption Growth?" *The Review of Economics and Statistics* **92(3)**: pp. 679–683.

Koop, G. (2003): *Bayesian Econometrics.* John Wiley & Sons.

Koopmans, T. C. (1960): "Stationary ordinal utility and impatience." *Econometrica* **28(2)**: pp. 287–309.

Korniotis, G. M. (2010): "Estimating Panel Models With Internal and External Habit Formation." *Journal of Business & Economic Statistics* **28(1)**: pp. 145–158.

Laforte, J.-P. (2007): "Pricing Models: A Bayesian DSGE Approach for the U.S. Economy." *Journal of Money, Credit and Banking* **39(s1)**: pp. 127–154.

Levin, A. T., A. Onatski, J. Williams, & N. M. Williams (2005): "Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models." In "NBER Macroeconomics Annual 2005, Volume 20," NBER Chapters, pp. 229–312.

Levine, P., J. Pearlman, G. Perendia, & B. Yang (2012): "Endogenous Persistence in an estimated DSGE Model Under Imperfect Information." *Economic Journal* **122(565)**: pp. 1287–1312.

Ley, E. & M. F. Steel (2009): "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression." *Journal of Applied Econometrics* **24(4)**: pp. 651–674.

Lubik, T. A. & F. Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *American Economic Review* **94(1)**: pp. 190–217.

Lubik, T. A. & W. L. Teo (2014): "Deep Habits in the New Keynesian Phillips Curve." *Journal of Money, Credit and Banking* **46(1)**: pp. 79–114.

Lucas, Robert E, J. (1978): "Asset Prices in an Exchange Economy." *Econometrica* **46(6)**: pp. 1429–45.

Madigan, D. & J. York (1995): "Bayesian graphical models for discrete data." *International Statistical Review* **63(2)**: pp. 215–232.

Magnus, J. R., O. Powell, & P. Prufer (2010): "A comparison of two model averaging techniques with an application to growth empirics." *Journal of Econometrics* **154(2)**: pp. 139–153.

Matheson, T. (2010): "Assessing the fit of small open economy DSGEs." *Journal of Macroeconomics* **32(3)**: pp. 906–920.

Maurer, J. & A. Meier (2008): "Smooth it Like the 'Joneses'? Estimating Peer-Group Effects in Intertemporal Consumption Choice." *Economic Journal* **118(527)**: pp. 454–476.

Mehra, R. & E. C. Prescott (1985): "The equity premium: A puzzle." *Journal of Monetary Economics* **15(2)**: pp. 145–161.

Mehra, Y. P. & E. W. Martin (2003): "Why does consumer sentiment predict household spending?" *Economic Quarterly* **89(4)**: pp. 51–67.

Mertens, K. & M. O. Ravn (2011): "Understanding the Aggregate Effects of Anticipated and Unanticipated Tax Policy Shocks." *Review of Economic Dynamics* **14(1)**: pp. 27–54.

Milani, F. (2007): "Expectations, learning and macroeconomic persistence." *Journal of Monetary Economics* **54(7)**: pp. 2065–2082.

Moral-Benito, E. (2015): "Model Averaging In Economics: An Overview." *Journal of Economic Surveys* **29(1)**: pp. 46–75.

Naik, N. Y. & M. J. Moore (1996): "Habit Formation and Intertemporal Substitution in Individual Food Consumption." *The Review of Economics and Statistics* **78(2)**: pp. 321–28.

Necker, S. (2014): "Scientific misbehavior in economics." *Research Policy* **43(10)**: p. 1747–1759.

Nickell, S. J. (1981): "Biases in Dynamic Models with Fixed Effects." *Econometrica* **49(6)**: pp. 1417–26.

Obstfeld, M. (1992): "International Adjustment with Habit-Forming Consumption: A Diagrammatic Exposition." *Review of International Economics* **1(1)**: pp. 32–48.

Pagano, P. (2004): "Habit persistence and the marginal propensity to consume in Japan." *Journal of the Japanese and International Economies* **18(3)**: pp. 316–329.

Rabanal, P. (2007): "Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model." *Journal of Economic Dynamics and Control* **31(3)**: pp. 906–937.

Raftery, A., D. Madigan, & J. Hoeting (1997): "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* **92(437)**: pp. 179–191.

Ravn, M., S. Schmitt-Grohé, & M. Uribe (2006): "Deep Habits." *Review of Economic Studies* **73(1)**: pp. 195–218.

Ravn, M. O., S. Schmitt-Grohe, M. Uribe, & L. Uuskula (2010): "Deep habits and the dynamic effects of monetary policy shocks." *Journal of the Japanese and International Economies* **24(2)**: pp. 236–258.

Rhee, W. (2004): "Habit formation and precautionary saving: evidence from the Korean household panel studies." *Journal of Economic Development* **29(2)**: pp. 1–19.

Rozen, K. (2010): "Foundations of Intrinsic Habit Formation." *Econometrica* **78(4)**: pp. 1341–1373.

Rustichini, A. & P. Siconolfi (2014): "Dynamic theory of preferences: Habit formation and taste for variety." *Journal of Mathematical Economics* **55(C)**: pp. 55–68.

Sahuc, J.-G. & F. Smets (2008): "Differences in Interest Rate Policy at the ECB and the Fed: An Investigation with a Medium-Scale DSGE Model." *Journal of Money, Credit and Banking* **40(2-3)**: pp. 505–521.

Sala-I-Martin, X., G. Doppelhofer, & R. I. Miller (2004): "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review* **94(4)**: pp. 813–835.

Schmitt-Grohé, S. & M. Uribe (2012): "What's News in Business Cycles." *Econometrica* **80(6)**: pp. 2733–2764.

Slanicay, M. & O. Vašíček (2011): "Habit Formation, Price Indexation and Wage Indexation in the DSGE Model: Specification, Estimation and Model Fit." *Review of Economic Perspectives* **11(2)**: pp. 71–91.

Smets, F. & R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* **1(5)**: pp. 1123–1175.

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

Sommer, M. (2007): "Habit Formation and Aggregate Consumption Dynamics." *The B.E. Journal of Macroeconomics* **7(1)**: pp. 1–25.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

Stock, J. H. & J. Wright (2000): "GMM with Weak Identification." *Econometrica* **68(5)**: pp. 1055–1096.

Sugo, T. & K. Ueda (2008): "Estimating a dynamic stochastic general equilibrium model for Japan." *Journal of the Japanese and International Economies* **22(4)**: pp. 476–502.

Trigari, A. (2009): "Equilibrium Unemployment, Job Flows, and Inflation Dynamics." *Journal of Money, Credit and Banking* **41(1)**: pp. 1–33.

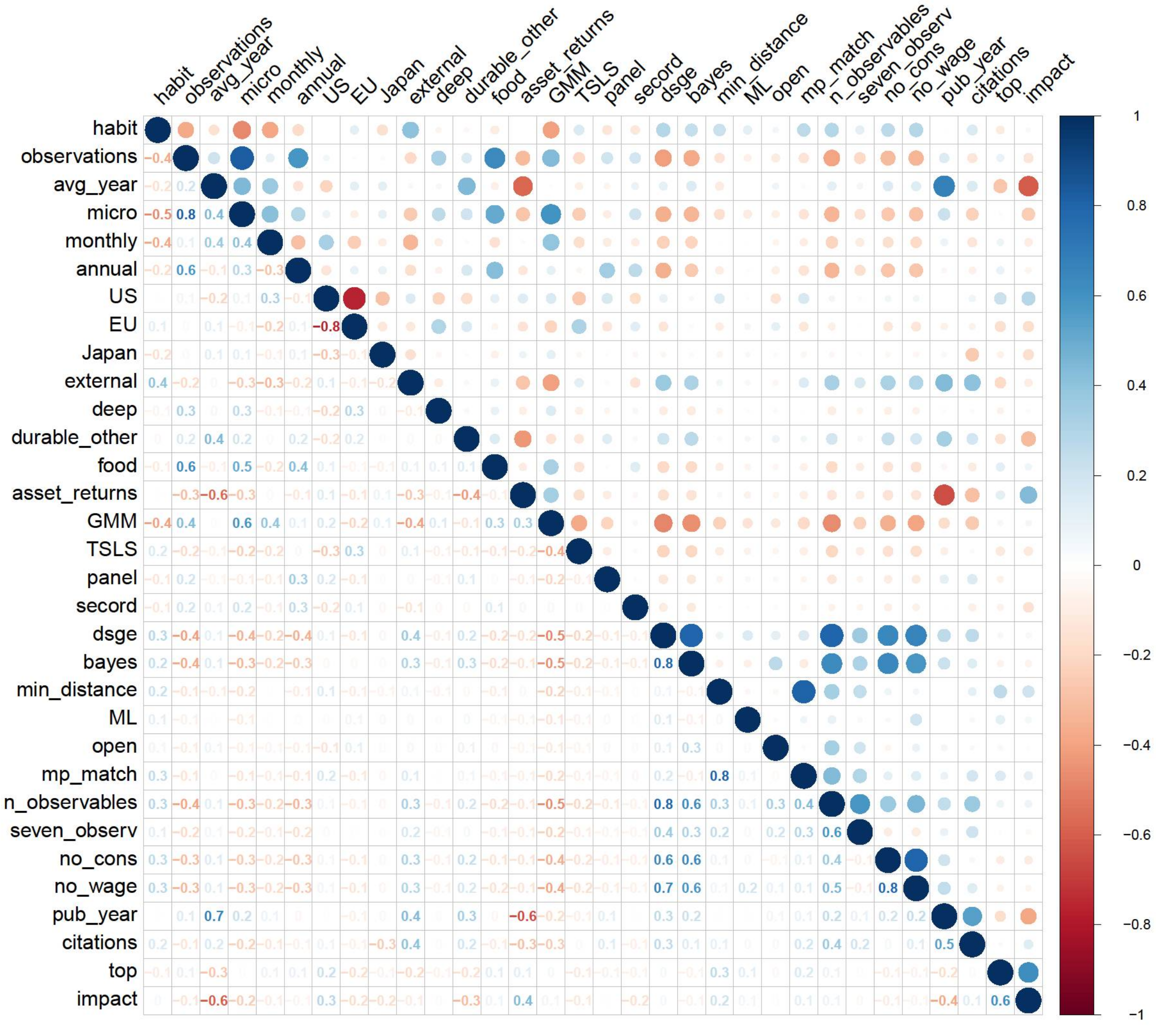
Wieland, V., T. Cwik, G. J. Müller, S. Schmidt, & M. Wolters (2012): "A new comparative approach to macroeconomic modeling and policy analysis." *Journal of Economic Behavior & Organization* **83(3)**: pp. 523–541.

Wouters, R. & F. Smets (2005): "Comparing shocks and frictions in US and euro area business cycles: a Bayesian DSGE Approach." *Journal of Applied Econometrics* **20(2)**: pp. 161–183.

4.A Supplementary Statistics and Analysis

4.A.1 Correlation of the Variables

Figure 4.6: Correlation matrix



Notes: A description of the variables is available in Table 4.4.

4.A.2 Diagnostics of BMA

Figure 4.7: Model size and convergence

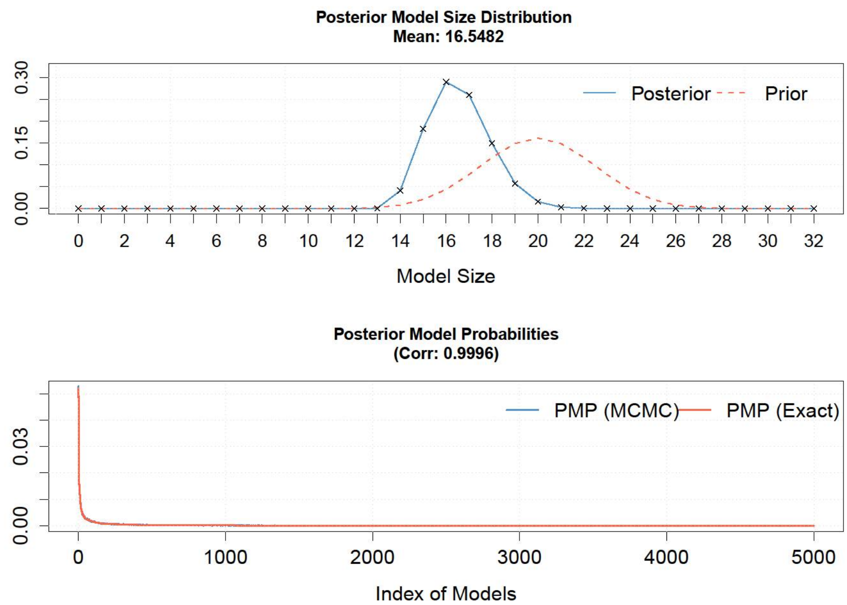


Table 4.6: Explaining the differences in the estimates of habit formation (frequentist approach)

Response variable:	Frequentist model averaging		
Estimate of habit formation	Coef.	Std. er.	p-value
Precision (1/SE)	0.563	0.457	0.218
Data characteristics			
No. of obs.	0.000	0.000	1.000
Average year	0.000	0.006	1.000
Micro	-0.836	0.365	0.022
Monthly	-0.000	0.276	1.000
Annual	0.356	0.136	0.009
Countries examined			
US	0.264	0.070	0.000
EU	-0.000	0.002	1.000
Japan	-0.000	0.076	1.000
Variable definition			
External	0.098	0.197	0.619
Deep	0.000	0.025	1.000
Durable	-0.000	0.007	1.000
Food	0.000	0.068	1.000
Asset returns	0.000	0.226	1.000
Estimation approach			
GMM	-0.000	0.051	1.000
TSLS	0.000	0.286	1.000
Panel	-0.405	0.120	0.001
Second-order approx.	-0.000	0.367	1.000
DSGE	-0.000	0.025	1.000
Bayes	-0.000	0.056	1.000
Minimum distance	0.000	0.725	1.000
ML	0.000	0.263	1.000
DSGE specification			
Open	0.342	0.056	0.000
Matching IR to mon. policy	-0.000	0.323	1.000
No. of observables	-0.000	0.022	1.000
Seven observables from SW	0.000	0.006	1.000
No consumption	0.000	0.051	1.000
No wage	0.000	0.019	1.000
Publication characteristics			
Publication year	0.000	0.002	1.000
Citations	0.000	0.108	1.000
Top journal	-0.414	0.185	0.025
Impact	0.062	0.105	0.552
Constant	-3.011	0.803	0.000
Studies	81		
Observations	597		

Notes: Frequentist model averaging requires full enumeration of models, which are weighted by information criteria. We employ Mallow's criterion to select the weights since it delivers weights that are asymptotically optimal. Because our model consists of 32 potential explanatory variables, the model space is huge, 2^{32}, and full enumeration would take a prohibitive amount of time. We therefore follow the approach suggested by Amini & Parmeter (2012), who build on Magnus *et al.* (2010), and use orthogonalization of the covariate space, thus reducing the number of models that need to be estimated from 2^{32} to 32.

Table 4.7: Summary of BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
16.5482	$3e+06$	$1e+06$	10.33624 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
$749,088$	$4.3e+09$	0.017%	92%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9996	597	uniform	UIP
Shrinkage-Stats			
Av= 0.9983			

Notes: In this specification we employ the priors suggested by Eicher *et al.* (2011) based on the predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of the data).

4.A.3 Alternative Priors for BMA

Table 4.8: Explaining the differences in the estimates of habit formation (alternative priors)

Response variable:	Bayesian model averaging		
Estimate of habit formation	Post. mean	Post. std. dev.	PIP
Precision (1/SE)	0.272	0.144	0.878
Data characteristics			
No. of obs.	0.000093	0.0002	1.000
Average year	0.006	0.002	1.000
Micro	-0.572	0.072	1.000
Monthly	-0.357	0.119	1.000
Annual	0.226	0.067	1.000
Countries examined			
US	0.201	0.047	1.000
EU	-0.002	0.021	1.000
Japan	-0.112	0.069	1.000
Variable definition			
External	0.073	0.093	0.437
Deep	0.000	0.004	0.020
Durable	-0.001	0.011	0.035
Food	0.012	0.047	0.079
Asset returns	0.006	0.039	0.062
Estimation approach			
GMM	-0.002	0.017	0.029
TSLS	0.055	0.109	0.238
Panel	-0.528	0.064	1.000
Second-order approx.	-0.392	0.112	0.983
DSGE	-0.011	0.045	0.086
Bayes	-0.002	0.017	0.038
Minimum distance	0.058	0.134	0.192
ML	-0.005	0.032	0.123
DSGE specification			
Open	0.216	0.100	0.907
Matching IR to mon. policy	-0.021	0.050	0.210
No. of observables	0.000	0.004	0.033
Seven observables from SW	-0.002	0.018	0.021
No consumption	0.001	0.012	0.037
No wage	0.000	0.009	0.021
Publication characteristics			
Publication year	0.002	0.005	0.124
Citations	0.012	0.037	0.111
Top journal	-0.563	0.117	1.000
Impact	0.119	0.062	0.901
Constant	-2.753	NA	1.000
Studies	81		
Observations	597		

Notes: PIP = posterior inclusion probability. We use an alternative to the unit information prior, the BRIC prior suggested by Fernandez *et al.* (2001), which takes into account the number of explanatory variables for the determination of the weight of the zero prior for the regression parameters. In this set of priors we also employ the random beta-binomial model prior (Ley & Steel 2009), which implies that each *model size* has the same prior probability.

4.B BMA and Model Uncertainty in Meta-Analysis

This section discusses how our application of Bayesian model averaging departs from the standard approach employed by the literature. In Section 4.4 we identify 31 factors that we believe could contribute to the heterogeneity in the reported estimates $\hat{\gamma}$ of habit formation, and we would like to quantify their effects by estimating the following model (4.6), already featured in Subsection 4.4.2:

$$\hat{\gamma}_{ij} = \alpha_0 + \sum_{k=1}^{31} \theta_k Z_{k,ij} + \varepsilon_{ij}. \tag{4.6}$$

Among the 31 explanatory variables Z_k, eight describe the critical features of the data generating process, and some of the remaining factors may have effects that are small or insignificant. As *a priori* we do not know which of these remaining elements have a systematic effect on the estimates, we are facing a total of 2^{31-8} possible models we could use to describe the variation in the estimates. Our use of BMA aims to resolve this type of uncertainty: uncertainty over which features of study design affect the estimates of the habit formation parameter, conditional on the effects being linear and on there being only 31 possible explanatory variables.

Let $B \equiv [B_1, B_2, ...B_{2^{31-8}}]$ denote all possible combinations of explanatory variables Z_k that could be included in regression (4.6). Let θ denote a vector of the 31 effects θ_k associated with regressors Z_k. The posterior of θ can then be written as

$$p(\theta|\hat{\gamma}, B) = \sum_{m=1}^{2^{31-8}} p(\theta|\hat{\gamma}, B_m) p(B_m|\hat{\gamma}, B), \tag{4.7}$$

where $p(\theta|\hat{\gamma}, B_m)$ is obtained by estimating model B_m on the set $\hat{\gamma}$, and $p(B_m|\hat{\gamma}, B)$ is a posterior model probability associated with combination B_m that can be calculated via

$$p(B_m|\hat{\gamma}, B) = \frac{p(B_m|B)p(\hat{\gamma}|B_m)}{\sum_{m=1}^{2^{31-8}} p(B_m|B)p(\hat{\gamma}|B_m)}, \tag{4.8}$$

where $p(B_m|B)$ and $p(\hat{\gamma}|B_m)$ are prior probability of model B_m and its marginal likelihood. We follow the standard BMA approach and use (4.8) to identify posterior model probabilities. But in doing so we treat $\hat{\gamma}$ as data points, not estimates—this treatment ignores a portion of uncertainty attached to the choices of data and methodology made by the authors of the primary studies.

In Section 4.4 we state that the literature studying consumption habits is very diverse, employing different data sets and methods. To be more precise, all studies in our sample employ unique data sets, and in some cases there is even variation in data used within one study. Furthermore, the structural models that primary studies rely on vary. Even though many studies employ approximations similar to (4.4), the estimated specifications differ substantially: (4.4) can be estimated on its own, or as part of some DSGE model. In Section 4.4 we pinpoint key differences between modeling approaches; however, with this strategy we cannot hope to fully capture the diversity of structural models employed in the literature and the uncertainty attached to the modeling choices. Below we sketch a strategy that could, if successfully implemented, resolve these issues. We thank an anonymous referee for providing the underlying idea.

Let $A \equiv [A_1, A_2, ...A_n]$ denote the sequence of structural models used to obtain estimates $\hat{\gamma} \equiv [\hat{\gamma}_1, \hat{\gamma}_2, ...\hat{\gamma}_n]$ where $n = 597$, and $Y \equiv [Y_1, Y_2, ...Y_n]$ be corresponding data sets. Let $\mathscr{A}$ denote the set of structural models used within the literature. (The number of elements in $\mathscr{A}$ is smaller than 597 because many studies apply same methodology to different data sets.) The probability of an auxiliary model B_m conditional on data Y and a collection of models A can then be expressed as

$$p(B_m|B, \mathscr{A}, Y) = \int p(B_m|B, \hat{\gamma})p(\hat{\gamma}|\mathscr{A}, Y)d\hat{\gamma}, \tag{4.9}$$

where $p(\hat{\gamma}|\mathscr{A}, Y)$ is probability of the set of habit parameter estimates conditional on estimating structural model set $\mathscr{A}$ on data Y. To account for uncertainty with respect to structural models, we would need to further decompose this probability as follows:

$$\begin{aligned} p(\hat{\gamma}|\mathscr{A}, Y) &= p(\hat{\gamma}_1, \hat{\gamma}_2, ...\hat{\gamma}_n|\mathscr{A}, Y) = \\ &= \prod_{i=1}^{n} \left[p(\hat{\gamma}_i|A_i, Y_i)p(A_i|\mathscr{A}, Y)\right]. \end{aligned} \tag{4.10}$$

where $p(A_i|\mathscr{A}, Y)$ is a probability attached to structural model A_i, conditional on model set $\mathscr{A}$ and data Y.

A study wishing to fully account for uncertainty over which structural models should be used to evaluate Euler equations with habits would need to assess $p(A_i|\mathscr{A}, Y)$, facing a variety of obstacles, some of which may prove insurmountable. As discussed before, structural models in $\mathscr{A}$ differ along many dimensions, which makes comparing their relative performance not straightforward. Furthermore, the data in Y have features that may affect the relative performance of each model. As we saw in Figure 4.2, it seems that estimates coming from macro and micro data and data of different frequencies are associated with distinct distributions. What is more, some structural models are meant to only be applied to certain types of data. For example, models that account for taste shifters are designed for micro studies, whereas DSGE models can only be estimated on aggregate data sets.

In our understanding these difficulties make complete Bayesian treatment of both sources model uncertainty infeasible. In this paper we address the model uncertainty associated with the choice of variables in the meta-analysis model. Since we do not address the other source of model uncertainty, related to structural models A, the resulting posterior standard deviations may be underestimated. For this reason, as a robustness check, we also estimate a frequentist model averaging specification, for which we can use the typical approach employed in estimated dependent variable models.

4.C Studies Included in the Data Set

Table 4.9: List of primary studies

Alessie & Kapteyn (1991)	Bover (1991)	Campbell & Mankiw (1991)
Ferson & Constantinides (1991)	Braun *et al.* (1993)	Heaton (1993)
Naik & Moore (1996)	Dynan (2000)	Fuhrer (2000)
Stock & Wright (2000)	Guariglia (2002)	Baltagi *et al.* (2003)
Mehra & Martin (2003)	Smets & Wouters (2003)	Gruber (2004)
Iacoviello (2004)	Lubik & Schorfheide (2004)	Pagano (2004)
Rhee (2004)	Bouakez *et al.* (2005)	Christiano *et al.* (2005)
Levin *et al.* (2005)	Wouters & Smets (2005)	Batini *et al.* (2006)
Boivin & Giannoni (2006)	Adolfson *et al.* (2007)	Collado & Browning (2007)
Del Negro *et al.* (2007)	Laforte (2007)	Milani (2007)
Rabanal (2007)	Smets & Wouters (2007)	Sommer (2007)
Auray & Feve (2008)	Del Negro & Schorfheide (2008)	Edge *et al.* (2008)
Guerron-Quintana (2008)	Maurer & Meier (2008)	Sahuc & Smets (2008)
Sugo & Ueda (2008)	Andrés *et al.* (2009)	Christoffel *et al.* (2009)
Dennis (2009)	Kano (2009)	Trigari (2009)
Alessie & Teppa (2010)	Bartolomeo *et al.* (2010)	Bekaert *et al.* (2010)
Castelnuovo & Nistico (2010)	Chib & Ramamurthy (2010)	Fernández-Villaverde (2010)
Guerron-Quintana (2010)	Hirose & Naganuma (2010)	Iacoviello & Neri (2010)
Justiniano & Preston (2010)	Kiley (2010)	Korniotis (2010)
Matheson (2010)	Ravn *et al.* (2010)	Altig *et al.* (2011)
Bjornland *et al.* (2011)	Carroll *et al.* (2011)	Fusaro & Dutkowsky (2011)
Mertens & Ravn (2011)	Slanicay & Vašíček (2011)	Levine *et al.* (2012)
Schmitt-Grohé & Uribe (2012)	Iwamoto (2013)	Everaert & Pozzi (2014)
Heaton (1995)	Flavin & Nakagawa (2008)	Andreasen (2012)
Ferson & Harvey (1992)	Cooley & Ogaki (1996)	Hirose (2008)
Allais (2004)	Gerali *et al.* (2010)	De Graeve (2008)
Lubik & Teo (2014)	Adolfson *et al.* (2008)	Eichenbaum & Hansen (1990)

Notes: Both published and unpublished studies are included if they control for vertical spillovers. We use all comparable estimates reported in the studies. The search for primary studies was terminated on March 31, 2010. A list of excluded studies, with reasons for exclusion, is available in the online appendix.

Chapter 5

Nowcasting Czech GDP in Real Time

Abstract

In this paper, we employ a Dynamic Factor Model (DFM) to nowcast Czech GDP. Using multiple vintages of historical data and taking into account the publication lags of various monthly indicators, we evaluate the real-time performance of the DFM over the 2005–2012 period. The main result of this paper is that the accuracy of model-based nowcasts is comparable to that of the nowcasts of the Czech National Bank (CNB). Moreover, combining the DFM and the CNB nowcasts results in more accurate performance than in the case of the individual nowcasts alone. Our results also suggest that foreign variables are crucial for the accuracy of the model, while omitting financial and confidence indicators does not worsen the nowcasting performance.

5.1 Introduction

Because of considerable publication delays in the release of GDP data, the current state of the economy is subject to sizeable uncertainty. Accurate and timely estimates of the current state of the economy are therefore especially important for policymakers, who make their decisions in real time. In the present turbulent sit-

The paper was published in the *Economic Modelling* [2016, 54(C), pp. 26–39]. We would like to thank Michal Franta, Tomáš Havránek, Roman Horváth, Gabriel Perez Quiros, Christian Schumacher, and Luboš Růžička for their useful comments and discussions, and Marta Bańbura for sharing her Matlab code. We would also like to thank the Editor of the *Economic Modelling* and three anonymous referees for their comments and suggestions that helped to improve the paper. We acknowledge the support of the Czech National Bank, the Grant Agency of Charles University (grant # 888413), and the Grant Agency of Czech Republic (grant P402/12/G097).

uation, obtaining the most up-to-date forecasts of GDP, possibly after each new data announcement, is becoming even more important, for example, in the event of irregular monetary policy meetings in the midst of a crisis or other unexpected developments in the economy. Such up-to-date forecasts of Czech GDP produced in real time are the objective of this study.

Forecasters face several problems when producing predictions in real time. Macroeconomic variables are announced in a non-synchronous manner, that is, they have different publication lags. As a result, forecasters have to work with datasets that contain many missing observations towards the end of the sample (the so-called ragged end problem). Another problem forecasters typically face is the fact that data are sampled at different frequencies. Most of the traditional forecasting models – such as leading indicators models and classical vector autoregressions –cannot easily deal with these issues: they cannot utilize the most up-to-date data releases in a model-consistent fashion.

Enter nowcasting. The nowcasting framework of Giannone *et al.* (2008) has become the workhorse model of short-term forecasters at many central banks and other institutions (for an extensive list of references see Bańbura *et al.* 2013). The framework is based on a dynamic factor model cast in the state-space representation and on the application of the Kalman filter to deal with mixed frequencies and unbalanced datasets.[1] The framework can accommodate a potentially large number of variables by summarizing the information with a few common factors, thus overcoming the so-called curse of dimensionality (Stock & Watson 2002b; Bernanke & Boivin 2003). An additional advantage of the framework is that it allows forecasters not only to predict variables of interest in real time, but also to interpret and comment on the sources of the changes in the forecasts. This provides a story-telling dimension and a deeper understanding of the forecast that is almost as important to policymakers as the accuracy of the forecast itself. This feature is missing from most of the statistical models that are currently used for near-term projections.

[1] Previous seminal contributions include Wallis (1986) and Evans (2005). See also Foroni & Marcellino (2013), who provide a survey of state-of-the-art mixed frequency models that can deal with ragged end problems.

An additional challenge for real-time forecasters is the presence of data revisions. Typically, the forecasting exercises and model selection are performed using revised data. It is well known, however, that the revisions to macroeconomic data are frequent and large (Faust *et al.* 2005; Garratt & Vahey 2006; Aruoba 2008; Croushore 2011; Fernandez *et al.* 2011). Therefore, working with the last available data may provide starkly different results than those obtained using real-time data (as documented by many studies: Robertson & Tallman 1998; Faust *et al.* 2003; Orphanides 2001; Kugler *et al.* 2005; Molodtsova *et al.* 2008; Marcellino & Musso 2011; Ince & Papell 2013). As for the properties of revisions to Czech GDP, in our previous research (Rusnak 2013), we find that the revisions are relatively large. Performing a proper real-time forecasting exercise using Czech data therefore seems to be greatly needed.

The short-term forecasting performance of various models of Czech GDP has been assessed before by many studies (Benda & Ruzicka 2007; Arnostova *et al.* 2011; Havranek *et al.* 2012; Horvath 2012). Unfortunately, most of these studies do not account for publication lags and data revisions, which renders the relevance of their results to policymakers rather limited.[2] Consider, for example, the official comments that the CNB makes after each release of GDP. Out of 32 comments published by the the Czech National Bank (CNB) during the 2005–2012 period, roughly 17 of them mention revisions to the national accounts as one of the sources of the deviation of the official CNB forecasts from the announced data. Obviously, revisions must be therefore considered an important issue to policymakers. Truly real-time exercises to evaluate the performance of dynamic factor models in the presence of data revisions are still relatively scarce. The exceptions are Schumacher & Breitung (2008) for Germany, Camacho & Perez-Quiros (2010) for the euro area, and Bańbura *et al.* (2013) and Lahiri & Monokroussos (2013) for the US. To the best of our knowledge, we are the first to investigate the performance of forecasts of Czech GDP in a truly real-time setting that employs unrevised vintages of historical data.

[2] Arnostova *et al.* (2011), in their replication of Rünstler *et al.* (2009), account for publication lags, but their analysis is based on a revised dataset.

In this paper, we focus on the performance of the DFM in obtaining accurate forecasts of the current quarter GDP growth (so-called nowcasts). Accurate nowcasts are important since they serve as inputs to the structural models that are used for medium to long-term prediction (the CNB uses a G3 DSGE model, see Andrle *et al.* 2009, for details). Furthermore, the CNB comments on the releases of the latest GDP growth figures and discusses the deviations from its official predictions. This makes the accuracy of CNB nowcasts of crucial importance.

Formal model-based forecasts are typically compared to naive benchmarks or to other competing models. Comparisons with official central bank forecasts are rare, but do exist, especially in the context of model combinations (Lees *et al.* 2007; Adolfson *et al.* 2007; Groen *et al.* 2009; Edge *et al.* 2010; McDonald & Thorsrud 2011). A common finding of these studies is that the accuracy of model-based forecasts of GDP is comparable to that of the official forecasts of the respective central banks.[3] In this paper, we contribute to this literature by evaluating the performance of the dynamic factor model using Czech real-time data and comparing it with the accuracy of the nowcasts of the Czech National Bank.

Finally, we show how one can use the methodology of Bańbura & Modugno (2010) to decompose the updates of Czech GDP nowcasts into the contributions of the individual variables – so-called *news*. This is possible since the dynamic factor model produces forecasts for all of the variables included. One can then interpret changes in the forecasts stemming from the differences between the actual data released and their predicted values. For example, it is reasonable to assume that a higher-than-expected value of industrial production will cause the forecast to be revised upwards. The dynamic factor model can quantify such statements. Similar decompositions of forecast updates are now regularly used by many central banks (see for example ECB 2008; Bundesbank 2009) to enhance the understanding of their short-term forecasts.

Our results suggest that the nowcasting performance of the medium-scale DFM is comparable to the nowcasts of the Czech National Bank. In addition, we find

[3]Note that not all of these papers use unrevised data, so the comparability should be interpreted with caution.

that the simple average of the DFM and CNB nowcasts is more accurate than the nowcasts of the DFM and CNB alone. We also find that the DFM nowcasts add value to the CNB nowcasts: conditional on the CNB nowcast, on average, GDP growth turns out to be higher when the DFM nowcast is higher. Similarly to D'Agostino & Giannone (2012) we find that the relative performance of the DFM is better at times of crisis, which are characterized by large comovements of variables. We also find that the inclusion of foreign variables is crucial: if we exclude foreign variables the performance worsens significantly, while the omission of financial variables or surveys does not result in a dramatic deterioration of the forecasting accuracy.

The remainder of this paper is organized as follows. Section 5.2 briefly discusses the dynamic factor model, Section 5.3 describes our real-time dataset and provides details of the empirical exercise together with its results. Section 5.4 presents examples of nowcast update decompositions, while Section 5.5 provides further results and sensitivity checks. Section 5.6 concludes.

5.2 Dynamic factor model

Dynamic factor models aim at capturing the most important features of the data while remaining parsimoniously specified. They do so by assuming that the bulk of the comovements in macroeconomic variables are driven by just a few common factors (this seems to be the case in the US, see Giannone *et al.* 2005). The technology of dynamic factor models has evolved over time. The first generation consisted of small-scale models estimated by maximum likelihood and the Kalman filter (Engle & Watson 1981; Mariano & Murasawa 2003; Camacho & Perez-Quiros 2010). These models were able to handle data irregularities, but were unable to utilize more than a few variables.

Forecasters and policymakers, however, monitor a large number of different time series (Bernanke & Boivin 2003). Because the time span of most of the series is rather short – a problem of even bigger importance in economies that transformed to a market economy relatively recently – applying traditional models to a large number of variables would result in parameter proliferation and imprecise forecasts. Therefore, the second generation of factor models uses nonparametric principal components estimation of factors from large cross sections (Chamberlain & Rothschild 1983; Forni & Reichlin 1998; Forni *et al.* 2000; Stock & Watson 2002a;b). However, principal components cannot deal with ragged ends on their own.

The third generation of factor models combines the first and second generations: factors approximated by principal components are utilized within a state-space framework (Giannone *et al.* 2008; Rünstler *et al.* 2009; Bańbura & Rünstler 2011). Thus, they constitute a model that can handle large data sets with data irregularities present in a real-time forecasting setting. The asymptotic properties of these models can be found in Doz *et al.* (2011).

Finally, the most recent papers use the expectation-maximization algorithm to obtain maximum likelihood estimates of large models that are able to deal with unbalanced datasets (Schumacher & Breitung 2008; Bańbura & Modugno 2010). On the whole, this approach consists of iterating between the two steps: estimating the parameters conditional on the factors, and estimating the factors conditional on the parameters from previous iterations. The asymptotic theory is provided in Doz *et al.* (2012).

An accessible survey of dynamic factor models can be found in Stock & Watson (2010), while Bai & Ng (2008) provide a more technical survey. Bańbura *et al.* (2010b) and Bańbura *et al.* (2013) survey the application of factor models with a focus on nowcasting.

In our empirical exercise we will use the latest generation dynamic factor model estimated by the expectation-maximization algorithm. We begin by specifying the model for monthly variables:

$$x_t = \Lambda f_t + \varepsilon_t \tag{5.1}$$

$$f_t = A_1 f_{t-1} + \cdots + A_p f_{t-p} + u_t, \tag{5.2}$$

where x_t is a vector of monthly variables transformed into stationary ones, f_t is a vector of r (unobserved) common factors, and u_t is a vector of idiosyncratic shocks. Λ denotes a matrix of factor loadings, while $A_1, \ldots, A_p$ denote the autoregressive coefficients for the factors.

Quarterly variables are modeled using the approximation of Mariano & Murasawa (2003). We adopt the convention that the quarterly GDP level, denoted by GDP_t^Q, is assigned to the third month of the quarter. The unobserved monthly counterpart of GDP is denoted by GDP_t^M.

$$GDP_t^Q = GDP_t^M + GDP_{t-1}^M + GDP_{t-2}^M \qquad t = 3, 6, 9, \ldots \tag{5.3}$$

We further define

$$Y_t^Q = 100 * log(GDP^Q) \tag{5.4}$$

$$Y_t^M = 100 * log(GDP^M), \tag{5.5}$$

where log denotes natural logarithm, and assume that the monthly growth rate of GDP, $y_t = Y_t^M - Y_{t-1}^M$, admits the same factor model representation as the monthly variables:

$$y_t = \Lambda_Q f_t + \varepsilon_t^Q \tag{5.6}$$

We link y_t with the observed GDP data by constructing the following partially observed monthly series:

$$y_t^Q = \begin{cases} Y_t^Q - Y_{t-3}^Q & t = 3, 6, 9, \ldots \\ \text{unobserved} & \text{otherwise} \end{cases} \tag{5.7}$$

Finally, we use the approximation suggested by Mariano & Murasawa (2003):

$$\begin{aligned} y_t^Q = Y_t^Q - Y_{t-3}^Q &\approx (Y_t^M + Y_{t-1}^M + Y_{t-2}^M) - (Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M) \\ &= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4} \end{aligned}$$

Direct numerical maximization of the likelihood can be computationally challenging and inefficient if a model contains more than a few variables. Therefore, the estimation is performed using the expectation-maximization algorithm (Shumway & Stoffer 1982; Watson & Engle 1983; Schumacher & Breitung 2008). We use the methodology of Bańbura & Modugno (2010), who generalize the method so that the DFM can deal with an arbitrary pattern of missing observations. In brief, the estimation can be described as consisting of iterations of two steps. In the first step, the expectation of the log-likelihood conditional on the estimates from the previous iteration is calculated. In the second step, the parameters are re-estimated using the expected likelihood from the previous step. The initial values are obtained by filling in the missing observations by draws from N(0,1) and estimating the principal components on the balanced part of the sample (similarly as in Giannone *et al.* (2008)). For further technical details of the EM iterations we use in this application, see Bańbura & Modugno (2010).

5.3 Real-time nowcasting exercise

5.3.1 Real-time data set

We compose a real-time database of 99 monthly vintages: the first vintage is from October 2004, and the last from December 2012. We collect a panel of 28 headline macroeconomic variables that covers headline hard data, financial variables, surveys, and foreign variables. Most data start in January 2000 and span up to the latest observation available in that particular vintage. The exceptions are the government bond yield and the services confidence indicator, which start in April 2000 and May 2002, respectively. Our dataset is relatively balanced in the number of series pertaining to each group. In particular, we have nine series of hard data covering the production, labor, and trade sectors of the economy. A further seven financial series cover exchange and interest rates, stock prices, and credit aggregates, while five survey series cover confidence indicators of business and consumers. Finally, we add six series of foreign variables covering hard, financial, and survey variables. The variables are transformed to stationarity by taking log-differences (or first differences in the case of several confidence indicators).[4] Plots of the series can be found in the Appendix. Further, before estimation, the variables are standardized to have zero mean and unit variance.

For the series that are subject to frequent revisions (ten overall, most of the hard data variables) we use the OECD Real-Time Database. In addition, we collect vintages of credit from CNB Monetary Statistics Monthly Bulletin publications. Most of the financial variables (interest and exchange rates) and surveys are not revised. The exception is the euro area business climate indicator, which is revised due to changes in the composition of the euro area. Therefore, for this variable we collect vintages from press releases available on the European Commission website. Unemployment is not revised, but it is published as not seasonally adjusted. Performing seasonal adjustment on the latest available series first and then using the data sequentially would probably introduce information about trends that was not available at the time of the forecast (see also Orphanides & van Norden 2002). Therefore, we perform seasonal adjustment sequentially, using only the information available at the time of the relevant forecast.[5]

The number of variables is relatively small compared to what is typically used in factor model applications.[6] However, Bańbura *et al.* (2010a;b) show that the gains

[4]Note that it is not clear whether one should also difference the confidence indicators: some authors prefer to keep them in levels (Camacho & Perez-Quiros 2010), while others do difference them (Giannone *et al.* 2008; Bańbura *et al.* 2013). We followed the suggestion of a referee and also estimated the specification with surveys in levels: the results suggest that the accuracy of the model deteriorated, so we decided to keep the surveys in differences. These results are available upon request.

[5]Seasonal adjustment was performed by employing Demetra software and using the Tramo-Seats procedure. Note that the real-time vintages of construction from the OECD Real-Time Database were also only available as not seasonally adjusted. Therefore, we adjusted them as well.

[6]Note that the Monte Carlo evidence by Doz *et al.* (2012) suggests that sufficient EM estimation robustness can be obtained with just a handful of variables.

from including more than 20–40 variables are rather modest and that disaggregate information does not improve the forecast accuracy. Arnostova *et al.* (2011) consider 98 indicators to forecast Czech GDP, but almost half of them are disaggregate information on industrial production and sales. We do not include this disaggregate information since this would probably result in contaminating the estimated common factor with idiosyncratic shocks to industrial production and sales (see also Boivin & Ng 2006, for a more general discussion). By including only headline variables, we are, in fact, also loosely following the recommendation of Alvarez *et al.* (2012) to include only one reference series for each economic concept. Note also that the set of the variables we use in this exercise coincides to the large extent with the data that are typically monitored by the market participants in the Czech Republic.

Other than dismissing the disaggregate sectoral information and omitting some variables due to unavailability of real-time vintages (such as fiscal data covering monthly government spending and tax revenues), we opt not to pre-select the indicators any further. We find pre-selection of indicators rather problematic. First, the existing procedures recommended by Boivin & Ng (2006) and Bai & Ng (2008) do not take into account the presence of ragged ends and differences in the timeliness of the variables. Arnostova *et al.* (2011) compute bivariate correlations with GDP and exclude those with a correlation lower than 0.5. We opt not to follow this practice since it neglects the ragged ends and potential dynamic cross-correlation between different variables. Second, it is well known that the predictive content of individual variables is not stable over time (De Mol *et al.* 2008; Rossi & Sekhposyan 2010; Stock & Watson 2012; Kuzin *et al.* 2013) and therefore pre-selecting the indicators might not be the optimal strategy. Third, a model that includes all of the key variables might be of greater interest to policymakers than a model with pre-selected indicators only, since policymakers might want to comment on various headline data releases. Fourth, pre-selecting indicators using data from tranquil periods might have a negative effect on the accuracy of forecasts during crisis periods. Finally, we believe that by not including too many variables (over)representing the same concept, the dynamic factor model will assign the correct weights to the variables included (see

Bańbura *et al.* 2013, for more details).

Table 5.1: Data set

No.	Group	Variable	Rev.	Pub. Lag	Unb. Pat.	Source
1	Hard	Real GDP	Y	68 to 71	4,5,3–4,5,3	OECD
2	Hard	Industrial production index	Y	37 to 45	2–2	OECD
3	Hard	Construction output	Y	37 to 45	2–2	OECD
4	Hard	Retail Sales	Y	35 to 49	2–2	OECD
5	Hard	Unemployment rate	N	8 to 11	1–1	MLSA
6	Hard	CPI total	N	8 to 11	1–1	CZSO
7	Hard	Exports (current prices)	Y	35 to 39	2–2	OECD
8	Hard	Imports (current prices)	Y	35 to 39	2–2	OECD
9	Hard	Export price index	N	43 to 47	3–2	CZSO
10	Hard	Import price index	N	43 to 47	3–2	CZSO
11	Financials	CZK/EUR exchange rate	N	0	1–0	CNB
12	Financials	M2	Y	30 to 31	2–1	OECD
13	Financials	Credit	Y	30 to 31	2–1	CNB MB
14	Financials	3M PRIBOR	N	0	1–0	CNB
15	Financials	1Y PRIBOR	N	0	1–0	CNB
16	Financials	PX-50 stock index	N	0	1–0	PSE
17	Financials	Czech gov. bond yield (10Y)	N	0	1–0	CNB
18	Surveys	Consumer confidence	N	-7 to -2	1–0	CZSO
19	Surveys	Industry confidence	N	-7 to -2	1–0	CZSO
20	Surveys	Construction confidence	N	-7 to -2	1–0	CZSO
21	Surveys	Trade confidence	N	-7 to -2	1–0	CZSO
22	Surveys	Services confidence	N	-7 to -2	1–0	CZSO
23	Foreign	EURIBOR 3M	N	0	1–0	ECB
24	Foreign	EURIBOR 1Y	N	0	1–0	ECB
25	Foreign	Oil price (Brent)	N	0	1–0	Datastream
26	Foreign	Ifo business climate Germany	N	-10 to -4	1–0	IFO
27	Foreign	Euro area business climate	Y	-4 to -1	1–0	EC
28	Foreign	Germany exports	Y	40	2–2	OECD

Notes: Rev. indicates whether a variable is typically revised, Pub. Lag stands for publication lag and indicates the typical publication delay of a variable in days (based on 2005–2012 publication calendars), and Unb. Pat. stands for unbalancedness patterns and indicates the number of missing observations for the middle of the month and the end of the month, respectively; for GDP (because it is released quarterly) the numbers correspond to the first, second, and third month of each quarter. CZSO denotes the Czech Statistical Office, CNB denotes the Czech National Bank's ARAD Database, CNB MB denotes the Czech National Bank's Monetary Statistics Monthly Bulletin, PSE denotes the Prague Stock Exchange, ECB denotes the European Central Bank's Statistical Data Warehouse, MLSA denotes the Ministry of Labor and Social Affairs, OECD denotes the OECD Real-time Database, and EC denotes the European Commission. All indicators except for GDP are at monthly frequency. All of the variables are in logarithms and differenced, except for the industry, construction, trade, and services confidence indicators, which are differenced only.

GDP data are released approximately ten weeks after the end of the reference quarter (in the first half of the third month of the subsequent quarter). Most of the hard data are published with varying delays ranging from one to seven weeks. On the other hand, with the exception of money and credit aggregates the financial variables are available with no lag. The surveys are, in fact, published several days before the end of the reference month. Details about the variables used, including their publication lags and sources, are summarized in Table 5.1.

5.3.2 Design of the nowcasting exercise

Our nowcasting exercise is designed as follows. We perform 31 nowcasting rounds, starting with 2005Q1. For each quarter we perform 14 forecast updates, which reflect the arrival of new information over time. Throughout the text we will refer to forecast origins during the preceding quarter (Q(-1)) as forecasting, those during the current quarter (Q(0)) as nowcasting, and those during the following quarter (Q(+1)) as backcasting. The first forecast is performed in the middle of the first month of the preceding quarter (Q(-1)M1 mid). We update the forecasts in the middle and at the end of each month. The last forecast update is performed at the end of the first month of the following quarter (Q(+1)M1 end). We do not perform any additional update, as the preliminary (flash) estimate of GDP is released in the first half of the second month of the following quarter.[7] Since at the time of writing this paper (December 2012), only 2012Q3 GDP growth is available for evaluation, the last nowcasting round we perform is for 2012Q3.

We could, in principle, perform more updates during a month, i.e., after each publication release. However, in practice trade, industry, construction, unemployment, and the CPI are released early at the beginning of the month – although the relative ordering of publication changes from month to month. Consequently, we prefer to model this as a simultaneous release, since we believe it is closer to reflecting the real-time situation.

As for the evaluation of forecasts, we use both the first release and the latest vintage available (December 2012). The argument for using the former is that the Czech National Bank officially discusses every first release value of GDP and explains the reasons behind the deviations from its nowcast. Therefore, the accuracy of the model with respect to this first release is of importance to the CNB. On the other

[7] The correlation between the preliminary and first releases and the preliminary and final releases of GDP over the 2007Q4–2012Q2 period is 0.95 and 0.84, respectively.

hand, the latest vintage data are arguably closest to reflecting the "true" value of GDP growth. As a result, we opt for using both series to evaluate the accuracy of the nowcasts.

Given that our time series dimension is rather short (beginning in Jan 2000) we opt for a parsimonious specification with regard to the number of factors and lags. We model comovements with one factor and the dynamics of the factor with two lags. While specifications with one or three lags give virtually same results, increasing the number of factors results in deterioration of the forecasting accuracy (see Section 5.C for more details).

5.3.3 In-sample properties

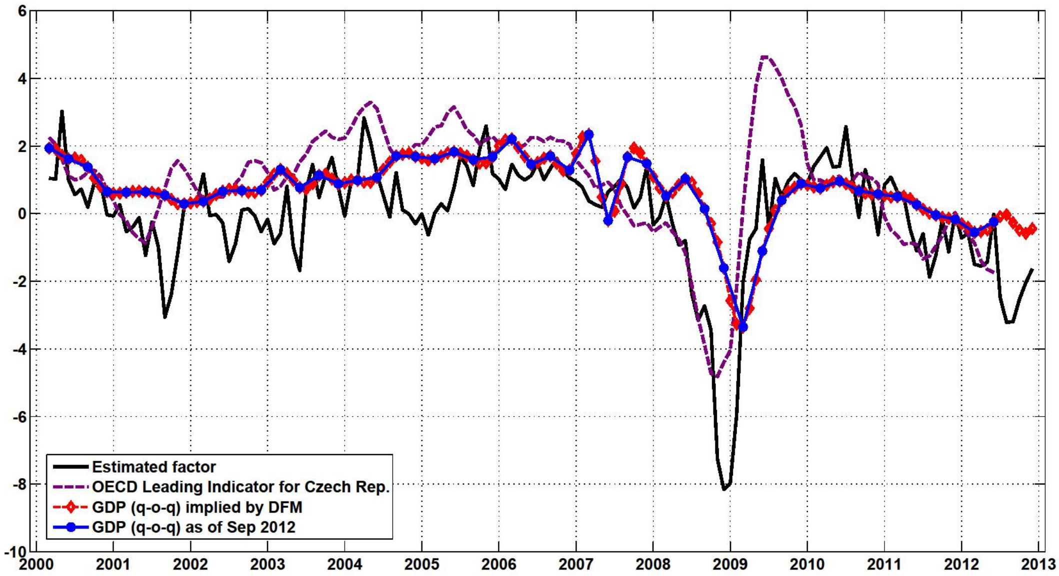
Before presenting the results of the out-of-sample exercise, we describe here several properties of the DFM estimated with the data available in September 2012, which corresponds to Q(0) M3 end of our last nowcasting round.

Figure 5.1 presents the estimated factor, which reflects the common element that drives the comovements of the variables included in our model. We compare the factor to the leading indicator produced by the OECD, which is designed to predict turning points in the Czech business cycle relative to the trend.[8] Overall, the factor and the OECD leading indicator are very similar and it seems that both track the business cycle dynamics in the Czech Republic quite well.

Next, to get more insight into the forces driving the DFM nowcasts we report the estimated loadings in Figure 5.2. Note that the loadings reflect mostly contemporaneous correlations, and we make no attempt to establish the causality. The loadings indicate that most of the series are procyclical, while unemployment, the exchange rate, export and import prices, and the government bond yield seem countercyclical. Except perhaps for the exchange rate, the loadings are in line with what one might

[8]The components of the OECD leading indicators are: the balance of payments, demand and production evolution surveys, the CPI, consumer confidence, exports, and share prices. For more details see http://stats.oecd.org/mei/default.asp?lang=e&subject=5&country=CZE. We present the vintage of the leading indicator as of September 2012. To facilitate comparison, we present the monthly growth rates of the indicator scaled by the mean and standard deviation of the factor estimated by the DFM.

Figure 5.1: Factor estimated in Sep 2012



expect a priori about the contemporaneous correlations with the business cycle.[9] As for the relative magnitudes, the foreign variables have the largest loadings. Notably, the trade variables along with interest rates also have high magnitudes. On the other hand, construction, M2, and government bond yields possess rather small loadings, but we prefer to keep them in the model since we do not want to select variables based on in-sample measures.

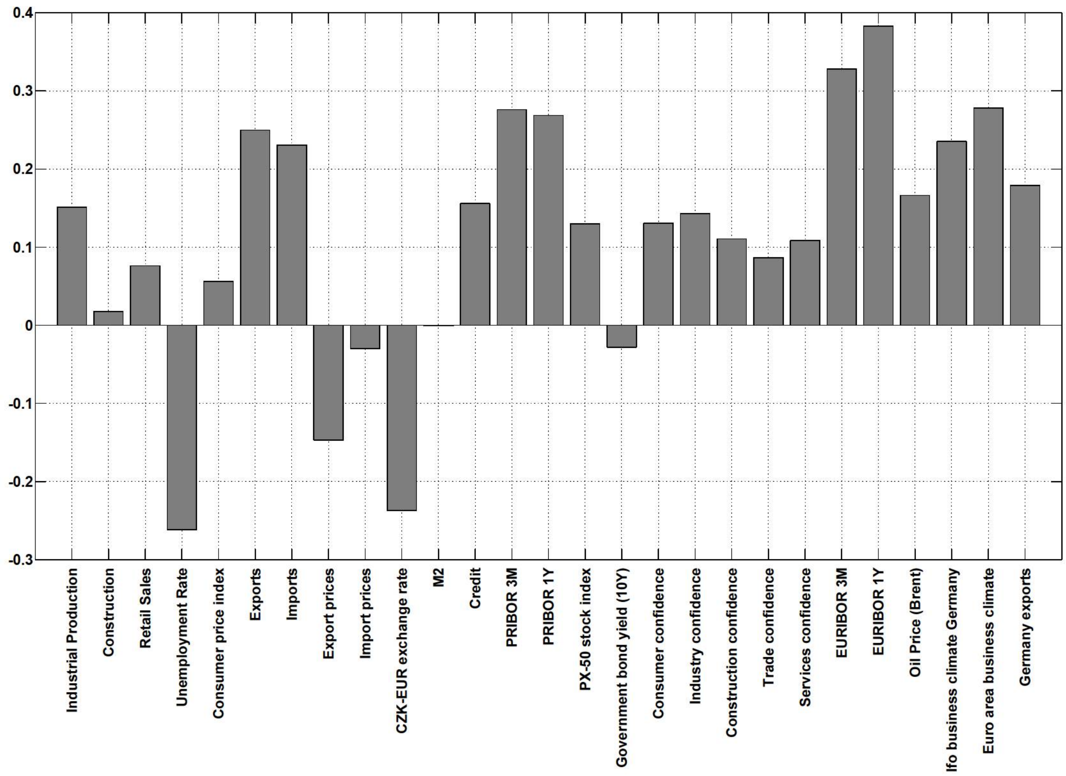
5.3.4 Nowcasting performance

Figure 5.3 reports the results from our real-time nowcasting exercise. For each forecasting round, i.e.,for each of the 14 different forecast origins starting from the middle of the first month of the preceding quarter until the end of the first month of the following quarter, we plot the corresponding root mean square error (RMSE). The RMSE gives us an idea of the out-of-sample forecast uncertainty that is tied to a given forecast origin.

First, we consider several naive benchmarks: a model where the last available

[9]The exchange rate is defined as the Czech koruna against the euro, hence an increase corresponds to a depreciation of the currency. Since there might be delays between the time of the exchange rate shock and the effect on trade or the economy as a whole, the negative contemporaneous correlation might be plausible. Alternatively, the loading might be a consequence of the fact that the Czech currency typically depreciates when investors are expecting an overall deterioration in economic activity in the region.

Figure 5.2: Loadings estimated in Sep 2012



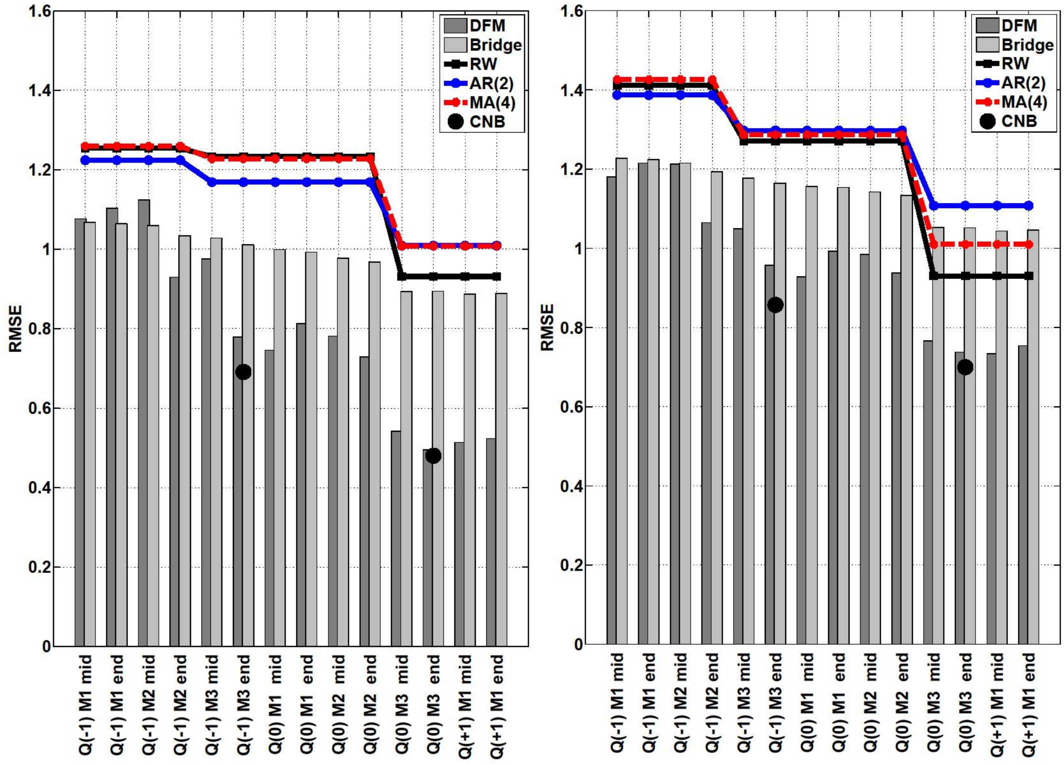
growth is a new forecast (random walk (RW)), an autoregressive model of order two (AR(2)), and a moving average of the last four available quarters (MA(4)). We observe that the DFM performs better than any of the naive benchmarks. On the whole, it seems that with the arrival of new information the forecasting errors seem to decrease, although not always. This fact was also shown succinctly by Lahiri & Monokroussos (2013).

We also compare the performance of the DFM to so-called bridge-equation models, which are the tools traditionally used in central banks (Kitchen & Monaco 2003; Baffigi *et al.* 2004).[10] In Figure 5.3, we present the RMSE of the mean of the individual bridge equation forecasts. We observe that the bridge equations become more precise with more information, but cannot really compete with the DFM. They are able to beat the naive benchmarks, except for the forecast origins at the end of the current quarter and the beginning of the next quarter, where they seem to perform worse than the RW and MA(4) benchmarks. The relatively worse performance of

[10]More information about the specification of the bridge equations is provided in the Appendix.

Figure 5.3: Root mean square errors of different forecasts

(a) Evaluated using first releases of GDP growth (b) Evaluated using latest vintage (Dec 2012)



bridge equation models is worth noting. This result might be the consequence of the fact that we only use bivariate specification, where in each model only one indicator and GDP growth is estimated and the models are than equally weighted. This suggests that more complicated model (such as DFM) that is able to capture the dynamic cross-correlations of the data might be needed to improve predictability of Czech GDP. See also Brunhes-Lesage & Darné (2012) for a comparison of forecasts from bridge and factor models.

In Figure 5.3, we also plot the RMSE of the CNB nowcasts. The CNB nowcasts are taken from the final forecast books that are prepared regularly by the Monetary Policy Department for the quarterly Situation Report. The CNB produces its GDP nowcast at the end of the last month of the reference quarter (Q(0)M3 end). The CNB nowcast is produced by a model that consists of a set of equations of expenditure components, estimated at quarterly frequency. The CNB nowcasts are adjusted by expert judgment, typically reflecting the latest developments of leading indicators or other subjective evaluation (see Arnostova *et al.* 2011, for more details). We also compare the accuracy of the one-quarter-ahead forecasts (Q(-1)M3 end). Overall, the performance of the model-based DFM nowcasts is comparable to that of the nowcasts of the CNB, while at the one-quarter-ahead horizon the DFM seems to fare rather worse than the CNB. While the CNB nowcasts are a result of a model and expert judgment, the nowcasts produced by the DFM are entirely model-based without imposing subjective judgment. The comparative performance is therefore good news, since the DFM nowcast might serve as a good cross-check of the nowcast.

Figure 5.4 presents the nowcasts made by the CNB and the DFM over the 2005–2012 period (nowcasts from Q(0)M3 end forecast origins). The first release GDP growth and the growth as of the latest available vintage are also plotted. The figure suggests that the nowcasts by the CNB and the DFM are very similar in the first half of the evaluation sample, while in the second half they often seem to point in different directions. This is likely the consequence of the increased overall uncertainty in the period after the global financial crisis.

Figure 5.4: Quarterly GDP growth and its nowcasts as of Q(0)M3 end

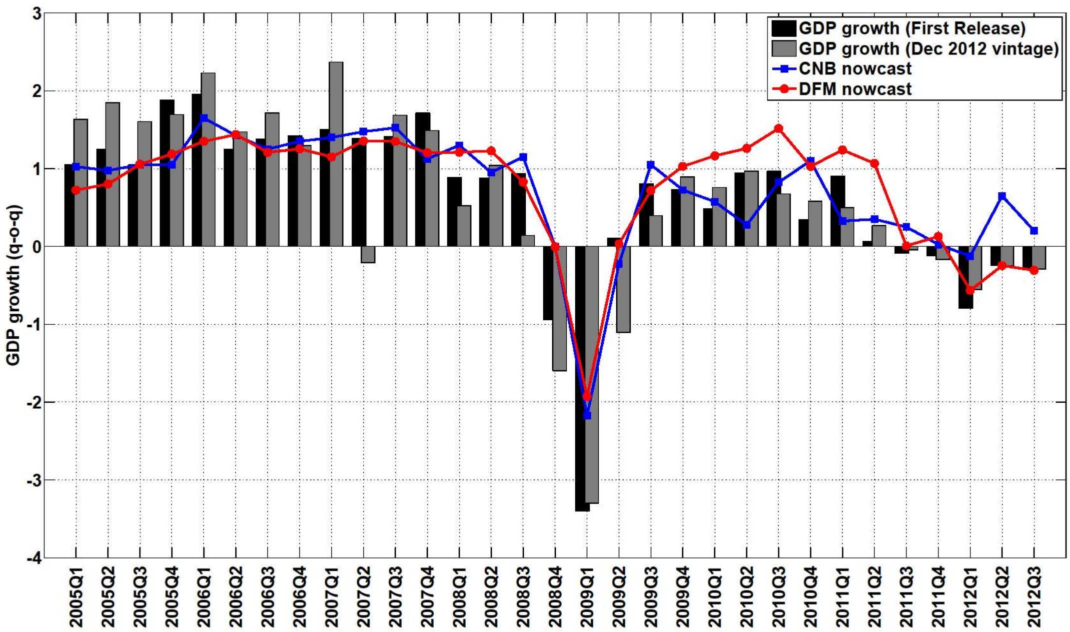


Table 5.2 reports the performance of the DFM and the CNB over the whole sample

and two subsamples: a pre-crisis subsample covering the 2005Q1–2008Q2 period and a crisis subsample covering 2008Q3–2012Q3.[11] We present the performance relative to the random walk. The performance of the three naive models is very similar, but we choose the random walk as the benchmark since it has the best performance for the horizon when the CNB nowcasts are produced (Q(0)M3 end).

The average forecasting error of the naive random walk model is 0.93 for the nowcast and almost 1.3 for the one-quarter-ahead horizon. The DFM and CNB are able to reduce the average forecast errors by 30 to 50% relative to the naive RW model. The results also suggest that indirect pooling of information as represented by the bridge equations is not as successful as direct pooling within a single dynamic factor model: the gains in the forecasting accuracy of the bridge equations relative to the naive model are small.

The forecasting improvements seem to come mainly from the crisis period, while the improvements in the pre-crisis period are more modest. This is in line with D'Agostino & Giannone (2012), who show that the performance of more complex models relative to simple benchmarks is better during more volatile periods characterized by large comovements. Kuzin *et al.* (2013) note that the forecasting errors of many models are larger in absolute terms during crises and that the improvements in relative performance stem from the fact that the naive benchmarks performed worse. This is also our case: the forecast errors of the naive models are approximately three times higher during crisis periods than in pre-crisis times. More evidence about the pre- and post-crises performance of factor models can be found in Dias *et al.* (2015).

We also report the performance of the combination of the CNB and DFM nowcasts, which is obtained as the simple mean of the two nowcasts. This combination of nowcasts might serve as insurance against uncertain instabilities, an issue even more important during times of crisis (Clark & McCracken 2010; Aiolfi *et al.* 2012). The results suggest that the combination performs better than the CNB or DFM nowcast alone. The gains are highest during the crisis period. This can be due to fact

[11] The subsample split also approximately corresponds to the date of change of the core forecasting model used by the CNB. In 2008, the CNB switched from a quarterly projection model to the G3 DSGE model. For more details, see Andrle *et al.* (2009).

Table 5.2: Root mean square errors

	Full Sample (2005Q1–2012Q3)		Pre-Crisis (2005Q1–2008Q2)		Crisis (2008Q3–2012Q3)	
	Q(-1)M3 end	Q(0)M3 end	Q(-1)M3 end	Q(0)M3 end	Q(-1)M3 end	Q(0)M3 end
Evaluated using first releases of GDP growth						
Random Walk (absolute RMSE)	1.23	0.93	0.48	0.39	1.61	1.21
RMSE relative to RW						
Bridge	0.82	0.96	1.07	1.22	0.80	0.93
DFM	0.63	0.53	1.01	0.94	0.59	0.48
CNB	0.56	0.52	0.94	0.84	0.52	0.48
Combination CNB & DFM	0.54	0.47	0.91	0.86	0.51	0.42
Evaluated using GDP growth in December 2012 vintage						
Random Walk (absolute RMSE)	1.27	0.93	0.78	0.74	1.56	1.06
RMSE relative to RW						
Bridge	0.92	1.13	1.08	1.12	0.88	1.13
DFM	0.75	0.79	1.05	1.04	0.68	0.67
CNB	0.67	0.74	1.02	0.96	0.58	0.65
Combination CNB & DFM	0.68	0.74	1.01	0.99	0.58	0.61

Notes: Bridge stands for the nowcast obtained as the average of the nowcasts from the individual bridge equations. *DFM* stands for the nowcast obtained from the dynamic factor model. *CNB* stands for the official nowcast of the Czech National Bank. *Combination CNB & DFM* stands for the nowcast obtained as the simple mean of the CNB and DFM nowcasts.

that forecast errors show different degree of correlation within the two subsamples. While in the pre-crisis period the forecast errors seem to be rather correlated, in the crisis period they are frequently going in different directions. This might be the consequence of the increased overall uncertainty in the period following the global financial crisis.

The success of the combination suggests that the purely model-based DFM might add value to the CNB nowcasts in the sense that it contains useful information missing from the CNB nowcasts. We further investigate this issue formally by running the following regression:

$$y_t = \alpha + \beta_1 \widehat{y}_t^{CNB} + \beta_2 \widehat{y}_t^{DFM} + \varepsilon_t, \tag{5.8}$$

where $\widehat{y}_t^{CNB}$ denotes the CNB forecast and $\widehat{y}_t^{DFM}$ denotes the model-based DFM forecast. Similar regressions are typically employed in the literature (Romer & Romer 2000; Bjornland *et al.* 2012).

The results in Table 5.3 suggest that the DFM could possibly have added value to the CNB nowcasts: conditional on the CNB's forecasts GDP growth turns out to be higher when the DFM nowcast is higher. The subsample results, however, suggest that this result was limited to the crisis period.

Table 5.3: Does the DFM add value to the CNB's GDP nowcasts?

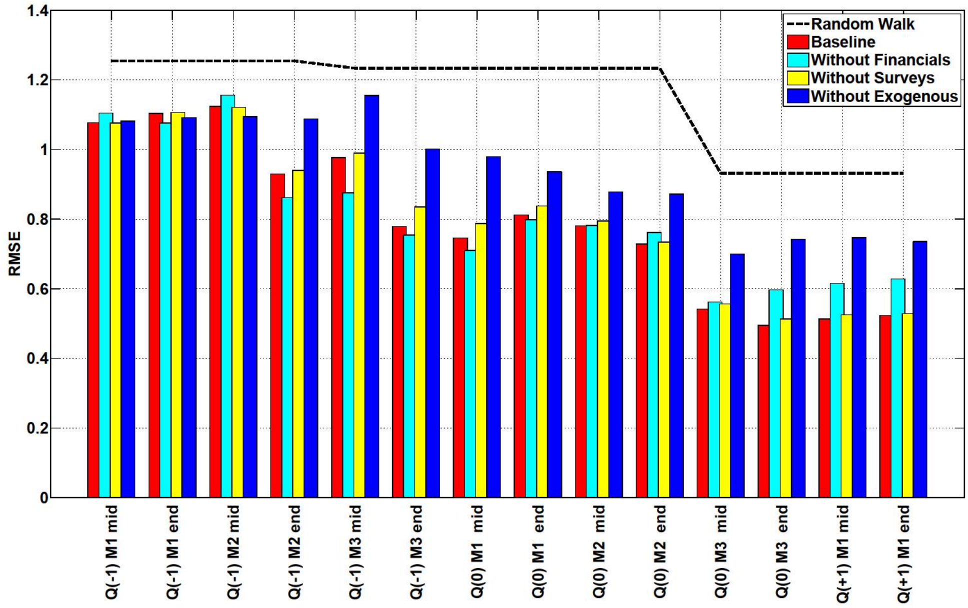
	$y_t = \alpha + \beta_1 \widehat{y_t}^{CNB} + \beta_2 \widehat{y_t}^{DFM} + \varepsilon_t$					
	Full Sample (2005Q1–2012Q3)		Pre-Crisis (2005Q1–2008Q2)		Crisis (2008Q3–2012Q3)	
	Q(-1)M3 end	Q(0)M3 end	Q(-1)M3 end	Q(0)M3 end	Q(-1)M3 end	Q(0)M3 end
Dependent variable: first releases of GDP growth						
α	-0.46*	-0.39***	1.14*	0.57**	-0.58*	-0.43***
	(0.27)	(0.06)	(0.57)	(0.22)	(0.28)	(0.07)
β_1	0.84***	0.72***	0.41	0.53	0.58*	0.61***
	(0.21)	(0.08)	(0.37)	(0.39)	(0.29)	(0.15)
β_2	0.71***	0.64***	-0.21	0.11	0.78***	0.68***
	(0.17)	(0.06)	(0.45)	(0.41)	(0.24)	(0.09)
R^2	0.67	0.90	0.13	0.17	0.57	0.90
Dependent variable: GDP growth in December 2012 vintage						
α	-0.53	-0.46***	2.48***	2.12**	-0.73*	-0.57***
	(0.36)	(0.14)	(0.30)	(0.79)	(0.39)	(0.19)
β_1	0.92***	0.86***	0.06	0.94	0.53**	0.57***
	(0.29)	(0.19)	(0.41)	(0.55)	(0.23)	(0.18)
β_2	0.69***	0.56***	-1.01**	-1.56**	0.81**	0.69***
	(0.24)	(0.16)	(0.36)	(0.69)	(0.28)	(0.22)
R^2	0.57	0.74	0.11	0.12	0.55	0.85

Notes: Autocorrelation and heteroskedasticity robust standard errors (Newey & West 1987) in parenthesis.

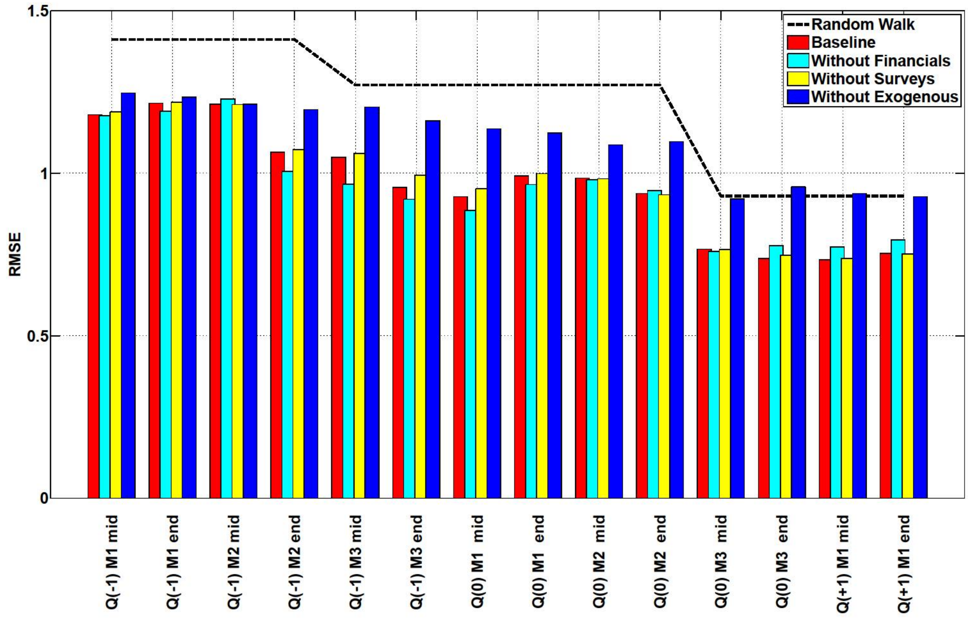
Finally, we look at the role of financials, surveys, and foreign variables for the performance of the DFM. Figure 5.5 presents the out-of-sample RMSE for models that exclude different groups of variables. Looking at the figure, several observations stand out. First, excluding surveys leaves the accuracy of the model intact, except perhaps for the early forecast origins. Second, the role of the financial variables is ambiguous. Finally, the foreign variables seem to be crucial for the performance of the model, since their exclusion from the model results in larger errors, consistently so across different forecasting origins. Our results are robust with regard to the actual series used for evaluating the forecasts. Our results corroborate the findings of Liu *et al.* (2012) who also find that foreign indicators are useful in improving forecast accuracy.

Figure 5.5: Performance of the model when excluding various groups of data

(a) Evaluated using first releases of GDP growth



(b) Evaluated using latest vintage (Dec 2012)



Note also that the role of the various groups of variables is different for different forecast origins, i.e., the timeliness of the variables matters. In our case, there is some role for surveys at the beginning of the nowcasted quarter (Q(0)M1), when little data for that quarter is actually available. In the case of financial variables, excluding them seems to actually decrease the forecast errors during forecasting (Q(-1)). See also Ferrara *et al.* (2014), who study the role of financial variables for the growth forecasts in more detail. On the other hand, they seem to be important during nowcasting (Q(0)) and backcasting (Q(+1)). The role of timeliness was also clearly demonstrated by Lahiri & Monokroussos (2013).

The importance of foreign variables is not surprising, as the Czech Republic is a small open economy (the share of exports and imports in GDP was roughly 146% in 2011).[12] Previous studies employing dynamic factor models seem to suggest that in different countries the inclusion of different blocks of variables is crucial for the accuracy of the DFM. Bańbura *et al.* (2013) and Bańbura & Rünstler (2011) show that the role of surveys is crucial using US and euro area data, respectively. Aastveit & Trovik (2012) find that the inclusion of foreign variables has a negative impact on the performance of the model using Norwegian data, while financial variables seem to be key to the accuracy of their model. Matheson (2010) finds that excluding surveys worsens the nowcasting performance of the DFM in New Zealand, while Yiu & Chow (2011) find that excluding interest rates increases the forecast errors of the DFM for China. Note, however, that none of the above-mentioned countries is as open as the Czech Republic. The share of exports and imports in GDP for Norway, the most open of these countries, is approximately 70% in 2011, barely half of the Czech Republic's figure. So, it is quite plausible that the shocks hitting the export-dependent Czech economy are different in nature and magnitude from those hitting more closed economies (e.g., terms of trade shocks).

Note that we also performed Diebold-Mariano tests to evaluate the statistical significance of the differences in the predictive abilities of the DFM and CNB nowcasts (Diebold & Mariano 1995).[13] However, the test results pointed in almost all cases to statistically insignificant (at the 5% level) differences between the competing models. This is not surprising given the small evaluation sample of only 31 observations.[14]

[12]See OECD Factbook 2013, available at http://dx.doi.org/10.1787/factbook-2013-en.

[13]We used the Newey & West (1987) estimator of the long-run variance of the difference between the squared prediction errors.

[14]See also Ashley (2003), who points out that typically more than 100 observations are needed to establish statistically significant differences in forecasting ability across models.

5.4 Interpreting new data releases through the lens of the DFM

In this section, we use the methodology of Bańbura & Modugno (2010) to show how the nowcasting framework can be used to read the flow of data releases through lens of the dynamic factor model. It is of interest to know the sources of changes in the nowcast that occur after new data are released. For example, when newly released data about the euro area business situation are worse than expected, the model-based nowcast of the GDP will be revised down. Because the DFM produces forecasts for all variables, we can precisely decompose the changes in the nowcasts. Similar decompositions are regularly used in central banks (ECB 2008; Bundesbank 2009) to complement their real-time nowcasting exercises with story-telling. In fact, the CNB performs similar decompositions for the interest rate within the core model (Andrle *et al.* 2009). But the core model is geared towards producing medium to long-term predictions, so it cannot be used directly to decompose the changes in the GDP nowcasts as a result of newly published data.

We denote Ω_v as the information set at the release v and $\mathbb{D}$ as the set of parameters estimated on the information set Ω_v. Further, we denote $\widetilde{\Omega}_{v+1}$ as the information set with the same unbalancedness pattern as Ω_v, but using the latest data vintage.

We can then decompose the change of the nowcast into three parts: the effect of re-estimation, the effect of data revisions, and the effect of *news*.

1. **The effect of re-estimation** is computed as the difference between the nowcast obtained using the old information set Ω_v using the new parameters $\mathbb{D}_{v+1}$ and the nowcast obtained using the old information set Ω_v and the old parameters $\mathbb{D}_v$:

$$\mathbb{E}[y|\Omega_v, \mathbb{D}_{v+1}] - \mathbb{E}[y|\Omega_v, \mathbb{D}_v].$$

2. **The effect of data revisions** is computed as the difference between the nowcast obtained using the new information set with the same unbalancedness

pattern as the old one $\widetilde{\Omega}_{v+1}$ and the new parameters $\mathbb{D}_{v+1}$ and the nowcast obtained using the old information set Ω_v and the new parameters $\mathbb{D}_{v+1}$:

$$\mathbb{E}[y|\widetilde{\Omega}_{v+1}, \mathbb{D}_{v+1}] - \mathbb{E}[y|\Omega_v, \mathbb{D}_{v+1}].$$

3. **The effect of *news*** (the unexpected component of the released data):

$$\mathbb{E}[y|\Omega_{v+1}, \mathbb{D}_{v+1}] - \mathbb{E}[y|\widetilde{\Omega}_{v+1}, \mathbb{D}_{v+1}].$$

In computing the effect of *news* we follow Bańbura & Modugno (2010). They show that one can find coefficients $\delta_{j,v+1}$ such that:

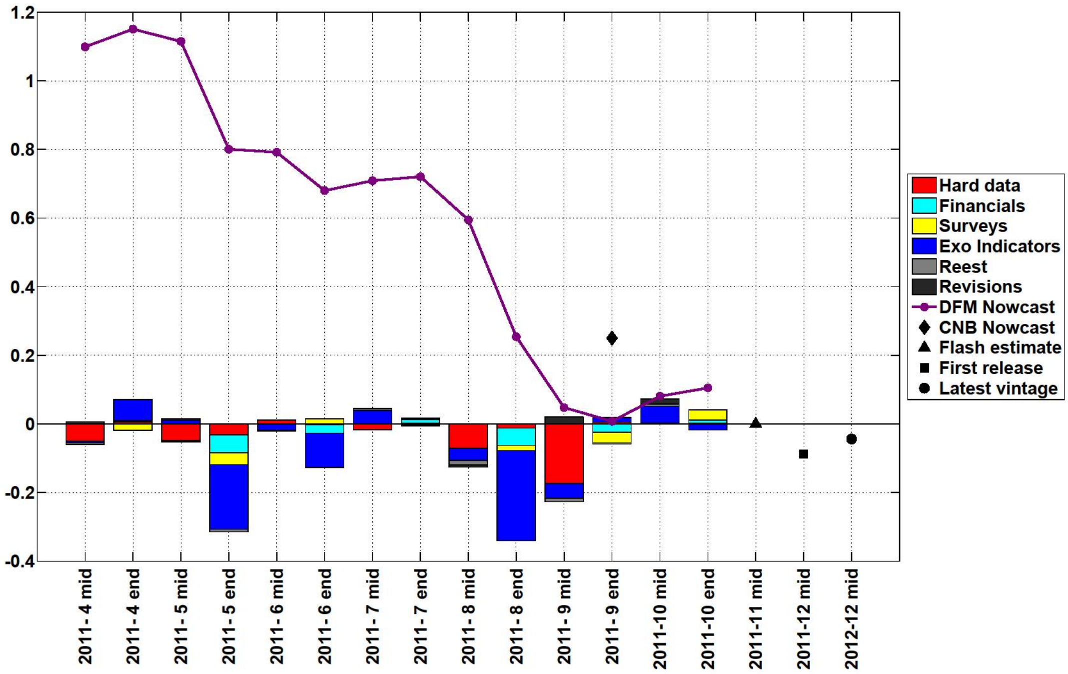
$$\mathbb{E}[y|\Omega_{v+1}, \mathbb{D}_{v+1}] - \mathbb{E}[y|\widetilde{\Omega}_{v+1}, \mathbb{D}_{v+1}] = \sum_j \delta_{j,v+1}(x_{j,T_{j,v+1}} - \mathbb{E}[x_{j,T_{j,v+1}}|\widetilde{\Omega}_{v+1}, \mathbb{D}_{v+1}]). \tag{5.9}$$

The nowcast revision is a weighted average of the *news*. The resulting revision stemming from a release of new data depends on the size as well as the weight of the given variable.

With equation (9) at hand, we are now able to use the lens of the dynamic factor model to interpret the information from the new data releases. Figure 5.6 presents the evolution of the nowcast as the new information arrives. At each nowcast update, we decompose the size of the update into the contributions of re-estimation, revision, and *news* from the respective variables. To keep the exposition clear, we group the news from individual variables into groups.

We illustrate the contribution of news to the nowcast updates on the example of 2011Q3 Czech GDP growth. We choose this example because the second quarter of 2011 marked the peak of the recovery from the 2009 recession. At the beginning of the preceding quarter the nowcast for 2011Q3 is still rather optimistic, probably reflecting the lack of data corresponding to the 2011Q3 period. The first sizeable downward update of the nowcast is caused by the release of data at the end of May

Figure 5.6: Contribution of news to nowcast updates for 2011Q3 Czech GDP growth (q-o-q)



2011: all of the new data point to a worsening of economic activity. This probably stems from the sharp rises in bond yields. This negative news is corroborated by further data releases pointing to a larger deterioration in expected growth. Following additional negative news coming mostly from the foreign indicators in August and from the hard data in September, the nowcast at the end of the quarter points to zero quarterly growth. The flash estimate released in the middle of November confirms the stall of the economy, while the first release of the national accounts even points to negative quarterly growth.

5.5 Further results and robustness checks

5.5.1 Nowcasting the expenditure components of national accounts

Dynamic factor models can be useful in nowcasting other policy-relevant quarterly variables, for example, the expenditure components of GDP. Indeed, several papers have employed dynamic factor models to successfully nowcast the components of GDP (Angelini *et al.* 2010; Godbout & Lombardi 2012; Lahiri *et al.* 2015).

To investigate the performance in the Czech case, we add five expenditure components of GDP to our baseline model: consumption, gross fixed capital formation, government consumption, exports, and imports (all at constant prices).[15] The source of the real-time data is again the OECD Real-Time Database. The CNB forecasts/nowcasts for the expenditure components are again taken from the forecast books prepared for the regular quarterly CNB Situation Reports. The forecasts for the components are available in the forecast books only from 2009Q1, so we confine ourselves to presenting the results for this period only.

Looking at the results, several observations emerge. The accuracy of the GDP forecasts is not worsened by adding additional variables. Table 5.4 presents the results when forecasting one quarter ahead (forecast origin Q(-1)M3 end). The DFM forecasts seem to perform worse than the CNB forecasts for Consumption, GFCF, and Gov. Cons. But the DFM still seems to add value, as suggested by the fact that the forecast combination improves the accuracy of the forecasts. The DFM seems to dominate the CNB when forecasting Exports and Imports.

Table 5.4: Root mean square errors, Forecasting GDP components at Q(-1)M3 end, 2009Q1–2012Q3

	GDP	Cons.	GFCF	Gov. Cons.	Exports	Imports
Evaluated using first releases of GDP growth						
Random Walk (absolute RMSE)	1.65	1.26	4.84	2.09	5.19	5.32
RMSE relative to RW						
DFM	0.53	1.02	0.80	0.99	0.75	0.82
CNB	0.50	0.95	0.69	0.90	0.87	0.96
Combination CNB & DFM	0.46	0.94	0.72	0.78	0.79	0.88
Evaluated using GDP growth in December 2012 vintage						
Random Walk (absolute RMSE)	1.52	1.49	5.08	1.88	4.09	3.85
RMSE relative to RW						
DFM	0.56	0.97	0.61	1.11	0.73	0.75
CNB	0.53	0.86	0.50	0.93	0.84	0.94
Combination CNB & DFM	0.49	0.90	0.52	0.84	0.74	0.82

Notes: DFM stands for the nowcast obtained from the dynamic factor model, and *CNB* stands for the official nowcast of the Czech National Bank. *Combination CNB & DFM* stands for the nowcast obtained as the simple mean of the CNB and DFM nowcasts.

Table 5.5 reports the results for nowcasting the current quarter, i.e., forecast ori-

[15] We could also impose a restriction that would reflect the national account identities. However, Angelini *et al.* (2010) find, using the euro area data, that the improvements from imposing this constraint are rather modest.

Table 5.5: Root mean square errors, Nowcasting GDP components at Q(0)M3 end, 2009Q1–2012Q3

	GDP	Cons.	GFCF	Gov. Cons.	Exports	Imports
Evaluated using first releases of GDP growth						
Random Walk (absolute RMSE)	1.19	1.01	6.12	2.07	4.86	5.15
RMSE relative to RW						
DFM	0.48	1.14	0.75	0.75	0.96	1.05
CNB	0.47	1.28	0.68	0.85	0.80	0.91
Combination CNB & DFM	0.42	1.10	0.68	0.66	0.75	0.83
Evaluated using GDP growth in December 2012 vintage						
Random Walk (absolute RMSE)	0.90	1.64	5.57	1.84	3.96	3.70
RMSE relative to RW						
DFM	0.69	0.78	0.74	0.86	0.90	1.12
CNB	0.62	0.80	0.63	0.95	0.89	1.04
Combination CNB & DFM	0.58	0.73	0.64	0.75	0.67	0.80

Notes: DFM stands for the nowcast obtained from the dynamic factor model, and *CNB* stands for the official nowcast of the Czech National Bank. *Combination CNB & DFM* stands for the nowcast obtained as the simple mean of the CNB and DFM nowcasts.

gin (Q(-1)M3 end). On the whole, the DFM seems to nowcast better for Consumption and Government Consumption. Note that in the case of exports and imports the dynamic factor model fares worse, but still seems to add value, as combining the DFM and CNB nowcasts decreases the nowcast errors.

Note that one could also perform the *news* exercise with components similar to those presented in the previous section.

5.5.2 Forecasting performance of the DFM at longer horizons

While the DFM is geared towards nowcasting, it might be of interest to evaluate the accuracy at longer horizons as well. Because the variables are transformed to stationarity, the forecast of the DFM at longer horizons will converge to the steady states (historical means). As for the CNB forecasts, these are also converging to the steady states implied by the DSGE model, but they are conditional on expected shocks (largely coming from external developments).

In Table 5.6 we report the average accuracy of the RW, DFM, and CNB forecasts at horizons two to six quarters ahead. These forecasts are produced at the forecast origin (Q(0)M3 end). The gains relative to the naive random walk forecasts are smaller than for nowcasting and one-quarter-ahead forecasting. Furthermore,

Table 5.6: Root mean square errors for longer horizons – 2005Q1–2012Q3

	Forecast				
	2Q ahead	3Q ahead	4Q ahead	5Q ahead	6Q ahead
Evaluated using first releases of GDP growth					
Random Walk (absolute RMSE)	1.29	1.38	1.52	1.56	1.52
RMSE relative to RW					
DFM	0.85	0.85	0.78	0.75	0.79
CNB	0.75	0.77	0.76	0.78	0.84
Combination CNB & DFM	0.77	0.79	0.76	0.75	0.81
Evaluated using GDP growth in December 2012 vintage					
Random Walk (absolute RMSE)	1.44	1.53	1.67	1.70	1.67
RMSE relative to RW					
DFM	0.83	0.85	0.80	0.78	0.80
CNB	0.74	0.78	0.79	0.81	0.86
Combination CNB & DFM	0.76	0.80	0.79	0.79	0.83

Notes: The forecasts are produced at the forecast origin Q(0)M3 end. The first forecast is produced in March 2005 and the last forecast in September 2012. *DFM* stands for the forecast obtained from the dynamic factor model, and *CNB* stands for the official forecast of the Czech National Bank. *Combination CNB & DFM* stands for the nowcast obtained as the simple mean of the CNB and DFM nowcasts.

the results suggest that the accuracy of the forecasts is comparable, with the CNB slightly dominating at the two to three-quarter horizon, while the DFM seems to be slightly more accurate at longer horizons.[16] Combining the forecasts does not result in any apparent improvements. Again, the Diebold-Mariano test of differences in the accuracy of the forecasts indicates no statistical differences between the CNB and DFM forecasts. Since the CNB's monetary policy horizon is four to six quarters ahead, it might be of interest to use forecasts from the DFM as a cross-check even at forecasting horizons beyond the current quarter.

5.6 Concluding remarks

In this paper, we evaluate the real-time accuracy of the nowcasts produced by the dynamic factor model over the 2005–2012 period. We find that the accuracy of the model-based nowcasts is comparable to the nowcasts of the Czech National Bank. The accuracy improves if the two nowcasts are combined. Furthermore, we find that

[16] We also tried a specification that includes outlooks for foreign demand, the foreign PPI, and EURIBOR. There were no improvements in the accuracy of the forecasts. These results are available upon request.

the role of foreign variables is crucial for the performance of the DFM: excluding them results in larger forecast errors. We also show how one can interpret the changes in the nowcasts as news contributions from new data releases. The framework might be useful in nowcasting other variables as well. We demonstrated good performance for nowcasting of the expenditure components of Czech GDP. Finally, the forecasting abilities of the DFM even at longer horizons (up to six quarters ahead) are also competitive with the CNB's forecasts.

Our results are in line with the anecdotal evidence provided by Sims (2002), who documents that the advantage of judgmental forecasts probably stems mainly from their ability to utilize disparate sources of data in real time and is largely limited to the current and one-quarter-ahead horizon. Our results suggest that, indeed, because of the ability of the dynamic factor model to exploit the latest releases of new data, it is able to compete successfully with the CNB forecasts.

Further research could focus on comparing the accuracy of the DFM with other recently developed mixed-frequency models, such as MIDAS (Andreou *et al.* 2012; Kuzin *et al.* 2011) or Mixed Frequency Bayesian VARs (Schorfheide & Song 2012). Moreover, with regard to the current period of increased uncertainty, accounting for stochastic volatility might bring further forecasting improvements (Marcellino *et al.* 2013; Carriero *et al.* 2012).

Finally, note that our analysis focused on the accuracy of point forecasts only. By focusing on the root mean square forecast errors, we assumed that the loss function of policymakers is quadratic or that the world is linear. Therefore, in future research, it might be of interest to focus on characterizing the uncertainty surrounding the nowcasts in a fashion similar to Aastveit *et al.* (2011).

References

Aastveit, K. & T. Trovik (2012): "Nowcasting norwegian GDP: the role of asset prices in a small open economy." *Empirical Economics* **42(1)**: pp. 95–119.

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, & L. A. Thorsrud (2011): "Nowcasting GDP in real-time: A density combination approach." *Working Paper 2011/11*, Norges Bank.

Adolfson, M., M. K. Andersson, J. Lindé, M. Villani, & A. Vredin (2007): "Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks." *International Journal of Central Banking* **3(4)**: pp. 111–144.

Aiolfi, M., C. Capistran, & A. Timmermann (2012): "Forecast Combinations." In M. Clements & D. Hendry (editors), "Forecast Handbook," Oxford.

Alvarez, R., M. Camacho, & G. Pérez-Quirós (2012): "Finite sample performance of small versus large scale dynamic factor models." *CEPR Discussion Papers 8867*, C.E.P.R. Discussion Papers.

Andreou, E., E. Ghysels, & A. Kourtellos (2012): "Forecasting with mixed-frequency data." In M. P. Clements & D. Hendry (editors), "Oxford Handbook of Economic Forecasting," .

Andrle, M., T. Hledik, O. Kamenik, & J. Vlcek (2009): "Implementing the New Structural Model of the Czech National Bank." *Working Papers 2009/2*, Czech National Bank.

Angelini, E., M. Banbura, & G. Rünstler (2010): "Estimating and forecasting the euro area monthly national accounts from a dynamic factor model." *OECD Journal: Journal of Business Cycle Measurement and Analysis* **2010(1)**: pp. 1–22.

Arnostova, K., D. Havrlant, L. Ruzicka, & P. Toth (2011): "Short-Term Forecasting of Czech Quarterly GDP Using Monthly Indicators." *Czech Journal of Economics and Finance (Finance a uver)* **61(6)**: pp. 566–583.

Aruoba, S. B. (2008): "Data Revisions Are Not Well Behaved." *Journal of Money, Credit and Banking* **40(2-3)**: pp. 319–340.

Ashley, R. (2003): "Statistically significant forecasting improvements: how much out-of-sample data is likely necessary?" *International Journal of Forecasting* **19(2)**: pp. 229–239.

Bańbura, M., D. Giannone, M. Modugno, & L. Reichlin (2013): "Now-Casting and the Real-Time Data Flow." *Working Paper Series 1564*, European Central Bank.

Bańbura, M., D. Giannone, & L. Reichlin (2010a): "Large Bayesian vector auto regressions." *Journal of Applied Econometrics* **25(1)**: pp. 71–92.

Bańbura, M., D. Giannone, & L. Reichlin (2010b): "Nowcasting." *Working Paper Series 1275*, European Central Bank.

Bańbura, M. & M. Modugno (2010): "Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data." *Working Paper Series 1189*, European Central Bank.

Bańbura, M. & G. Rünstler (2011): "A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP." *International Journal of Forecasting* **27(2)**: pp. 333–346.

Baffigi, A., R. Golinelli, & G. Parigi (2004): "Bridge models to forecast the euro area GDP." *International Journal of Forecasting* **20(3)**: pp. 447–460.

Bai, J. & S. Ng (2008): "Forecasting economic time series using targeted predictors." *Journal of Econometrics* **146(2)**: pp. 304–317.

Benda, V. & L. Ruzicka (2007): "Short-term Forecasting Methods Based on the LEI Approach: The Case of the Czech Republic." *Research and Policy Notes 2007/01*, Czech National Bank.

Bernanke, B. S. & J. Boivin (2003): "Monetary policy in a data-rich environment." *Journal of Monetary Economics* **50(3)**: pp. 525–546.

Bjornland, H. C., K. Gerdrup, A. S. Jore, C. Smith, & L. A. Thorsrud (2012): "Does Forecast Combination Improve Norges Bank Inflation Forecasts?" *Oxford Bulletin of Economics and Statistics* **74(2)**: pp. 163–179.

Boivin, J. & S. Ng (2006): "Are more data always better for factor analysis?" *Journal of Econometrics* **132(1)**: pp. 169–194.

Brunhes-Lesage, V. & O. Darné (2012): "Nowcasting the French index of industrial production: A comparison from bridge and factor models." *Economic Modelling* **29(6)**: pp. 2174–2182.

Bundesbank (2009): "Short-Term forecasting methods as instruments of business cycle analysis." In "Monthly Report," April 2009, pp. 31–44. Deutsche Bundesbak.

Camacho, M. & G. Perez-Quiros (2010): "Introducing the euro-sting: Short-term indicator of euro area growth." *Journal of Applied Econometrics* **25(4)**: pp. 663–694.

Carriero, A., T. E. Clark, & M. Marcellino (2012): "Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility." *Working Paper 1227*, Federal Reserve Bank of Cleveland.

Chamberlain, G. & M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." *Econometrica* **51(5)**: pp. 1281–304.

Clark, T. E. & M. W. McCracken (2010): "Averaging forecasts from VARs with uncertain instabilities." *Journal of Applied Econometrics* **25(1)**: pp. 5–29.

Croushore, D. (2011): "Frontiers of Real-Time Data Analysis." *Journal of Economic Literature* **49(1)**: pp. 72–100.

D'Agostino, A. & D. Giannone (2012): "Comparing Alternative Predictors Based on Large-Panel Factor Models." *Oxford Bulletin of Economics and Statistics* **74(2)**: pp. 306–326.

De Mol, C., D. Giannone, & L. Reichlin (2008): "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics* **146(2)**: pp. 318–328.

Dias, F., M. Pinheiro, & A. Rua (2015): "Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence." *Economic Modelling* **44(C)**: pp. 266–272.

Diebold, F. X. & R. S. Mariano (1995): "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* **13(3)**: pp. 253–63.

Doz, C., D. Giannone, & L. Reichlin (2011): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering." *Journal of Econometrics* **164(1)**: pp. 188–205.

Doz, C., D. Giannone, & L. Reichlin (2012): "A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models." *The Review of Economics and Statistics* **94(4)**: pp. 1014–1024.

ECB (2008): "Short-term forecasts of economic activity in the euro area." In "Monthly Bulletin," pp. 69–74. European Central Bank.

Edge, R. M., M. T. Kiley, & J.-P. Laforte (2010): "A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model." *Journal of Applied Econometrics* **25(4)**: pp. 720–754.

Engle, R. & M. Watson (1981): "A one-factor multivariate time series model of metropolitan wage rates." *Journal of American Statistical Association* **76**: pp. 774–781.

Evans, M. D. D. (2005): "Where Are We Now? Real-Time Estimates of the Macroeconomy." *International Journal of Central Banking* **1(2)**.

Faust, J., J. H. Rogers, & J. H. Wright (2003): "Exchange rate forecasting: the errors we've really made." *Journal of International Economics* **60(1)**: pp. 35–59.

Faust, J., J. H. Rogers, & J. H. Wright (2005): "News and Noise in G-7 GDP Announcements." *Journal of Money, Credit and Banking* **37(3)**: pp. 403–19.

Fernandez, A. Z., E. F. Koenig, & A. Nikolsko-Rzhevskyy (2011): "A real-time historical database for the OECD." *Globalization and Monetary Policy Institute Working Paper 96*, Federal Reserve Bank of Dallas.

Ferrara, L., C. Marsilli, & J.-P. Ortega (2014): "Forecasting growth during the Great Recession: is financial volatility the missing ingredient?" *Economic Modelling* **36(C)**: pp. 44–50.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2000): "The Generalized Dynamic-Factor Model: Identification And Estimation." *The Review of Economics and Statistics* **82(4)**: pp. 540–554.

Forni, M. & L. Reichlin (1998): "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics." *Review of Economic Studies* **65(3)**: pp. 453–73.

Foroni, C. & M. Marcellino (2013): "A survey of econometric methods for mixed-frequency data." *Norges Bank Working Papers 06/2013*, Norges Bank.

Garratt, A. & S. P. Vahey (2006): "UK Real-Time Macro Data Characteristics." *Economic Journal* **116(509)**: pp. F119–F135.

Giannone, D., L. Reichlin, & L. Sala (2005): "Monetary Policy in Real Time." In "NBER Macroeconomics Annual 2004, Volume 19," NBER Chapters, pp. 161–224. National Bureau of Economic Research.

Giannone, D., L. Reichlin, & D. Small (2008): "Nowcasting: The real-time informational content of macroeconomic data." *Journal of Monetary Economics* **55(4)**: pp. 665–676.

Godbout, C. & M. J. Lombardi (2012): "Short-term forecasting of the Japanese economy using factor models." *Working Paper Series 1428*, European Central Bank.

Groen, J. J., G. Kapetanios, & S. Price (2009): "A real time evaluation of Bank of England forecasts of inflation and growth." *International Journal of Forecasting* **25(1)**: pp. 74–80.

Havranek, T., R. Horvath, & J. Mateju (2012): "Monetary transmission and the financial sector in the Czech Republic." *Economic Change and Restructuring* **45**: pp. 135–155.

Horvath, R. (2012): "Do Confidence Indicators Help Predict Economic Activity? The Case of the Czech Republic." *Czech Journal of Economics and Finance (Finance a uver)* **62**: pp. 398–412.

Ince, O. & D. H. Papell (2013): "The (un)reliability of real-time output gap estimates with revised data." *Economic Modelling* **33(C)**: pp. 713–721.

Kitchen, J. & R. Monaco (2003): "Real-Time Forecasting in Practice: The U.S. Treasury Staff's Real-Time GDP Forecast System." *Business Economics* pp. 10–19.

Kugler, P., T. J. Jordan, C. Lenz, & M. R. Savioz (2005): "GDP data revisions and forward-looking monetary policy in Switzerland." *The North American Journal of Economics and Finance* **16(3)**: pp. 351–372.

Kuzin, V., M. Marcellino, & C. Schumacher (2011): "MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area." *International Journal of Forecasting* **27(2)**: pp. 529–542.

Kuzin, V., M. Marcellino, & C. Schumacher (2013): "Pooling versus model selection for nowcasting GDP with many predictors: empirical evidence for six industrialized countries." *Journal of Applied Econometrics* **28(3)**: p. 392–411.

Lahiri, K. & G. Monokroussos (2013): "Nowcasting US GDP: The role of ISM business surveys." *International Journal of Forecasting* **29(4)**: pp. 644–658.

Lahiri, K., G. Monokroussos, & Y. Zhao (2015): "Forecasting consumption: the role of consumer confidence in real time with many predictors." *Journal of Applied Econometrics* .

Lees, K., T. Matheson, & C. Smith (2007): "Open economy DSGE-VAR forecasting and policy analysis - head to head with the RBNZ published forecasts."

Reserve Bank of New Zealand Discussion Paper Series DP2007/01, Reserve Bank of New Zealand.

Liu, P., T. Matheson, & R. Romeu (2012): "Real-time forecasts of economic activity for Latin American economies." *Economic Modelling* **29(4)**: pp. 1090–1098.

Marcellino, M. & A. Musso (2011): "The reliability of real-time estimates of the euro area output gap." *Economic Modelling* **28(4)**: pp. 1842–1856.

Marcellino, M., M. Porqueddu, & F. Venditti (2013): "Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility." *Temi di discussione (Economic working papers) 896*, Bank of Italy, Economic Research and International Relations Area.

Mariano, R. S. & Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series." *Journal of Applied Econometrics* **18(4)**: pp. 427–443.

Matheson, T. D. (2010): "An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys." *Economic Modelling* **27(1)**: pp. 304–314.

McDonald, C. & L. A. Thorsrud (2011): "Evaluating density forecasts: model combination strategies versus the RBNZ." *Reserve Bank of New Zealand Discussion Paper Series DP2011/03*, Reserve Bank of New Zealand.

Molodtsova, T., A. Nikolsko-Rzhevskyy, & D. H. Papell (2008): "Taylor rules with real-time data: A tale of two countries and one exchange rate." *Journal of Monetary Economics* **55**: pp. S63–S79.

Newey, W. K. & K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* **55(3)**: pp. 703–08.

Orphanides, A. (2001): "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* **91(4)**: pp. 964–985.

ORPHANIDES, A. & S. VAN NORDEN (2002): "The Unreliability of Output-Gap Estimates in Real Time." *The Review of Economics and Statistics* **84(4)**: pp. 569–583.

RÜNSTLER, G., K. BARHOUMI, S. BENK, R. CRISTADORO, A. DEN REIJER, A. JAKAITIENE, P. JELONEK, A. RUA, K. RUTH, & C. VAN NIEUWENHUYZE (2009): "Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise." *Journal of Forecasting* **28(7)**: pp. 595–611.

ROBERTSON, J. C. & E. W. TALLMAN (1998): "Data Vintages and Measuring Forecast Model Performance." *Federal Reserve Bank of Atlanta Economic Review* **b**: pp. 4–20.

ROMER, D. H. & C. D. ROMER (2000): "Federal Reserve Information and the Behavior of Interest Rates." *American Economic Review* **90(3)**: pp. 429–457.

ROSSI, B. & T. SEKHPOSYAN (2010): "Have economic models' forecasting performance for US output growth and inflation changed over time, and when?" *International Journal of Forecasting* **26(4)**: pp. 808–835.

RUSNAK, M. (2013): "Revisions to the Czech National Accounts: Properties and Predictability." *Czech Journal of Economics and Finance (Finance a uver)* **63(3)**: pp. 244–261.

SCHORFHEIDE, F. & D. SONG (2012): "Real-time forecasting with a mixed-frequency VAR." *Working Papers 701*, Federal Reserve Bank of Minneapolis.

SCHUMACHER, C. & J. BREITUNG (2008): "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data." *International Journal of Forecasting* **24(3)**: pp. 386–398.

SHUMWAY, R. H. & D. S. STOFFER (1982): "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm." *Journal of Time Series Analysis* **3(4)**: pp. 253–264.

SIMS, C. A. (2002): "The Role of Models and Probabilities in the Monetary Policy Process." *Brookings Papers on Economic Activity* **33(2)**: pp. 1–62.

Stock, J. & M. W. Watson (2010): "Dynamic Factor Models." In M. Clements & D. Hendry (editors), "Oxford Handbook of Economic Forecasting," Oxford Unversity Press.

Stock, J. H. & M. W. Watson (2002a): "Forecasting Using Principal Components From a Large Number of Predictors." *Journal of the American Statistical Association* **97**: pp. 1167–1179.

Stock, J. H. & M. W. Watson (2002b): "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics* **20(2)**: pp. 147–62.

Stock, J. H. & M. W. Watson (2012): "Generalized Shrinkage Methods for Forecasting Using Many Predictors." *Journal of Business & Economic Statistics* **30(4)**: pp. 481–493.

Wallis, K. F. (1986): "Forecasting with an econometric model: The "ragged edge" problem." *Journal of Forecasting* **5(1)**: pp. 1–13.

Watson, M. W. & R. F. Engle (1983): "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models." *Journal of Econometrics* **23(3)**: pp. 385–400.

Yiu, M. S. & K. K. Chow (2011): "Nowcasting Chinese GDP: Information Content of Economic and Financial Data." *Working Papers 042011*, Hong Kong Institute for Monetary Research.

5.A Description of benchmark models

We denote quarterly GDP growth as y_t. In all cases only the data available at the time of the forecast are used. Therefore, in the following equations, $k = 1$ for forecast origins from *Q(0)M3 mid* to *Q(+1)M1 end*, $k = 2$ for forecast origins from *Q(-1)M3 mid* to *Q(0)M2 end*, and $k = 3$ for forecast origins from *Q(-1)M1 mid* to *Q(-1)M2 end*. The lag for AR process was selected so as to strike balance between adding enough lags to capture important business cycle properties and possible over-parametrization. Additionally, selecting other lags does not change the performance substantially.

Random walk (RW)

$$y_t = y_{t-k} + \varepsilon_t$$

Autoregressive model (AR(2))

$$y_t = \rho_0 + \rho_1 y_{t-k} + \rho_2 y_{t-k-1} + \varepsilon_t$$

Moving average (MA(4))

$$y_t = \frac{1}{4}(y_{t-k} + y_{t-k-1} + y_{t-k-2} + y_{t-k-3}) + \varepsilon_t$$

Bridge equations Forecasting with bridge equations is performed in two steps:

1. First step: Forecasting of monthly indicators to get rid of ragged ends, using an AR process, where the lag is chosen using the AIC.

2. Second step: The monthly predictors are averaged to quarterly frequency and the following equation is estimated:

$$y_t = \alpha + \sum_{i=1}^{k} \beta_i^j(L) x_{it}^j + \varepsilon_t$$

The lag is chosen using the AIC.

5.B State-space representation of the DFM model

Our dynamic factor model can be then cast in a state-space form:

$$\begin{pmatrix} x_t \\ y_t^Q \end{pmatrix} = \begin{pmatrix} \Lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q & 1 & 2 & 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \xi_t^Q \end{pmatrix}$$

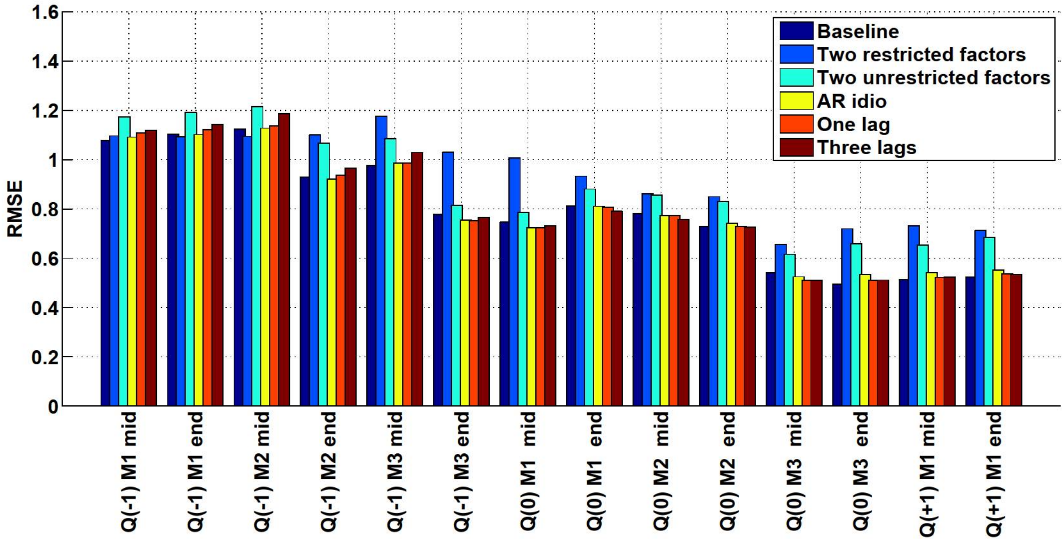
$$\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{pmatrix} = \begin{pmatrix} A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ f_{t-5} \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \\ \varepsilon_{t-5}^Q \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ 0 \\ 0 \\ 0 \\ \varepsilon_t^Q \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

5.C Performance of the DFM under different specifications

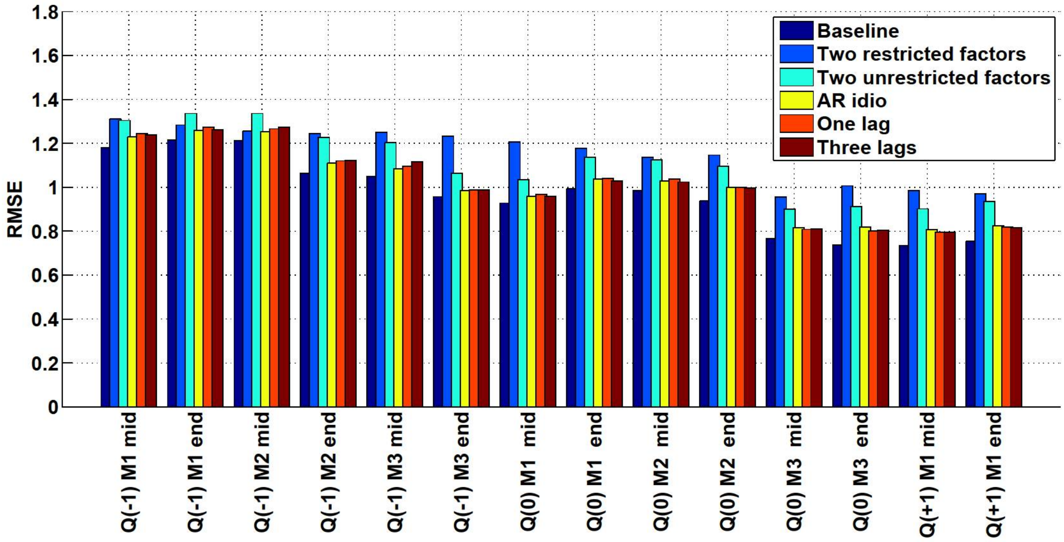
Motivated by the short sample available for the Czech Republic, we largely opted for the simple parsimonious specification of our dynamic factor model. While the results of our baseline model seem to be satisfactory and comparable to the CNB nowcasts, it might be of interest to investigate the sensitivity of the results to the specification of the number of factors or the number of lags. Furthermore, we also consider several extensions, such as modeling the dynamics of the idiosyncratic component or restricting the factors to a domestic and a foreign one.

Figure 5.7: Performance of the DFM model under various specifications

(a) Evaluated using first releases of GDP growth



(b) Evaluated using the latest vintage (Dec 2012)



In Figure 5.7, we present the results for several variations of the baseline model. First, we consider the possibility that Czech GDP is driven by two distinct factors: a factor extracted from domestic variables (Hard data, Financials, Surveys) and a factor extracted from foreign variables. This specification is labeled *Two restricted factors.* The restrictions are imposed as zeros in the loadings matrix. Next, we consider two factors, but both of them are extracted from all of the monthly indicators. This specification is labeled *Two unrestricted factors.* As an additional extension, we consider modeling the idiosyncratic shock ε_t as an autoregressive process of order one, to capture possible persistence in these shocks. This specification is labeled *AR idio.* Finally, we consider two variations of the modeling of the factor dynamics: *One lag* and *Three lags* denote the specification where the factor follows an autoregressive process of order one and three, respectively.

The results suggest that the results of various specifications are comparable with the baseline model. Specifications with two factors seem to perform slightly worse, while modeling the dynamics of factors and idiosyncratic components matters only marginally.

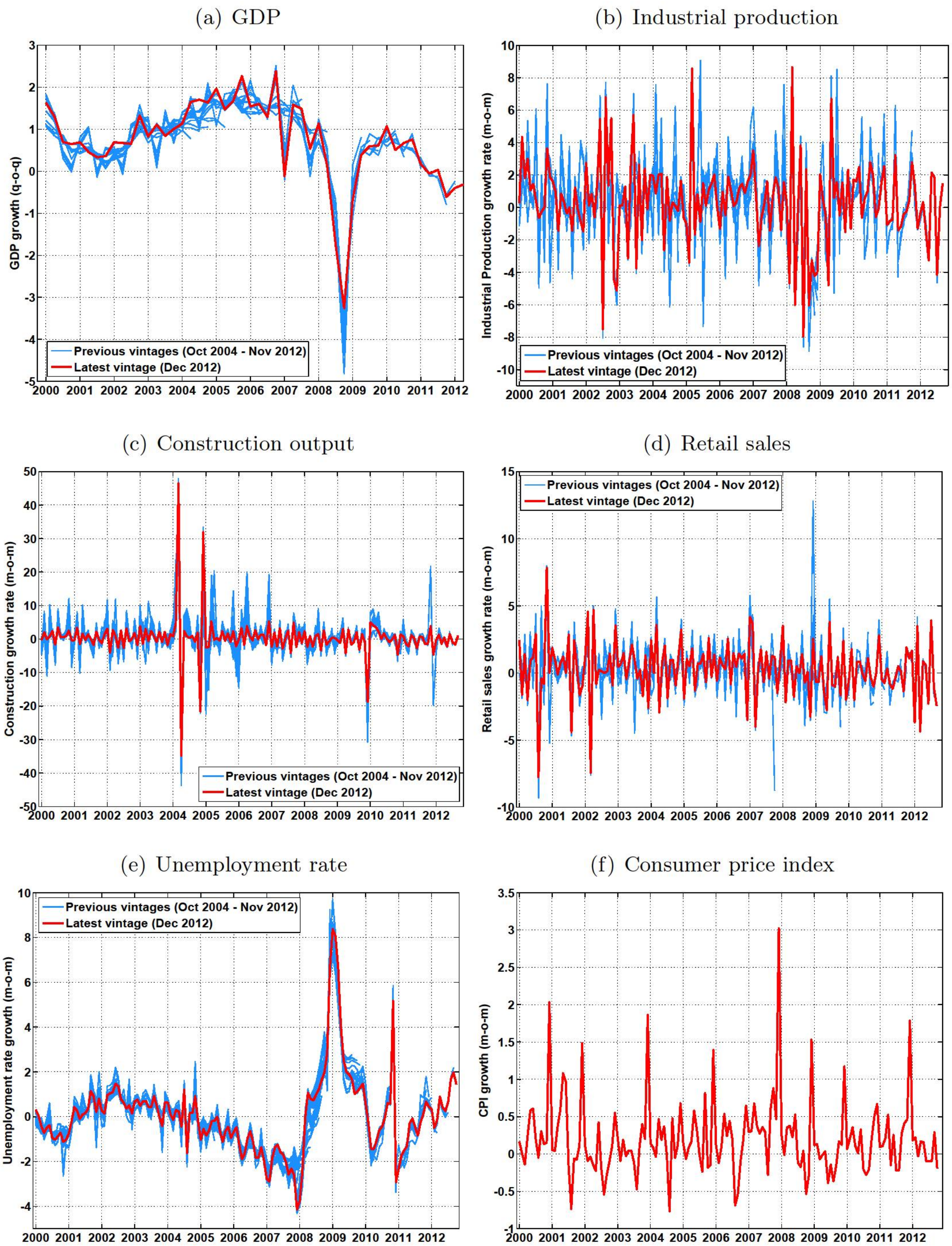
It is also worth checking how does the baseline model performs when we use last available data as opposed to the real-time vintages. Table 5.7 shows that the forecasting performance is not significantly better when one uses the last vintage data instead of real-time vintages, which suggests that the model is robust to the revisions (Giannone *et al.* 2008, find similar results).

Table 5.7: RMSE (last-vintage data vs. real-time data)

RMSE	Q(-1) M1 mid	Q(-1) M1 end	Q(-1) M2 mid	Q(-1) M2 end	Q(-1) M3 mid	Q(-1) M3 end	Q(0) M1 mid	Q(0) M1 end	Q(0) M2 mid	Q(0) M2 end	Q(0) M3 mid	Q(0) M3 end	Q(+1) M1 mid	Q(+1) M1 end
Real-time	1.18	1.22	1.21	1.07	1.05	0.96	0.93	0.99	0.98	0.94	0.77	0.74	0.73	0.75
Last-vintage	1.17	1.21	1.21	1.07	1.04	0.96	0.93	1.00	0.99	0.95	0.75	0.73	0.73	0.76
Rel. RMSE	0.99	1.00	1.00	1.01	0.99	1.00	1.01	1.00	1.01	1.01	0.99	1.00	1.00	1.00

5.D Data used in the nowcasting exercises

Figure 5.8: Data



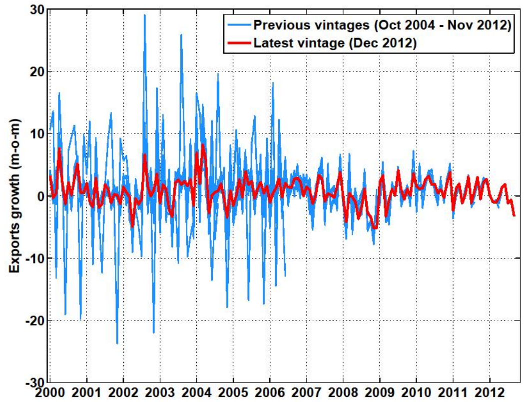
(a) GDP (b) Industrial production

(c) Construction output (d) Retail sales

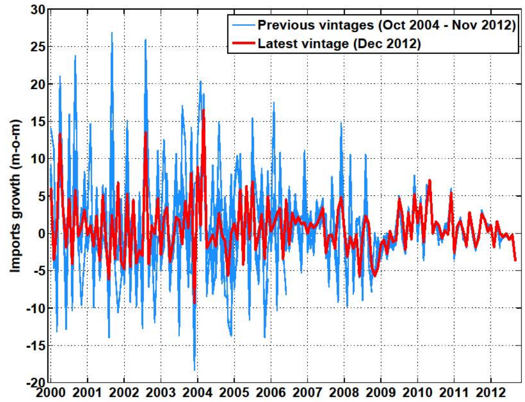
(e) Unemployment rate (f) Consumer price index

Figure 5.9: Data (continued)

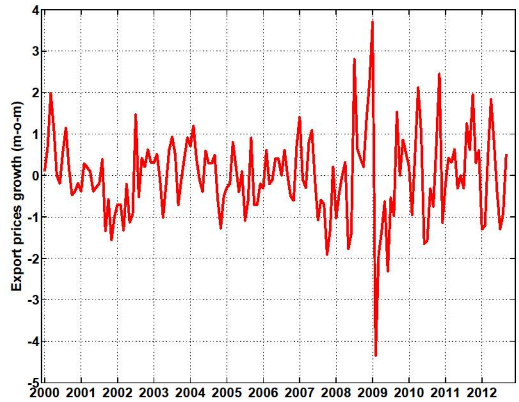
(a) Exports (current prices)



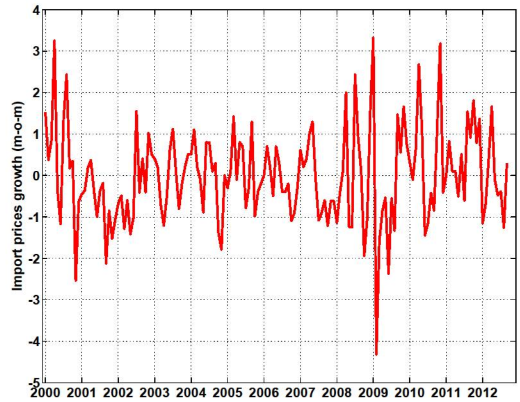
(b) Imports (current prices)



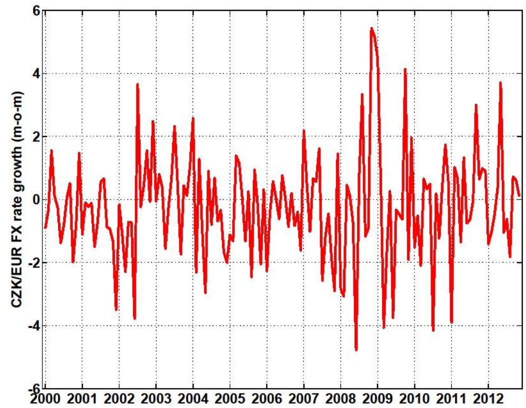
(c) Export prices



(d) Import prices



(e) CZK/EUR exchange rate



(f) M2

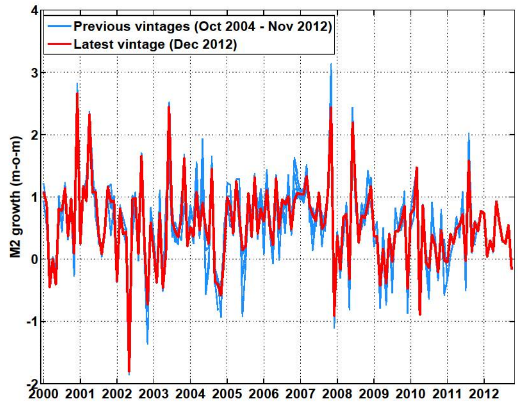
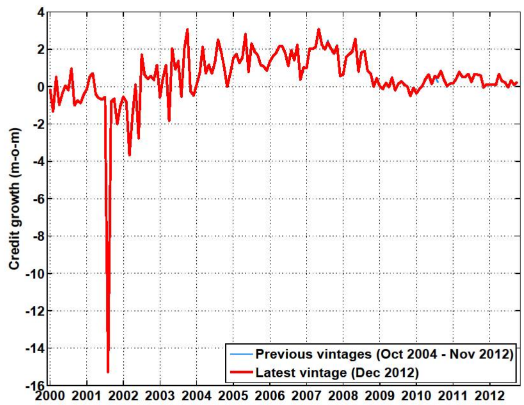
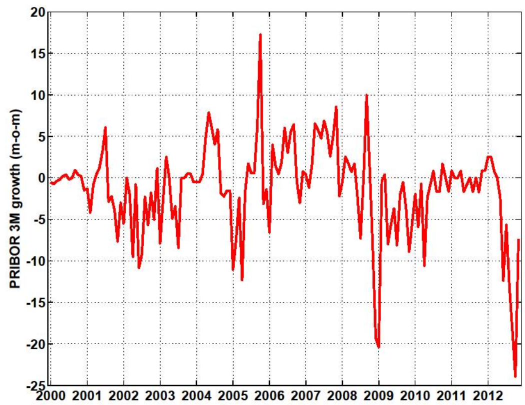


Figure 5.10: Data (continued)

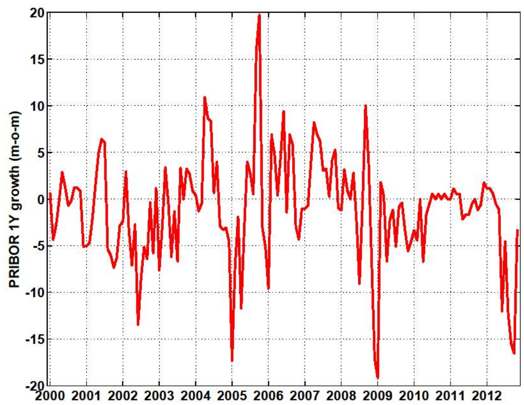
(a) Credit



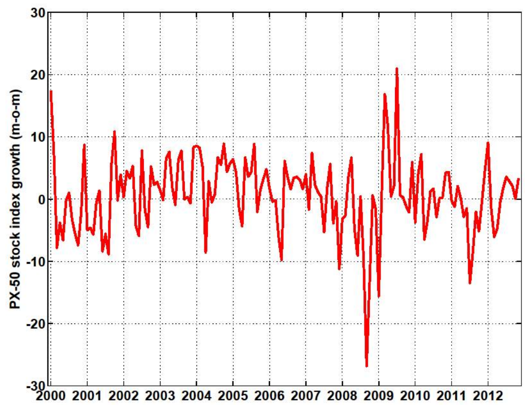
(b) 3M PRIBOR



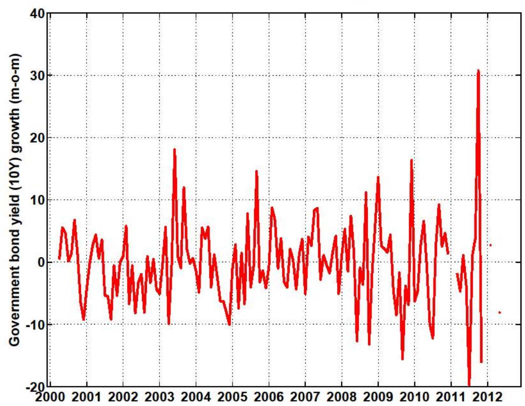
(c) 1Y PRIBOR



(d) PX-50 stock index



(e) Czech government bond yield (10Y)



(f) Consumer confidence indicator

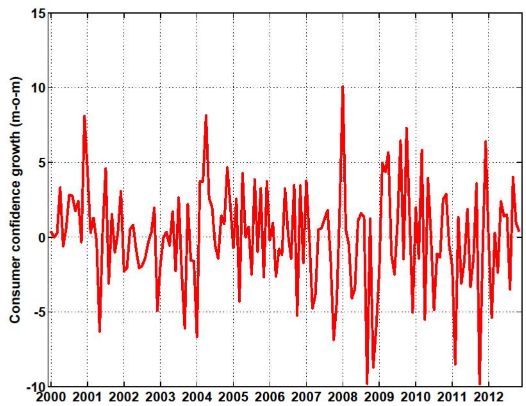
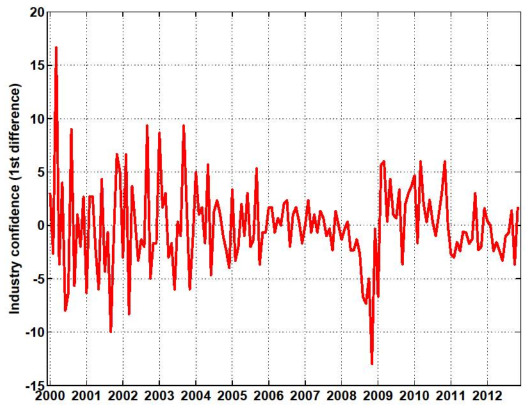
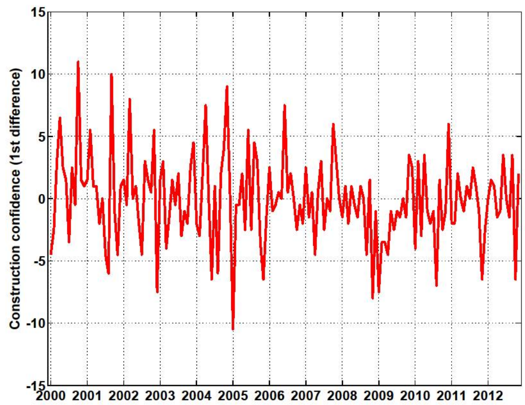


Figure 5.11: Data (continued)

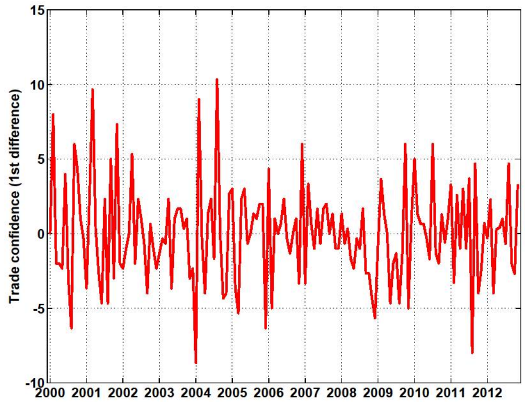
(a) Industry confidence



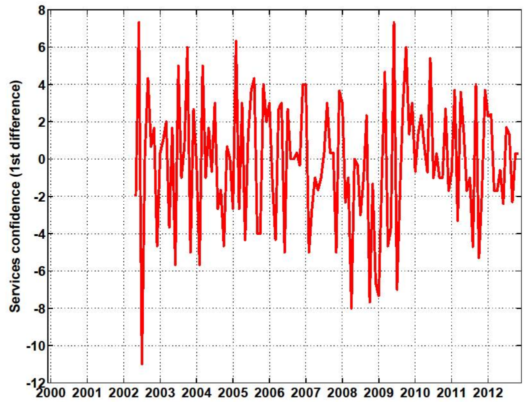
(b) Construction confidence



(c) Trade confidence



(d) Services confidence



(e) 3M EURIBOR

(f) 1Y EURIBOR

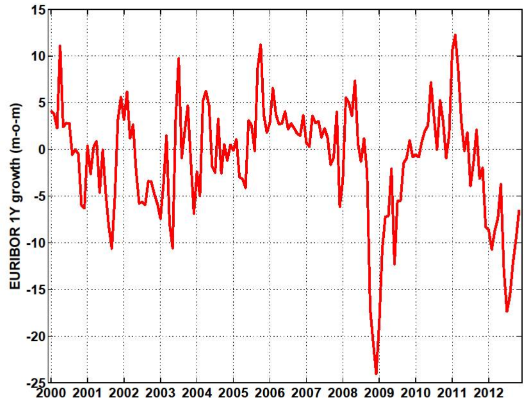
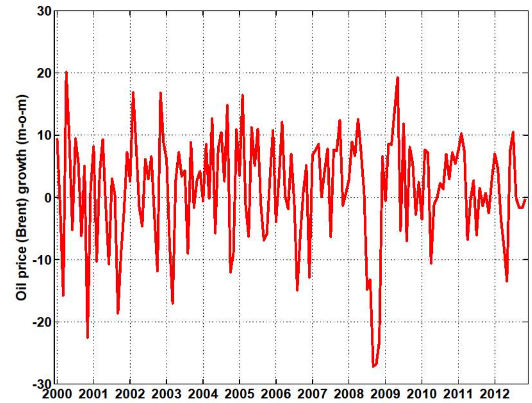
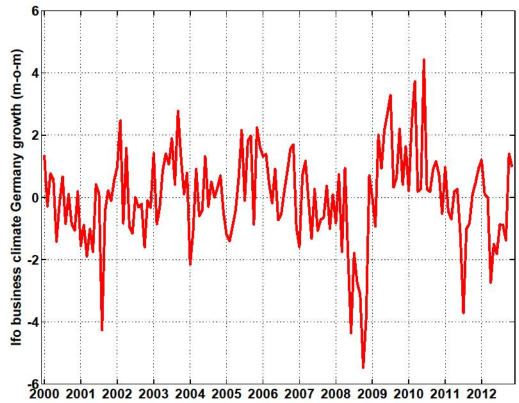


Figure 5.12: Data (continued)

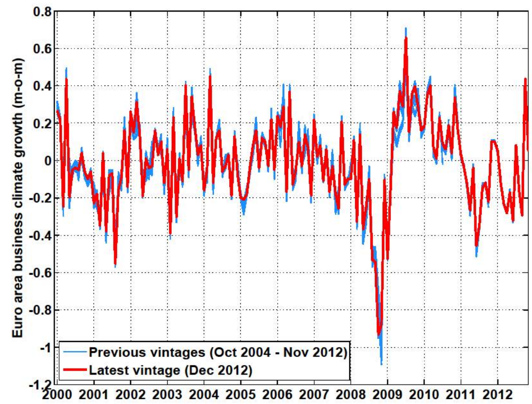
(a) Oil price (Brent)



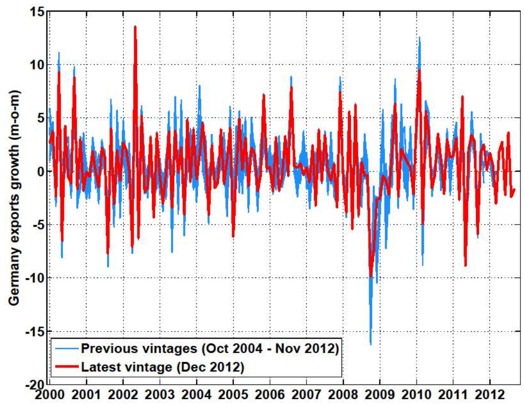
(b) Ifo business climate Germany



(c) Euro area business climate



(d) Germany exports



Chapter 6

Revisions to Czech National Accounts: Properties and Predictability

Abstract

Frequent revisions to the GDP and its components cause policymakers to face considerable uncertainty about the current state of the economy. In this paper, we provide stylized facts about the magnitude of revisions to the Czech national accounts. Using data over the 2003–2012 period, we find that the revisions are rather large. Revisions to real GDP growth are on average 1.4 for annualized quarterly growth rate and 0.7 percentage points for annual growth rate. Revisions to other variables are even larger: the average size of revisions range from 1 to 12 percentage points for annualized quarterly growth rates and from 0.5 to 4 percentage points for annual growth rates. We investigate whether the revisions could have been predicted using the information available at the time of announcement. We find evidence for in-sample predictability for most of the variables, suggesting that the first releases of these variables are not efficient predictors of the actual values. In a real-time out-of-sample exercise, however, we find that the revisions to real GDP, gross fixed capital formation and government consumption are not predictable. Only revisions to GDP deflator can be predicted with substantial gains relative to zero revisions forecasts.

The paper was published in the *Czech Journal of Economics and Finance (Finance a uver)* [2013, 63(3), pp. 244–261]. We would like to thank Michal Franta and the seminar participants at the Charles University in Prague for helpful comments and discussions. We acknowledge the support of Czech National Bank and financial support from the Grant Agency of Charles University (grant #888413) and the Grant Agency of Czech Republic (grant P402/12/1993). paper are only those of authors and do not necessarily reflect the position and views of the Czech National Bank or any other institution with which authors are associated.

6.1 Introduction

Crucial indicators about the state of the economy – GDP and its components – are measured imperfectly. They are available only after a significant lag and are often subject to revisions. The fact that important macro aggregates are imprecisely measured is of importance for policymakers who must make decisions that depend vitally on the current state of the economy. If data are often revised, a question arises as to how much weight should policymakers attribute to initial data releases. The pursuit of optimal policy might be jeopardized by an over-reaction to current data (Orphanides 2003; Kugler *et al.* 2005). Indeed, policymakers regularly discuss the expected revisions to the new data in their monetary policy deliberations.[1] Additionally, the revisions are often one of the main sources of Czech National Bank (CNB) forecast errors.[2]

The importance of using real-time data is already well recognized in the literature on forecasting and monetary policy analysis (Robertson & Tallman 1998; Croushore & Stark 2001; Croushore 2011). To that end several real-time databases are established (Croushore & Stark 2001; McKenzie 2006; Fernandez *et al.* 2011; Giannone *et al.* 2012). An increasing number of papers point out that many results obtained using revised data are sensitive to real-time data issues (Swanson & White 1997; Amato & Swanson 2001; Orphanides 2001; Orphanides & van Norden 2002; Christoffersen *et al.* 2002; Molodtsova *et al.* 2008).

The reasons why statistical offices perform revisions are discussed in McKenzie (2006). The Czech Statistical Office (CZSO) revises the data of the current year with each regular release of quarterly data. The CZSO states that more complete and updated information are the main reason for the revisions. In addition to that, twice a year revisions are made because of annual account compilation affecting the data as far as three years ago. Furthermore, revisions originating from seasonal adjustment are made each quarter to the current year data, and once a year to the

[1] See, for example, recent minutes of the Bank Board Meeting on 27 September 2012, where the expected revisions to investment data were discussed (http://www.cnb.cz/en/monetary policy/bank board minutes/2012/amom 120927.html).

[2] See, for example: http://www.cnb.cz/en/public/media service/comments/2012/12 hdp 2q.html.

whole series.[3] In addition to these regular revisions, the CZSO occasionally performs also benchmark revisions, which reflect changes in the methodology and affect the whole time series.

Mankiw *et al.* (1984) propose that the revision process might be characterized as either reflecting measurement error (revisions are then referred to as *noise*) or reflecting new information (revisions are referred to as *news*). If revisions are noise, the first release of a variable is an imperfect measure of the true variable. We can therefore make use of other information available at the time of the release to produce better forecast of the true value. The optimal forecast of the true value is then a weighted average of the first release value and the conditional mean of other observable data. For example, we can use the mean of the underlying variable itself (in such cases the optimal forecast of future revisions is related to the deviation of the value of the first release from the mean of the underlying variable). The larger is the variance of the measurement error, the smaller weight should be attributed to the first release observation.

If revisions contain news, they are not predictable using the information available at the time of first release. Therefore, it is optimal to put a weight of one to the value of first release and a weight of zero to other observable data. In other words, the optimal forecast of future revisions is zero. When revisions are news, the first releases are often referred to as rational or efficient forecasts of the true value of a variable.

Mankiw *et al.* (1984) find that the revisions to U.S. money aggregates can be characterized as noise. Mankiw & Shapiro (1986) find that revisions to U.S. nominal and real output can be characterized as news. Croushore & Stark (2001) find that short-term revisions to U.S. GDP contain news, while long-term revisions seem to reduce noise. Using a longer sample, Aruoba (2008) provides evidence showing that revisions to most U.S. macro variables are biased and cannot be characterized as reflecting news. Garratt & Vahey (2006) comes to similar conclusions for the U.K.

[3] See the description of methods used for the compilation of national accounts by the Czech Statistical Office available at http://www.czso.cz/eng/redakce.nsf/i/gross domestic product (gdp).

Faust *et al.* (2005) document that revisions to GDP are predictable in most G-7 countries. Recently, de Castro *et al.* (2011) shows that the revisions to releases of budget deficit data in EU-15 are biased downward and cannot be considered as news.

The above mentioned literature attempts to characterize the revisions either as news or noise. There also exists related a strand of literature that focuses on optimal forecasting and inference in the presence of revisions (Sargent 1989; Kapetanios & Yates 2010). Jacobs & van Norden (2011) try to connect these two strands of literature and model the news and noise simultaneously within a state-space framework.

There is virtually no evidence about the properties of Czech real-time data.[4] The main objective of this paper is to fill this void, to enhance our understanding about the size and the properties of revisions to Czech national accounts. We gather real-time vintages of Czech GDP and its components over 2003-2012 and provide evidence about their statistical properties. Moreover, in line with the above mentioned literature, we test whether the revisions to Czech national accounts can be viewed as noise or news. We therefore investigate the predictability of revisions both in-sample and in a real-time out-of-sample exercise. Note, that the analysis is not meant to criticize statistical agency: the CZSO certainly has limited resources and tries to minimize revisions subject to its operational constraints. The main objective of our analysis is to improve our understanding of the properties of revisions.

Our results suggest that the revisions to real GDP and its components are largely unbiased, with the exception of a positive bias in short-term revisions to annual growth rates of exports and imports. Revisions to GDP deflator, on the other hand, are biased downward for both quarterly and annual growth rates. The revisions are rather large: the mean absolute revision to annualized quarterly GDP growth is roughly 1.4 percentage points and roughly 0.7 for annual growth rate. Revisions for other variables are even larger. Judging by the size of revisions relative to the size of the original variables, the largest relative revisions seem to occur in consumption and

[4]Exceptions are the two boxes in the CNB Inflation Reports that, however, consider only the effects of benchmark revisions: in 2004 (http://www.cnb.cz/en/monetary policy/inflation reports/2004/2004 october/boxes annexes/zpinflace 04 october b3.html) and in 2011 (http://www.cnb.cz/en/monetary policy/inflation reports/2012/2012 I/boxes and annexes/zoi 2012 I box 2.html).

gross fixed capital formation. On the other hand, exports and imports have smallest noise to signal ratio from the components of GDP. Next, we find that revisions for GDP are predictable in-sample and thus the first releases cannot be characterized as news. In addition to in-sample evidence, we also investigate whether the revisions are predictable in a real-time out-of-sample exercise. We find evidence of out-of-sample predictability for the revisions of GDP deflator, and to a lesser extent to consumption and year-on-year growth rates of exports and imports, while for other variables zero revision forecasts seems to perform better in the real-time exercises. The results from the real-time exercise should be, however, viewed with caution since the out-of-sample period is very small and covers the recent crisis period.

The remainder of the paper is structured as follows. Section 6.2 describes the data. Section 6.3 provides stylized facts about the revisions while Section 6.4 examines their in-sample and out-of-sample predictability. Section 6.5 summarizes the implications of our key results.

6.2 The Data

We gather historical vintages of Czech data on nominal GDP, GDP deflator, real GDP and its expenditure components: real private consumption, real gross fixed capital formation (GFCF), real government consumption, real exports, and real imports. The source of our data is the *OECD Real-Time Data & Revisions Database.*[5] Note that the Czech Statistical Office provides a history of its announcements on its website starting from 2003. We prefer to work with the OECD database because it provides us with the whole time series of data for each component of GDP at each vintage, while the CZSO archive does not always provide time series for the components. Nevertheless, we double-checked the data for first releases obtained from the OECD Real-Time database with the archive of the Czech Statistical Office to ensure consistency.[6]

[5] Available at http://stats.oecd.org/mei/default.asp?rev=1.

[6] There was only one inconsistency: the 2004Q1 value was missing from the release of the same data in the OECD Real-Time database, perhaps because the CZSO did not release seasonally adjusted quarterly series. Therefore, for the annual growth value we used the value stated at the CZSO

Table 6.1 illustrates how the real-time data are typically structured in a so-called revision triangle. The subscript denotes the reference period which the observation captures, while the superscript denotes the period in which the release is made. Notation emphasizes the fact that national accounts data are only available with a 1 quarter lag.

Table 6.1: Revision triangle

Vintage				
1	$\cdots$	$t-v$	$\cdots$	$t+1$
y_0^1	$\cdots$	y_0^{t-v}	$\cdots$	y_0^{t+1}
	$\ddots$	$\vdots$		$\vdots$
		y_{t-v-1}^{t-v}	$\cdots$	y_{t-v-1}^{t+1}
			$\ddots$	$\vdots$
				y_t^{t+1}

The real-time vintages for seasonally adjusted real GDP and its components are available from September 2003 till March 2013. That means we have 39 observations available for the period 2003Q2-2012Q4. We do not consider preliminary (flash) estimates, since they are announced only starting from the fourth quarter of 2007.[7]

Figure 6.1 plots all of the available historical vintages of GDP in both quarter-on-quarter and year-on-year growth rates. The figure illustrates the uncertainty about the real GDP growth caused by revisions. Note that, on some occasions first release data point to acceleration of the growth relative to its previous (first release) value while the later revisions suggest the opposite.[8] This might be especially problematic for policymakers because they usually consider potential inflationary vs. anti-inflationary pressures. Such data uncertainty therefore hinders the decision

website, and for the quarter-on-quarter growth we used value from the next release. The data at other releases were exactly the same in both sources.

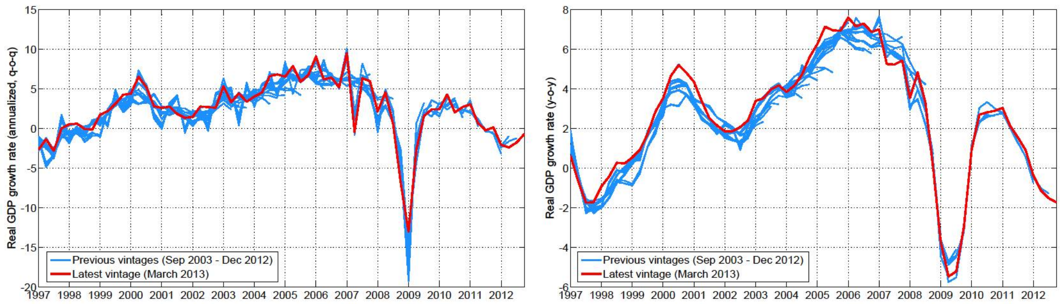
[7]The correlation of the preliminary estimate of the quarterly growth rate of Real GDP with the first release and the preliminary estimate with the final value (as of March 2012) is 0.95 and 0.87, respectively. Similarly, the correlation of the preliminary estimate of the annual growth rate of Real GDP with the first release and the preliminary estimate with the final value (as of March 2012) is 0.99 and 0.97, respectively. The correlations are based on 21 observations. We provide scatter plots in the Appendix.

[8]See also Figure 6.3 and 6.4 in the Appendix for better illustration. For example, there are 12 such occasions in case of short-term revisions to quarter-on-quarter growth, and 8 in case of long-term revisions to quarter-on-quarter growth. As for the year-on-year growth rates, there are 9 and 8 occasions for short-term and long-term revisions, respectively.

making about the optimal policy. For an illustration of data uncertainty of other components and time-series of short-term and long-term revisions, see supplementary figures in the Appendix.

Figure 6.1: Growth rates of various GDP vintages

(a) Annualized quarterly growth rates (q-o-q) (b) Annual growth rates (y-o-y)



To mitigate the effect of benchmark revisions we take the standard practice of analyzing the growth rates of the variables. We analyze both the annualized quarter-on-quarter growth rates and annual year-on-year growth rates. We decided to work with both transformation, since both of them might be of interest for macroeconomists and policymakers. Year-on-year growth rates are smoother, while quarter-on-quarter might be more noisy. On the other hand, quarter-on-quarter rates might be more efficient for forecasting, while year-on-year growth rates are autocorrelated by definition.

Since the data are continually revised, it is not clear which observations should be considered the "true" or final ones. The last published data may/should be closest to the "true" values since they reflect the latest information available to the statistical office and should reflect the latest, most up-to-date methodologies for computing the released data. On the other hand, revisions coming from changes in definitions and methodologies might not be of interest since they are not telling us about the efficiency of data releases under the current measurement system (Faust *et al.* 2005). Furthermore, policymakers might be more interested in revisions to recent data only, since these might have direct effect on their decision making.

To account for that, we compute two measures of revisions. First, we compute short-term revisions as the difference between the growth rate of a variable a year after the first release was published and the original first release growth rate ($r_t = y_t^{t+5} - y_t^{t+1}$). Second, we compute long-term revisions as the difference between the growth rate in the last vintage available and the original first release growth rate ($r_t = y_t^T - y_t^{t+1}$, where $T =$ March 2013). To make sure that each data series has at least some revisions we use only data through to 2011Q4, thus leaving 35 observations for the analysis. This might be considered low, but on the other hand, the practice and the methodology of statistical office have evolved over time, so having a longer sample could raise doubts about the presence of structural breaks.

6.3 Statistical Properties of Revisions

Aruoba (2008) puts forward three basic statistical properties that well-behaved revisions should satisfy. First, the mean of the revisions should be zero. In that case, the first releases of the statistical agency can be considered an unbiased estimates of the final values. Second, the variance of the revisions should be small relative to the variance of the underlying variable. Third, the first release of the statistical office should be an optimal forecast of the final value. That means that the revisions should not be predictable using information available at the time of the first release, i.e. revisions should be news. In this section, we focus on the first two of these properties, while the investigation of the predictability of revisions is deferred to the next section.

Table 6.2 provides information about the mean revisions, the maximum and minimum revision, the standard deviation of revisions, the mean absolute revisions, and the autocorrelation of revisions. Because absolute values may not be very informative about the size of the revisions relative to the size of original variables, we also report noise to signal ratios. The noise to signal ratio is defined as the standard deviation of revisions divided by the standard deviation of the final value of the variable.

Short-term revisions to quarter-on-quarter growth rates are generally unbiased: their mean revisions are very close to zero. The exception is the bias in GDP deflator revisions. Short-term quarter-on-quarter growth rates of GDP deflator seem to be systematically revised downward: on average by -0.57 percentage points and statistically significant at a 10 % significance level. We use t-statistics based on Newey-West heteroscedasticity and autocorrelation consistent standard errors with a lag parameter of four to test whether the mean revisions are zero. As for the long-term revisions to quarter-on-quarter growth rates, we find that they are largely unbiased.

The bottom panel of Table 6.2 presents the statistics about year-on-year growth rates. We find statistically significant bias in both short-term and long term revisions of GDP deflator. In addition, short-term revisions to exports and imports are biased upward and the bias is statistically significant. This suggests a tendency towards more pessimistic values in the first releases of exports and imports.

The short-term revisions in the variables with biased revisions cannot be considered as news. The biases in short-term revision suggest the potential predictability for these three variables. We investigate how this predictability may be exploited to improve forecasts of national accounts.

The magnitudes of the revisions are rather large. Minimum and maximum short-term revisions range from -4 to 2 percentage points for Real GDP for annualized quarter-on-quarter growth rates and from -2 to 1.4 for year-on-year growth rates. The range is even larger for other variables and long-term revisions. This is confirmed by the standard deviation of revision and the mean absolute revision. The smallest mean absolute revisions for short-term revisions to quarter-on-quarter growth rates are for GDP Deflator (1.2 percentage point in annualized growth rates) while largest mean revisions are made to Exports (6.8 percentage points in annualized growth rates). As for year-on-year growth rates, the mean absolute short-term revisions range from 0.6 (GDP Deflator) to 2.6 (Gross Fixed Capital Formation). The mean absolute long term revisions are smallest for real GDP and largest for Gross Fixed Capital Formation.

The fact that the revisions are very large is also corroborated by the noise to signal ratios which range from 0.4 to 1.1 for quarter-on-quarter growth rates and from 0.2 to 0.8 for year-on-year growth rates. The smallest relative revisions to quarter-on-quarter growth rates are made to real GDP (noise to signal ratio approximately

Table 6.2: Summary Statistics of Revisions

	NGDP	RGDP	GDPD	C	I	G	E	I
	Annualized quarterly growth rates (q-o-q)							
Short-Term Revisions								
Mean	-0.62	-0.04	-0.57	-0.30	0.45	0.53	1.98	2.02
p-value	0.10	0.89	0.02	0.59	0.71	0.30	0.07	0.03
Max	2.85	2.02	2.28	5.45	24.16	8.26	21.64	26.25
Min	-9.10	-4.04	-5.00	-7.53	-27.69	-9.51	-12.23	-10.77
Std. Deviation	2.38	1.53	1.66	2.63	9.23	4.13	8.56	7.67
Mean Absolute Revision	1.83	1.30	1.21	1.94	6.27	3.36	6.88	5.46
Noise to Signal	0.53	0.37	0.63	0.99	0.70	0.73	0.49	0.41
AR(1)	-0.18	0.07	-0.29	0.15	-0.17	-0.15	-0.30	-0.17
Long-Term Revisions								
Mean	-0.50	0.15	-0.63	0.10	1.05	0.17	-0.46	-0.30
p-value	0.26	0.75	0.03	0.86	0.63	0.80	0.69	0.81
Max	8.20	3.53	5.40	10.65	58.75	13.65	27.58	21.28
Min	-9.90	-5.99	-7.20	-8.35	-27.73	-16.64	-38.27	-33.71
Std. Deviation	3.64	2.02	2.97	3.96	17.36	6.62	11.22	12.15
Mean Absolute Revision	2.64	1.52	2.24	3.09	11.92	5.11	7.71	10.02
Noise to Signal	0.67	0.46	0.71	1.10	1.10	0.82	0.67	0.67
AR(1)	-0.47	0.05	-0.55	-0.01	-0.16	-0.27	-0.54	-0.66
	Annual growth rates (y-o-y)							
Short-Term Revisions								
Mean	-0.15	0.17	-0.33	-0.03	-0.94	0.52	1.13	0.87
p-value	0.48	0.41	0.03	0.91	0.19	0.14	0.01	0.01
Max	1.49	1.36	1.12	1.42	6.36	5.25	6.23	4.73
Min	-2.49	-2.05	-1.69	-2.17	-5.97	-2.08	-2.47	-2.27
Std. Deviation	0.94	0.76	0.63	0.90	3.18	1.65	2.23	1.74
Mean Absolute Revision	0.74	0.61	0.56	0.69	2.64	1.28	1.88	1.44
Noise to Signal	0.24	0.24	0.34	0.41	0.59	0.70	0.22	0.18
AR(1)	0.42	0.56	0.40	0.54	0.37	0.34	0.33	0.23
Long-Term Revisions								
Mean	-0.29	0.31	-0.58	0.05	0.17	0.50	-0.96	-1.32
p-value	0.28	0.27	0.00	0.87	0.90	0.25	0.14	0.10
Max	2.74	2.13	1.32	2.13	17.75	5.53	6.36	3.49
Min	-3.04	-1.80	-2.80	-2.51	-11.44	-1.65	-6.78	-9.43
Std. Deviation	1.28	0.93	0.88	1.16	5.65	1.91	2.81	3.21
Mean Absolute Revision	1.03	0.76	0.79	0.96	4.07	1.34	2.37	2.55
Noise to Signal	0.32	0.27	0.44	0.64	0.80	0.63	0.30	0.35
AR(1)	0.34	0.66	0.30	0.59	0.59	0.35	0.44	0.48

Notes: The summary statistics are based on 2003Q2-2011Q4 revisions. NGDP denotes nominal GDP, RGDP denotes real GDP, GDPD denotes GDP deflator, C denotes real consumption, I denotes real gross fixed capital formation, G denotes real government consumption, E denotes real exports, and M denotes real imports. The short-term revision is the value from a year after the first release minus first-release value, the long-term revision is the final value minus the first-release value. Noise to Signal is defined as the standard deviation of revisions divided by the standard deviation of the final value of the variable. *p*-value is a statistics from a test that the mean revision is zero using autocorrelation and heteroscedasticity consistent standard errors. *AR*(1) denotes autocorrelation coefficient of the first order.

0.4). The largest relative revisions seems to occur to Consumption and Gross Fixed Capital Formation. On the other hand, exports and imports seem to have smallest relative revisions among the components of GDP. The revisions to the components are generally larger than to GDP as a whole. Research on the revisions of the GDP expenditure components is generally very scarce. But our results are in line with U.S. evidence provided by Aruoba (2008), who also finds that the revisions to components are larger than to aggregate GDP.

The revisions to quarter-on-quarter are generally not very persistent, as indicated by low first-order autocorrelation coefficients. Coefficients are mostly below 0.5. Revisions to quarter-on-quarter Real GDP growth rate seem to be not autocorrelated. Note, that the low autocorrelation suggests that autoregressive models will be relatively uninformative for predicting the revisions. As for the year-on-year growth rates, as expected, the order of autocorrelation is generally higher: most of the variables have autocorrelation coefficient higher than 0.4.

As for the international comparison of the magnitudes of revisions, McKenzie (2006) reports mean absolute revisions to quarter-on-quarter growth of GDP for 18 OECD countries over 1995-2004 period. The short-term mean absolute revisions range from 0.1 for Spain to almost 0.7 for Norway. The average mean absolute revision in these 18 OECD countries is 0.3, which is 1.2 percentage points at an annualized rate. Bearing in mind that the sample period for which the revisions are computed differs relative to our study, it seems that the short-term revisions to Czech GDP are on average similar in magnitude to those in OECD countries.

6.4 News and Noise in Czech National Accounts Revisions

We can decompose the first release data (y_t^{t+1}) as being equal to the final data (y_t^f) plus an error term (ε_t):

$$y_t^{t+1} = y_t^f + \varepsilon_t. \tag{6.1}$$

The literature views the revision process in two ways: revisions can be viewed as capturing noise or they can reflect news. Under the noise view, first release data contain a measurement error that is uncorrelated with the true values: y_t^f is orthogonal to ε_t. Under the news view, revisions reflect new information that is becoming available to the statistical office over time: y_t^{t+1} is orthogonal to ε_t.

We will run so called forecast efficiency regressions (Mincer & Zarnowitz 1969; Mankiw *et al.* 1984; Mankiw & Shapiro 1986; Faust *et al.* 2005; Aruoba 2008) to determine whether revisions to Czech national accounts can be viewed as noise or news. We test for noise by running the following regression:

$$r_t = \alpha_1 + \beta_1 y_t^f + \varepsilon_{1t}. \tag{6.2}$$

The null hypothesis is that the revisions reduce noise, that means that they are not related to the true values of the variable ($\alpha_1 = \beta_1 = 0$). If data revisions are noise, it would be optimal to discount the first release observation. More precisely, the optimal forecast for the variable would be a weighted average of the preliminary announcement and the conditional mean of the underlying variable.

To test the news hypothesis, we run the following regression:

$$r_t = \alpha_2 + \beta_2 y_t^{t+1} + \varepsilon_{2t} \tag{6.3}$$

The null hypothesis in this case is that revisions reflect the new information and thus are not predictable by the information available at the time of release ($\alpha_2 = \beta_2 = 0$). When the revisions are news, the first release observation is an efficient forecast of the variable, and thus it is optimal to assign it a full weight.

As noted by Aruoba (2008) these hypotheses are mutually exclusive but not collectively exhaustive - one can reject both hypotheses (e.g. if the constant is significant in both regressions). In small samples, one can reject or fail to reject both hypotheses because of sampling errors. Note also that in reality, it is likely that the revisions contain both noise and news components.

6.4.1 Testing for Noise

The upper panel of Table 6.3 presents the results of noise regressions for the variables in quarter-on-quarter growth rates. It shows that we are able to reject the noise hypothesis for almost all variables in quarter-on-quarter growth rates. Exceptions are short-term revisions of Nominal GDP and GDP Deflator.

The bottom panel of Table 6.3 presents the results of noise regressions for the variables in year-on-year growth rates. The results are mixed: for short-term revisions to most variables the noise hypothesis cannot be rejected, while for long-term revisions we reject the noise hypothesis for most variables except for Consumption, Exports and Imports.

6.4.2 Testing for News: Baseline Regressions

To investigate whether the revisions behave as news we run regression (3) and test whether $\alpha_2 = \beta_2 = 0$. The results are reported in Table 6.4.

The upper panel of Table 6.4 presents the results for variables in quarter-on-quarter growth rates. The F-statistics suggest we can only reject the hypothesis of news for GDP deflator and consumption. Short-term revisions in other variables seem to be unpredictable in these baseline naive regressions. The degree of predictability for GDP deflator and consumption is relatively high, with R^2=0.25 and R^2=0.35, respectively. When looking at long-term revisions we reject (at 10% significance level) the news hypothesis for consumption, gross fixed capital formation and exports and imports. The bottom panel of Table 6.4 shows the results for variables in year-on-year growth rates. We are able to reject the news hypothesis for short-term revisions in government consumption and exports and long-term revisions in consumption.

6.4.3 Testing For News: Augmented Regressions

If revisions are to be deemed news, revisions should not be predictable using any data available at the time of the announcement of first release. We therefore test whether some additional variables could be used to enhance the predictability of revisions. Mankiw & Shapiro (1986) use equity prices and short-term interest rates as a business cycle indicators. In addition to that, Faust *et al.* (2005) use also oil prices. Therefore, in our exercise, we follow the previous literature and add four more explanatory variables: the lagged value of a revision to capture potential persistence,

Table 6.3: Testing for Noise

	NGDP	RGDP	GDPD	C	I	G	E	M
	Annualized quarterly growth rates (q-o-q)							
Short-Term Revisions								
Final Release	0.001	0.089**	-0.082	0.165*	0.356***	0.170*	0.182**	0.079
	(0.054)	(0.036)	(0.052)	(0.090)	(0.099)	(0.091)	(0.073)	(0.055)
Constant	-0.626*	-0.326	-0.463*	-0.702	-0.560	0.507	0.580	1.534
	(0.319)	(0.232)	(0.242)	(0.475)	(0.712)	(0.479)	(1.050)	(0.933)
F	0.00	5.97	2.49	3.35	12.96	3.50	6.23	2.03
p-value	0.98	0.02	0.12	0.08	0.00	0.07	0.02	0.16
R^2	0.00	0.07	0.04	0.05	0.37	0.11	0.13	0.03
Long-Term Revisions								
Final Release	0.441***	0.243**	0.488***	0.730***	0.964***	0.642***	0.307	0.229*
	(0.070)	(0.095)	(0.091)	(0.096)	(0.154)	(0.100)	(0.187)	(0.131)
Constant	-0.621***	-0.158	-0.316***	-0.424**	-0.424	0.0203	-0.706	-0.428
	(0.170)	(0.110)	(0.112)	(0.181)	(0.430)	(0.152)	(0.586)	(0.495)
F	39.60	6.49	29.12	57.88	39.26	41.15	2.70	3.05
p-value	0.00	0.02	0.00	0.00	0.00	0.00	0.11	0.09
R^2	0.43	0.28	0.48	0.44	0.77	0.62	0.21	0.12
	Annual growth rates (y-o-y)							
Short-Term Revisions								
Final Release	0.048	0.031	0.017	0.072	0.114**	-0.008	0.093**	0.020
	(0.036)	(0.024)	(0.070)	(0.125)	(0.045)	(0.133)	(0.041)	(0.041)
Constant	-0.372*	0.069	-0.345*	-0.221	-1.240*	0.527	0.334	0.734*
	(0.199)	(0.162)	(0.184)	(0.410)	(0.715)	(0.328)	(0.480)	(0.383)
F	1.74	1.57	0.06	0.34	6.35	0.00	4.99	0.23
p-value	0.20	0.22	0.81	0.56	0.02	0.95	0.03	0.64
R^2	0.04	0.02	0.00	0.02	0.07	0.00	0.15	0.01
Long-Term Revisions								
Final Release	0.111**	0.119***	0.168***	-0.055	0.540***	0.308*	0.021	0.006
	(0.052)	(0.038)	(0.055)	(0.163)	(0.143)	(0.179)	(0.056)	(0.052)
Constant	-0.804***	-0.097	-0.781***	0.195	-1.257	0.305	-1.145**	-1.360**
	(0.193)	(0.167)	(0.219)	(0.544)	(1.258)	(0.371)	(0.561)	(0.613)
F	4.46	9.65	9.22	0.11	14.33	2.97	0.14	0.01
p-value	0.04	0.00	0.00	0.74	0.00	0.09	0.71	0.90
R^2	0.12	0.20	0.15	0.01	0.46	0.24	0.01	0.00

Notes: NGDP denotes nominal GDP, RGDP denotes real GDP, GDPD denotes GDP deflator, C denotes real consumption, I denotes real gross fixed capital formation, G denotes real government consumption, E denotes real exports, and M denotes real imports. The short-term revision is the value from a year after the first release minus first-release value, the long-term revision is the final value minus the first-release value. Autocorrelation and heteroscedasticity consistent standard errors (Newey & West 1987) in parenthesis.

Table 6.4: Testing for News: Baseline Regressions

	NGDP	RGDP	GDPD	C	I	G	E	M
	Annualized quarterly growth rates (q-o-q)							
Short-Term Revisions								
First Release	-0.065	0.013	-0.274***	-0.496***	0.137	-0.205	0.001	-0.045
	(0.097)	(0.054)	(0.065)	(0.167)	(0.314)	(0.183)	(0.090)	(0.062)
Constant	-0.295	-0.081	-0.040	0.874**	0.207	0.523	1.968	2.312**
	(0.539)	(0.359)	(0.211)	(0.346)	(1.353)	(0.494)	(1.495)	(0.938)
F	0.46	0.05	18.03	8.86	0.19	1.26	0.00	0.54
p-value	0.50	0.82	0.00	0.01	0.66	0.27	0.99	0.47
R^2	0.01	0.00	0.25	0.35	0.02	0.06	0.00	0.01
Long-Term Revisions								
First Release	-0.013	0.044	-0.021	-0.644***	-0.875*	-0.061	-0.169*	-0.224**
	(0.095)	(0.045)	(0.113)	(0.165)	(0.507)	(0.185)	(0.098)	(0.108)
Constant	-0.441	0.012	-0.589*	1.624**	2.625	0.165	0.918	1.144
	(0.605)	(0.413)	(0.319)	(0.640)	(2.986)	(0.652)	(1.608)	(1.636)
F	0.02	0.96	0.03	15.21	2.99	0.11	2.96	4.29
p-value	0.90	0.33	0.86	0.00	0.09	0.74	0.09	0.05
R^2	0.00	0.01	0.00	0.26	0.18	0.00	0.05	0.11
	Annual growth rates (y-o-y)							
Short-Term Revisions								
First Release	0.002	-0.016	-0.062	-0.092	-0.170	-0.305***	0.065*	-0.005
	(0.036)	(0.035)	(0.091)	(0.088)	(0.105)	(0.077)	(0.032)	(0.027)
Constant	-0.160	0.223	-0.215	0.218	-0.515	0.564*	0.514	0.906***
	(0.191)	(0.172)	(0.231)	(0.292)	(0.870)	(0.294)	(0.376)	(0.292)
F	0.00	0.21	0.47	1.07	2.63	15.73	4.11	0.03
p-value	0.95	0.65	0.50	0.31	0.11	0.00	0.05	0.86
R^2	0.00	0.00	0.03	0.05	0.08	0.00	0.08	0.00
Long-Term Revisions								
First Release	0.011	0.058	-0.030	-0.306***	-0.172	-0.109	-0.067	-0.106
	(0.059)	(0.055)	(0.074)	(0.068)	(0.260)	(0.091)	(0.045)	(0.077)
Constant	-0.344	0.130	-0.529***	0.859**	0.599	0.518	-0.319	-0.460
	(0.264)	(0.182)	(0.174)	(0.352)	(1.591)	(0.420)	(0.523)	(0.643)
F	0.04	1.14	0.16	20.29	0.44	1.42	2.24	1.87
p-value	0.85	0.29	0.69	0.00	0.51	0.24	0.14	0.18
R^2	0.00	0.04	0.00	0.35	0.03	0.02	0.05	0.10

Notes: NGDP denotes nominal GDP, RGDP denotes real GDP, GDPD denotes GDP deflator, C denotes real consumption, I denotes real gross fixed capital formation, G denotes real government consumption, E denotes real exports, and M denotes real imports. The short-term revision is the value from a year after the first release minus the first-release value, the long-term revision is the final value minus the first-release value. Autocorrelation and heteroscedasticity consistent standard errors (Newey & West 1987) in parentheses.

growth rate of oil prices (EUCRBRDT index), growth rate of stock prices (PX index) and short-term interest rate (PRIBOR 3M).

The results from the augmented news regressions for quarter-on-quarter growth rates are reported in Table 6.5. The results suggest that there is evidence of the in-sample predictability of short-term revisions in most variables. F-tests reject hypothesis of forecast efficiency for GDP deflator, consumption, and government consumption, exports, and imports. Business cycle seem to be important especially for short-term revisions to GDP deflator (the oil prices are statistically significant), and for long-term revisions to consumption (oil and stock prices are statistically significant). For real GDP and gross capital fixed formation we are unable to reject the news hypothesis, suggesting that the revisions to quarter-on-quarter real GDP and gross capital formation growth rates are capturing news information thus they are unpredictable.

In Table 6.6 we present the results of the augmented news regressions for year-on-year growth rates. The results are similar to the results for quarter-on-quarter growth rates. In addition, we reject the news hypothesis for revisions to year-on-year growth rates of real GDP and gross capital formation. Especially, short-term revisions to real GDP now seem to depend heavily on business cycle: stock prices, oil prices and interest rates are statistically significant.

We also test the possible effects of quarterly dummies, which might be expected as a consequence of regular revisions due to the compilation of national accounts for the first and third quarter release and as a consequence of revisions to seasonal adjustment methodologies for the first quarter release. Overall, the quarterly dummies were not jointly significant at a 10% level for most of the variables.

Table 6.5: Testing for News: Augmented Regressions

	NGDP	RGDP	GDPD	C	I	G	E	M
	Annualized quarterly growth rates (q-o-q)							
Short-Term Revisions								
First Release	-0.018	0.010	-0.353***	-0.487**	0.143	-0.213	-0.030	-0.028
	(0.110)	(0.104)	(0.071)	(0.189)	(0.337)	(0.252)	(0.109)	(0.058)
Revision (t-1)	-0.342*	-0.050	-0.267**	0.140	-0.178	-0.067	-0.334*	-0.163**
	(0.185)	(0.271)	(0.112)	(0.107)	(0.163)	(0.264)	(0.186)	(0.074)
Oil Price	-0.080**	-0.023	-0.066***	0.009	0.035	-0.003	0.044	0.000
	(0.038)	(0.028)	(0.021)	(0.037)	(0.130)	(0.054)	(0.113)	(0.117)
Stock price	0.083	0.029	0.031	-0.001	-0.038	0.130*	0.085	-0.124
	(0.055)	(0.043)	(0.026)	(0.042)	(0.136)	(0.075)	(0.147)	(0.111)
Interest Rate	-0.436	-0.341	0.171	0.052	-1.265	0.782	-1.374	-1.676
	(0.398)	(0.264)	(0.226)	(0.431)	(1.319)	(0.704)	(1.603)	(1.322)
Constant	0.465	0.757	-0.179	0.762	2.945	-1.554	5.877	6.894*
	(1.002)	(0.649)	(0.610)	(0.989)	(3.043)	(1.885)	(4.613)	(3.734)
F	1.08	0.59	0.49	3.29	0.86	7.15	2.48	2.66
p-value	0.39	0.71	0.00	0.02	0.52	0.00	0.06	0.04
R^2	0.19	0.07	0.49	0.35	0.06	0.20	0.15	0.09
Long-Term Revisions								
First Release	0.087	0.025	-0.112	-0.506**	-0.872	0.121	-0.226***	-0.183***
	(0.121)	(0.092)	(0.125)	(0.222)	(0.581)	(0.212)	(0.053)	(0.058)
Revision (t-1)	-0.583***	-0.070	-0.595***	-0.030	-0.127	-0.326**	-0.564***	-0.660***
	(0.135)	(0.329)	(0.135)	(0.086)	(0.136)	(0.144)	(0.140)	(0.122)
Oil Price	-0.033	-0.002	-0.033	0.142***	-0.037	0.156	0.185	0.191
	(0.051)	(0.027)	(0.035)	(0.043)	(0.388)	(0.109)	(0.150)	(0.186)
Stock price	0.101	0.026	0.048	-0.123**	0.066	-0.125	-0.027	-0.173
	(0.084)	(0.038)	(0.039)	(0.057)	(0.335)	(0.138)	(0.214)	(0.184)
Interest Rate	-0.334	-0.519	0.459	-0.195	-1.357	-0.414	-2.660	-3.240*
	(0.580)	(0.360)	(0.371)	(0.690)	(3.287)	(0.995)	(1.607)	(1.639)
Constant	-0.597	1.202*	-1.832	1.202	5.953	0.628	6.461	7.281
	(1.074)	(0.682)	(1.139)	(1.495)	(5.330)	(2.383)	(4.509)	(4.407)
F	5.04	1.48	11.28	7.50	2.12	4.61	16.63	11.87
p-value	0.00	0.23	0.00	0.00	0.09	0.00	0.00	0.00
R^2	0.32	0.09	0.34	0.41	0.20	0.14	0.41	0.55

Notes: NGDP denotes nominal GDP, RGDP denotes real GDP, GDPD denotes GDP deflator, C denotes real consumption, I denotes real gross fixed capital formation, G denotes real government consumption, E denotes real exports, and M denotes real imports. The short-term revision is the value from a year after the first release minus the first-release value, the long-term revision is the final value minus the first-release value. Oil and stock prices are in quarter-on-quarter growth rates. Autocorrelation and heteroscedasticity consistent standard errors (Newey & West 1987) in parentheses.

Table 6.6: Testing for News: Augmented Regressions

	NGDP	RGDP	GDPD	C	I	G	E	M
				Annual growth rates (y-o-y)				
Short-Term Revisions								
First Release	-0.033	-0.015	-0.231**	-0.066	-0.179	-0.271***	-0.027	-0.014
	(0.058)	(0.057)	(0.096)	(0.097)	(0.110)	(0.068)	(0.061)	(0.040)
Revision (t-1)	0.137	0.039	0.159	0.469***	0.270	0.153	0.310**	0.212
	(0.236)	(0.237)	(0.205)	(0.136)	(0.164)	(0.103)	(0.127)	(0.145)
Oil Price	-0.012*	-0.011**	-0.009	0.011**	-0.015	-0.008	0.004	0.005
	(0.007)	(0.004)	(0.005)	(0.004)	(0.013)	(0.007)	(0.013)	(0.009)
Stock price	0.016*	0.014**	0.006	-0.001	0.026	0.019**	0.023	-0.001
	(0.009)	(0.006)	(0.006)	(0.005)	(0.018)	(0.007)	(0.017)	(0.013)
Interest Rate	0.024	-0.253*	0.483**	-0.006	-0.510	0.340*	-0.205	-0.261
	(0.224)	(0.146)	(0.191)	(0.174)	(0.487)	(0.177)	(0.261)	(0.286)
Constant	0.022	0.817**	-0.862**	-0.044	0.682	-0.459	1.089	1.318
	(0.417)	(0.336)	(0.372)	(0.357)	(1.106)	(0.485)	(0.814)	(0.827)
F	4.45	21.72	6.53	8.97	13.39	8.24	2.86	0.6
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.70
R^2	0.32	0.52	0.36	0.47	0.34	0.40	0.22	0.07
Long-Term Revisions								
First Release	-0.028	-0.038	-0.000	-0.247***	-0.388**	-0.089	-0.171***	-0.127**
	(0.068)	(0.055)	(0.107)	(0.081)	(0.173)	(0.090)	(0.030)	(0.051)
Revision (t-1)	0.291**	0.591***	0.272*	0.422***	0.562***	0.279	0.276	0.340**
	(0.135)	(0.185)	(0.138)	(0.135)	(0.067)	(0.195)	(0.180)	(0.137)
Oil Price	0.005	0.004	0.005	0.009	0.010	-0.006	0.027**	0.030*
	(0.005)	(0.004)	(0.005)	(0.006)	(0.030)	(0.008)	(0.010)	(0.017)
Stock price	0.007	0.004	-0.003	0.004	0.041	0.014*	0.019	-0.006
	(0.009)	(0.004)	(0.005)	(0.006)	(0.027)	(0.007)	(0.020)	(0.024)
Interest Rate	-0.017	-0.088	0.108	0.147	0.659	0.232	-0.910*	-0.746
	(0.271)	(0.137)	(0.158)	(0.229)	(0.738)	(0.188)	(0.445)	(0.578)
Constant	-0.253	0.263	-0.724**	0.068	-1.340	-0.402	2.119**	1.256
	(0.504)	(0.291)	(0.269)	(0.450)	(1.366)	(0.378)	(0.882)	(1.232)
F	1.63	11.27	2.13	43.21	24.24	2.35	17.54	3.02
p-value	0.19	0.00	0.09	0.00	0.00	0.07	0.00	0.03
R^2	0.17	0.51	0.14	0.63	0.43	0.22	0.39	0.36

Notes: NGDP denotes nominal GDP, RGDP denotes real GDP, GDPD denotes GDP deflator, C denotes real consumption, I denotes real gross fixed capital formation, G denotes real government consumption, E denotes real exports, and M denotes real imports. The short-term revision is the value from a year after the first release minus first-release value, the long-term revision is the final value minus the first-release value. Oil and stock prices are in year-on-year growth rates. Autocorrelation and heteroscedasticity consistent standard errors (Newey & West 1987) in parenthesis.

6.4.4 Real-Time Forecasting Exercise

The results from the previous section suggest that there is some evidence of predictability in many variables - especially GDP deflator and consumption. However, the results are for in-sample predictability. It is well known that in-sample fit by no means guarantees good out-of-sample performance, especially in the presence of uncertain parameter instabilities or structural breaks in the data. Therefore, we perform the following real-time out-of-sample forecasting exercise. We focus on forecasting short-term revisions, since long-term revisions are not known until the last data observation. We design our real-time exercise as follows. After the release of GDP and national accounts for 2008Q1, we want to forecast its future short-term revision i.e. the revision as of 2009Q1. We have at our disposal a history of short-term revisions until 2007Q1.

We then use several models (summarized in Table 6.7). The first model is a naive model that assumes no revisions are made. The second model explores the potential presence of bias in revisions: the forecast of the second model is the mean bias computed from the sample available at the time the forecast is made (for the first forecast 2003Q2-2007Q1). The third model uses the constant and the announced value of the first release of a variable. The fourth model uses the lagged value of past revisions to exploit potential persistence. The fifth model augments the third model by adding the first principal component of stock prices, oil prices, interest rate, and the first-release value of the respective variable to capture the common factor that should represent the state of the business cycle.[9]

Overall, we perform 16 recursive out-of-sample forecasts for revisions to 2008Q1-2011Q4 announcements. The results are presented in Table 6.7. We present the root mean square errors of our forecast revisions and compare them with the benchmark model that assumes zero revisions. We also present the results of the test of Clark

[9] As for the real-time out-of-sample performance of augmented regressions, we find that forecast performance deteriorates greatly relative to baseline forecasts. Since we are working with small samples, it might be that the additional regressors are very imprecisely estimated. Therefore, we perform a slightly different exercise that uses principal component analysis. Variables used for common factor are standardized in mean and variance.

& West (2007).[10] Relative root mean square errors that are lower than one and statistically significant are presented in bold.

The results show evidence of real-time out-of-sample predictability of GDP deflator: most models are able to beat zero-revisions benchmark. Furthermore, for quarter-on-quarter growth rates we find some evidence of predictability for Consumption (Model 3) and Imports (Model 2). In case of year-on-year growth rates, exports and imports seem to be predictable as well (Models 2 and 5). Finally, note that the sample over which the forecasting exercise is performed is rather small (16 observations) and covers the crisis period (2008Q1-2011Q4). Therefore, the results should be interpreted with caution.

[10]This test allows us to compare nested forecasts by accounting for the noise term that is caused by the estimation of additional parameters.

Table 6.7: Real-Time Forecasts of Short-Term Revisions

Summary of forecasting models	
Model 1 (benchmark, zero revisions):	$r_t^f = \varepsilon_t$
Model 2 (constant only):	$r_t^f = \alpha + \varepsilon_t$
Model 3 (constant and first release):	$r_t^f = \alpha + \beta y_t^{t+1} + \varepsilon_t$
Model 4 (lagged):	$r_t^f = \alpha + \delta r_{t-4} + \varepsilon_t$
Model 5 (factor):	$r_t^f = \alpha + \gamma f_t + \varepsilon_t$

	NGDP	RGDP	GDPD	C	I	G	E	M
	Annualized quarterly growth rates (q-o-q)							
Model 1:								
$RMSE_1$	3.06	1.75	1.83	2.67	12.43	3.93	9.09	9.58
Model 2:								
$RMSE_2/RMSE_1$	**0.98**	1.12	**0.96**	1.02	1.01	1.02	1.00	**0.97**
Clark-West test p-value	0.07	0.96	0.06	0.47	0.71	0.53	0.19	0.04
$RMSE_3/RMSE_1$	1.22	1.85	**0.87**	**0.85**	1.17	1.20	1.05	0.99
Clark-West test p-value	0.80	0.96	0.00	0.00	0.61	0.69	0.67	0.18
$RMSE_4/RMSE_1$	1.09	1.13	1.00	1.04	1.11	1.14	1.00	1.03
Clark-West test p-value	0.74	0.93	0.27	0.63	0.99	0.83	0.24	0.50
$RMSE_5/RMSE_1$	0.98	1.12	**0.92**	1.06	1.06	1.39	1.02	1.00
Clark-West test p-value	0.16	0.87	0.05	0.49	0.50	0.72	0.23	0.11
	Annual growth rates (y-o-y)							
Model 1:								
RMSE1	1.00	0.74	0.61	1.02	3.62	1.06	2.70	2.39
$RMSE_2/RMSE_1$	1.07	1.32	**0.84**	1.04	0.98	1.55	**0.92**	**0.89**
Clark-West test p-value	0.81	0.85	0.01	0.95	0.16	0.92	0.04	0.00
$RMSE_3/RMSE_1$	1.33	2.35	1.30	1.15	1.65	1.27	1.07	0.95
Clark-West test p-value	0.90	0.72	0.63	0.88	0.90	0.27	0.11	0.13
$RMSE_4/RMSE_1$	1.23	1.26	**0.91**	1.10	0.99	1.34	1.05	1.03
Clark-West test p-value	0.90	0.82	0.00	0.90	0.18	0.99	0.21	0.36
$RMSE_5/RMSE_1$	0.94	1.19	**0.95**	**0.97**	1.63	1.37	**0.82**	**0.92**
Clark-West test p-value	0.11	0.38	0.01	0.09	0.93	0.41	0.01	0.00

Notes: NGDP denotes nominal GDP, RGDP denotes real GDP, GDPD denotes GDP deflator, C denotes real consumption, I denotes real gross fixed capital formation, G denotes real government consumption, E denotes real exports, and M denotes real imports. $RMSE_i$ denotes the root mean square error of model i. Forecasting performance computed over 2008Q1-2011Q2 period. Factor f_t in *Model 5* is computed as the first principal component of oil price growth, stock price growth, interest rate, and first-release value of the respective variable.

6.5 Concluding Remarks

In this paper, we investigate the properties and predictability of revisions to Czech national accounts over 2002–2012 period. The results show that the revisions are sizeable, which implies that for the results from analyses of Czech macroeconomic policy or forecasting exercises to be relevant one should use real-time data (as also stressed by Croushore 2011). Revisions are large enough that they appear to be of economic significance, for policy-makers for example: the average mean absolute short-term revision to GDP is roughly 1.4 and 0.7 percentage points for annualized quarter-on-quarter and year-on-year growth rates, respectively. Moreover, the standard deviation of revisions is roughly 1.6 - 2 percentage points at an annualized quarter-on-quarter growth rate and roughly 0.8 percentage points for year-on-year growth rate.

If these revisions reflect new information that was not available at the time of the initial release – the revisions are news - then there is little that can be done about the revisions. But when the revisions are not news and can be predicted, we would like to do so, to improve our understanding of the state of the economy. The predictability of such revisions could be used to improve decision-making of agents and policymakers, since their optimal choices depend on the state of the economy. By using the information available at the time of initial announcement, we found that many variables are predictable in-sample. To see whether we would be able to utilize the in-sample predictability in real-time we performed a proper out-of-sample exercise. On the whole, revisions are not easily predictable in real-time: for most variables zero-revision forecast works best. Subject to the caveats of small sample size and the inclusion of the crisis period, we found that only revisions to GDP deflator can be predicted with substantial gains over zero-revisions benchmark. Revisions to consumption and to year-on-year exports and imports can be predicted with some gains. There are no gains in predicting real GDP, gross fixed capital formation and government consumption.

Our analysis is a first step toward a deeper understanding of the size and the nature of revisions to Czech macroeconomic data. A natural extension (once there are more observations of revisions available) would be to use a state-space model in the spirit of Jacobs & van Norden (2011) to characterize the revision process.

References

Amato, J. D. & N. R. Swanson (2001): "The real time predictive content of money for output." *Journal of Monetary Economics* **48**: pp. 3–24.

Aruoba, S. B. (2008): "Data Revisions Are Not Well Behaved." *Journal of Money, Credit and Banking* **40(2-3)**: pp. 319–340.

de Castro, F., J. J. Pérez, & M. Rodríguez-Vives (2011): "Fiscal data revisions in Europe." *Working Paper Series 1342*, European Central Bank.

Christoffersen, P., E. Ghysels, & N. Swanson (2002): "Let's Get 'Real' About Using Economic Data." *Journal of Empirical Finance* **9(3)**: pp. 343–360.

Clark, T. E. & K. D. West (2007): "Approximately normal tests for equal predictive accuracy in nested models." *Journal of Econometrics* **138(1)**: pp. 291–311.

Croushore, D. (2011): "Frontiers of Real-Time Data Analysis." *Journal of Economic Literature* **49(1)**: pp. 72–100.

Croushore, D. & T. Stark (2001): "A real-time data set for macroeconomists." *Journal of Econometrics* **105(1)**: pp. 111–130.

Faust, J., J. H. Rogers, & J. H. Wright (2005): "News and Noise in G-7 GDP Announcements." *Journal of Money, Credit and Banking* **37(3)**: pp. 403–19.

Fernandez, A. Z., E. F. Koenig, & A. Nikolsko-Rzhevskyy (2011): "A real-time historical database for the OECD." *Globalization and Monetary Policy Institute Working Paper 96*, Federal Reserve Bank of Dallas.

Garratt, A. & S. P. Vahey (2006): "UK Real-Time Macro Data Characteristics." *Economic Journal* **116(509)**: pp. F119–F135.

GIANNONE, D., J. HENRY, M. LALIK, & M. MODUGNO (2012): "An Area-Wide Real-Time Database for the Euro Area." *The Review of Economics and Statistics* **94(4)**: pp. 1000–1013.

JACOBS, J. P. & S. VAN NORDEN (2011): "Modeling data revisions: Measurement error and dynamics of "true" values." *Journal of Econometrics* **161(2)**: pp. 101–109.

KAPETANIOS, G. & T. YATES (2010): "Estimating time variation in measurement error from data revisions: an application to backcasting and forecasting in dynamic models." *Journal of Applied Econometrics* **25(5)**: pp. 869–893.

KUGLER, P., T. J. JORDAN, C. LENZ, & M. R. SAVIOZ (2005): "GDP data revisions and forward-looking monetary policy in Switzerland." *The North American Journal of Economics and Finance* **16(3)**: pp. 351–372.

MANKIW, N. & M. SHAPIRO (1986): "News or Noise: an Analysis of GNP Revisions." *Survey of Current Business* **(May 1986)**: pp. 20–25.

MANKIW, N. G., D. E. RUNKLE, & M. D. SHAPIRO (1984): "Are Preliminary Announcements of the Money Stock Rational Forecasts?" *Journal of Monetary Economics* **14**: pp. 15–27.

MCKENZIE, R. (2006): "Undertaking Revisions and Real-Time Data Analysis using the OECD Main Economic Indicators Original Release Data and Revisions Database." *OECD Statistics Working Papers 2006/2*, OECD Publishing.

MINCER, J. A. & V. ZARNOWITZ (1969): "The evaluation of economic forecasts." In "Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance," NBER Chapters, pp. 1–46. National Bureau of Economic Research.

MOLODTSOVA, T., A. NIKOLSKO-RZHEVSKYY, & D. H. PAPELL (2008): "Taylor rules with real-time data: A tale of two countries and one exchange rate." *Journal of Monetary Economics* **55**: pp. S63–S79.

Newey, W. K. & K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* **55(3)**: pp. 703–08.

Orphanides, A. (2001): "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* **91(4)**: pp. 964–985.

Orphanides, A. (2003): "Monetary policy evaluation with noisy information." *Journal of Monetary Economics* **50(3)**: pp. 605–631.

Orphanides, A. & S. van Norden (2002): "The Unreliability of Output-Gap Estimates in Real Time." *The Review of Economics and Statistics* **84(4)**: pp. 569–583.

Robertson, J. C. & E. W. Tallman (1998): "Data Vintages and Measuring Forecast Model Performance." *Federal Reserve Bank of Atlanta Economic Review* **b**: pp. 4–20.

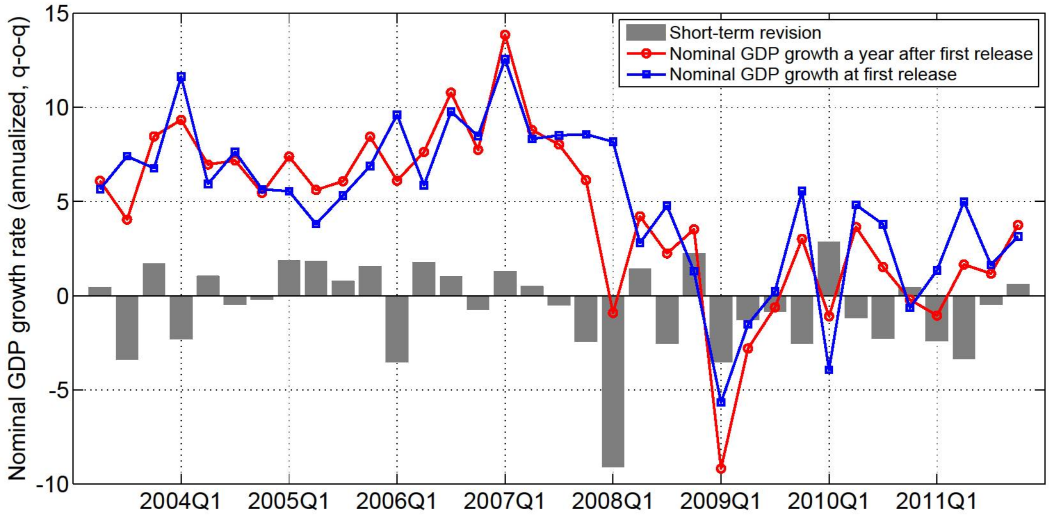
Sargent, T. J. (1989): "Two Models of Measurements and the Investment Accelerator." *Journal of Political Economy* **97(2)**: pp. 251–87.

Swanson, N. R. & H. White (1997): "A Model Selection Approach To Real-Time Macroeconomic Forecasting Using Linear Models And Artificial Neural Networks." *Review of Economics and Statistics* pp. 540–550.

6.A Short-Term and Long-Term Revisions

Figure 6.2: Revisions to Nominal GDP (Annualized quarterly growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

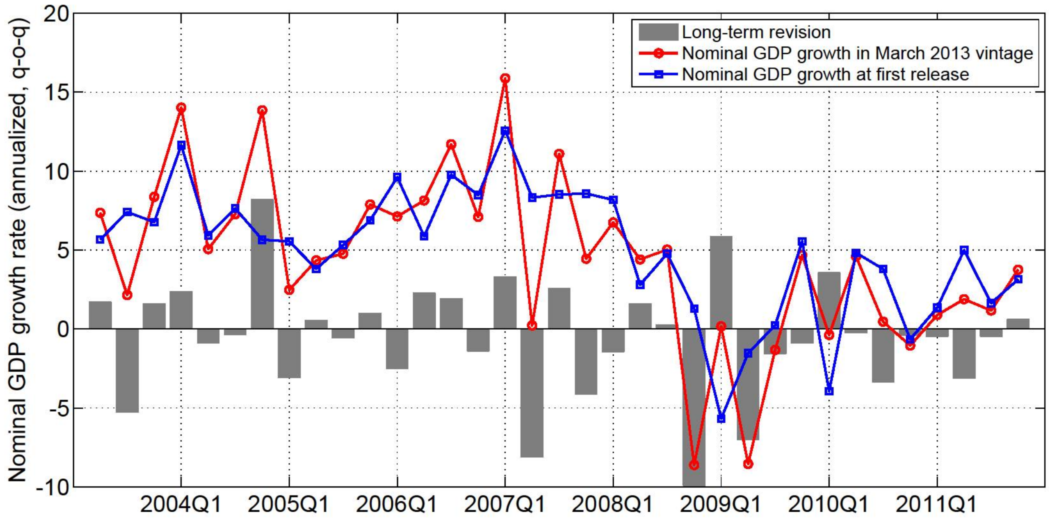
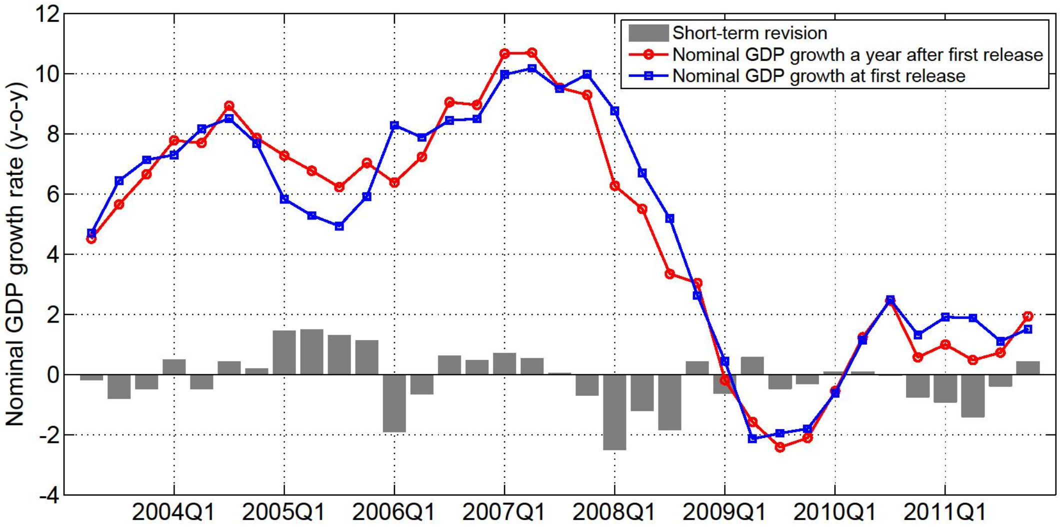


Figure 6.3: Revisions to Nominal GDP (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

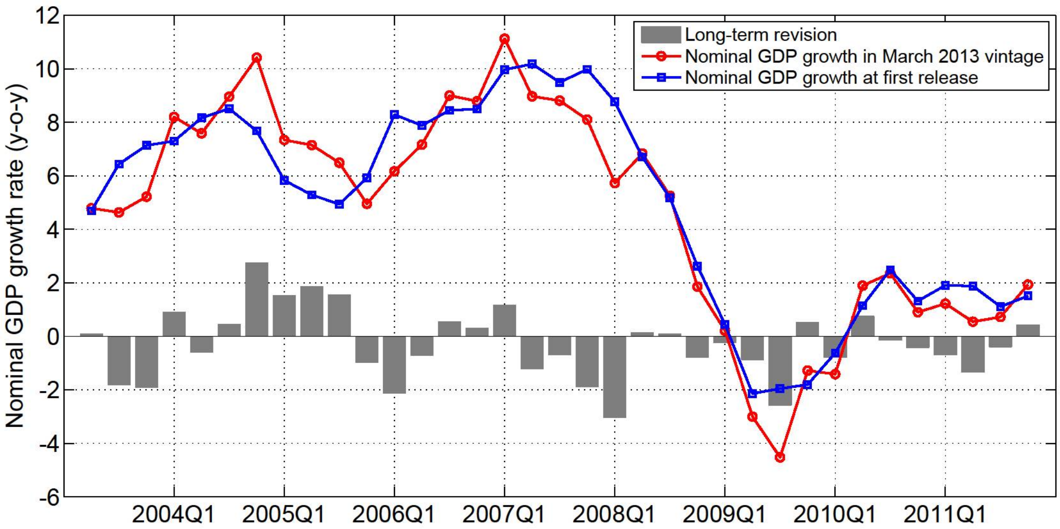
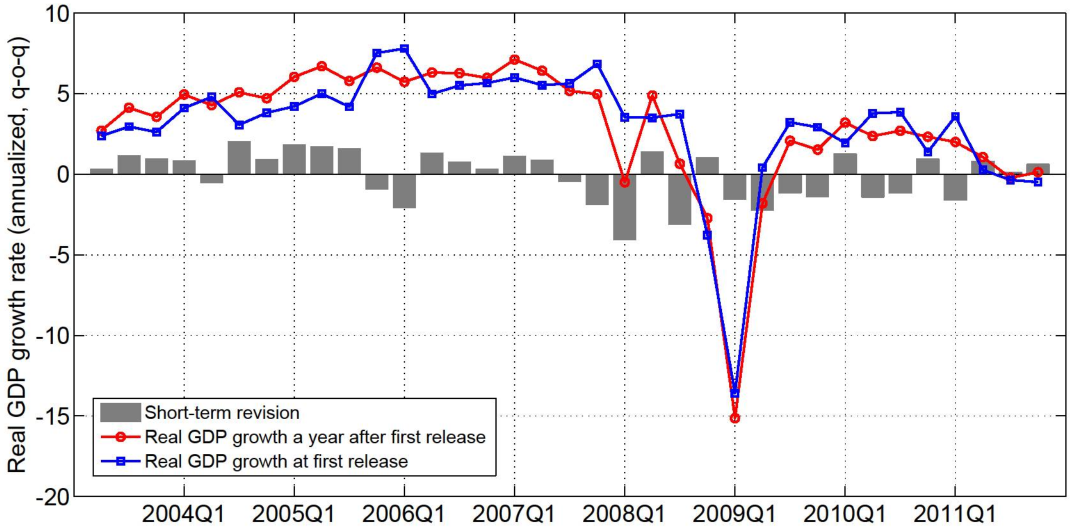


Figure 6.4: Revisions to Real GDP (Annualized quarterly growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

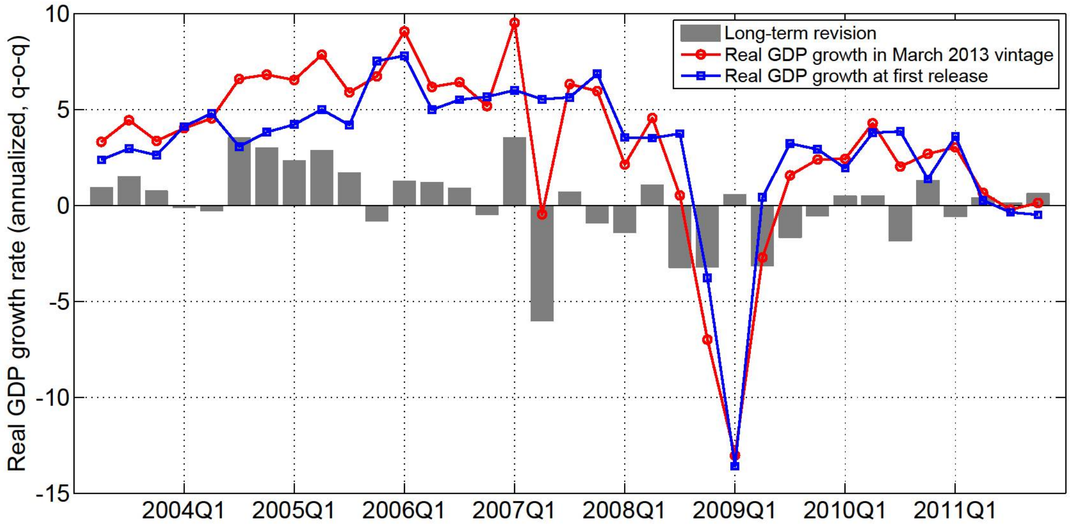
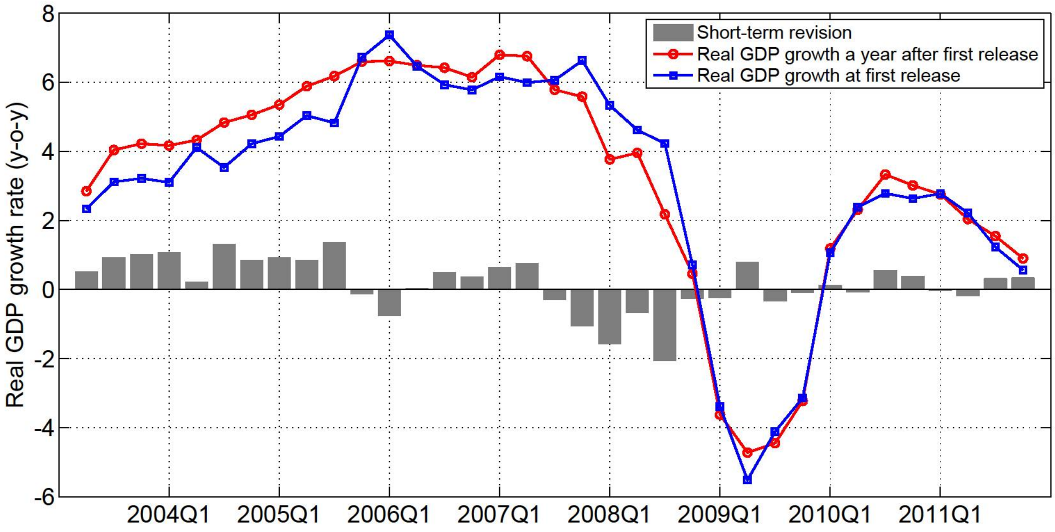


Figure 6.5: Revisions to Real GDP (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

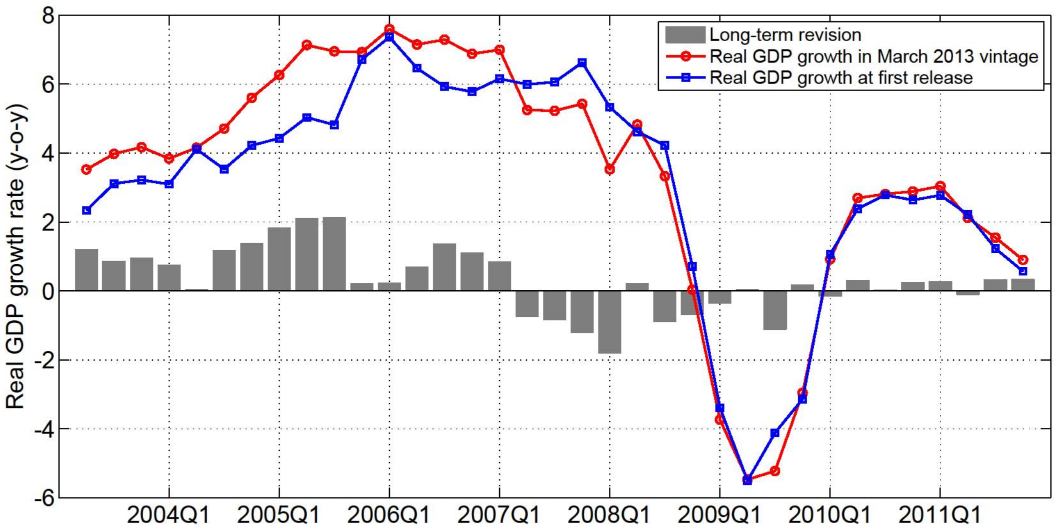
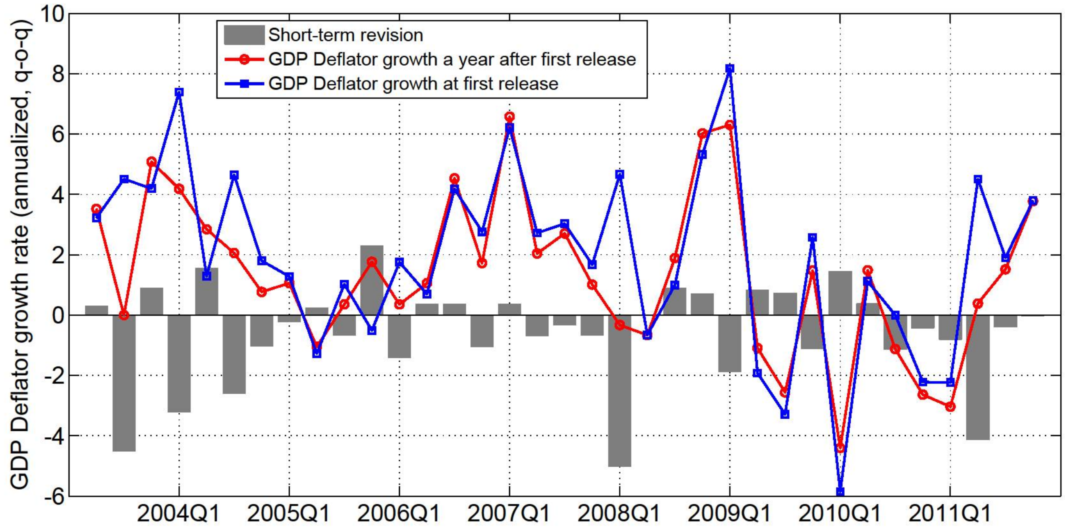


Figure 6.6: Revisions to GDP Deflator (Annualized quarterly growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

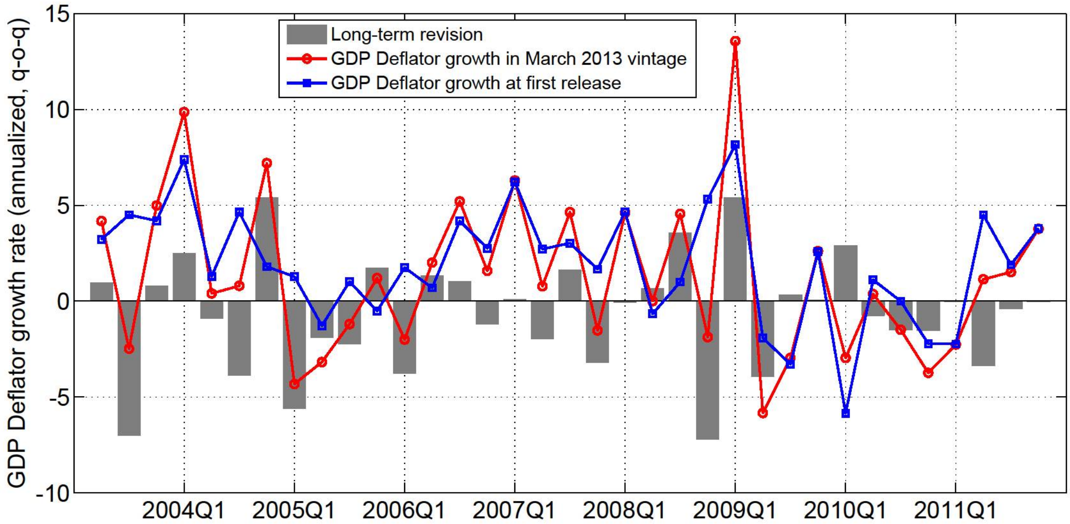
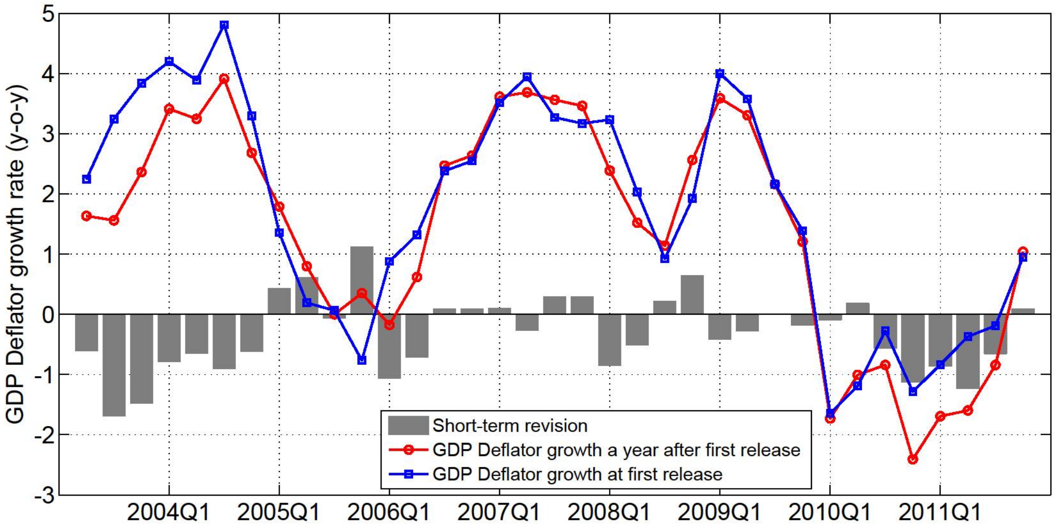


Figure 6.7: Revisions to GDP Deflator (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

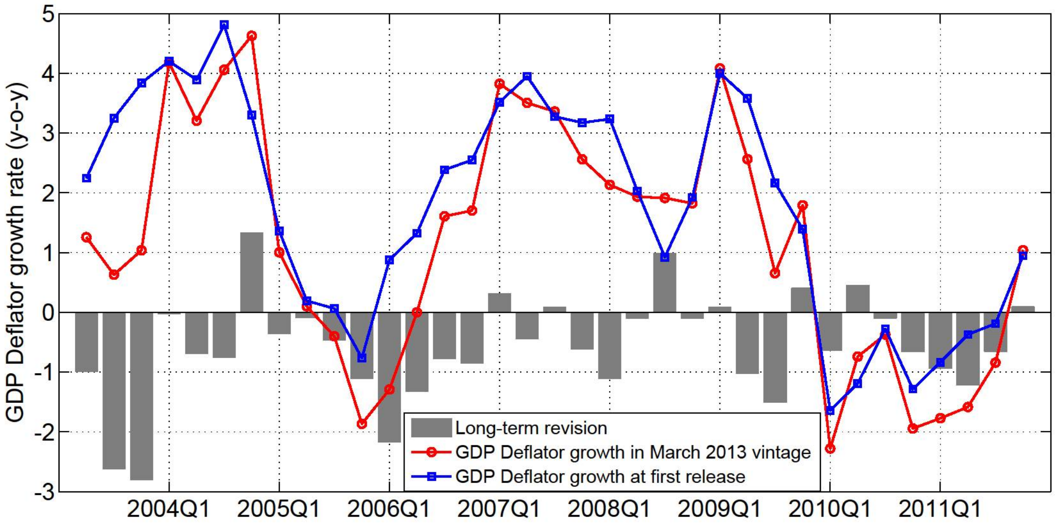
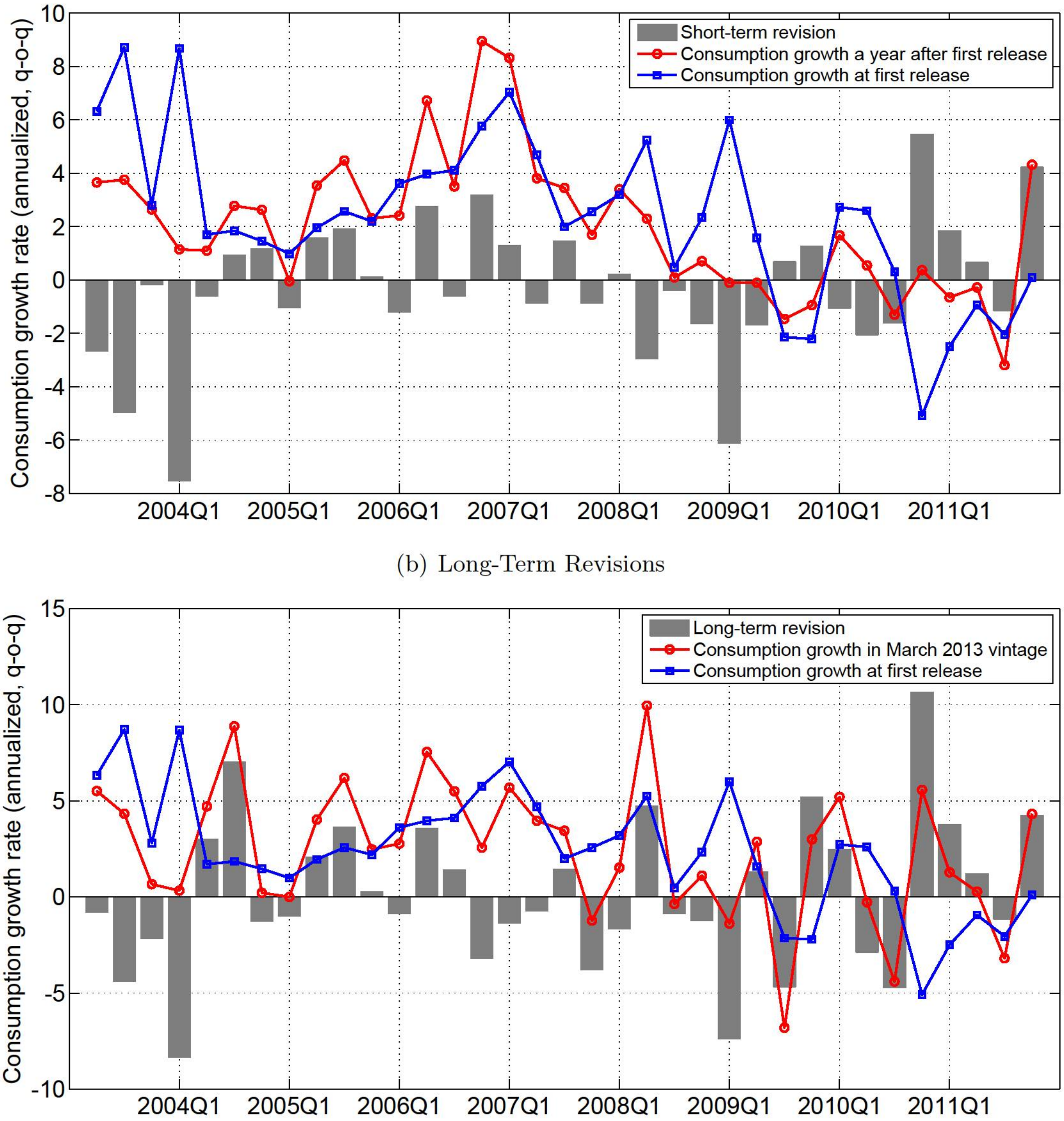


Figure 6.8: Revisions to Consumption (Annualized quarterly growth rates)

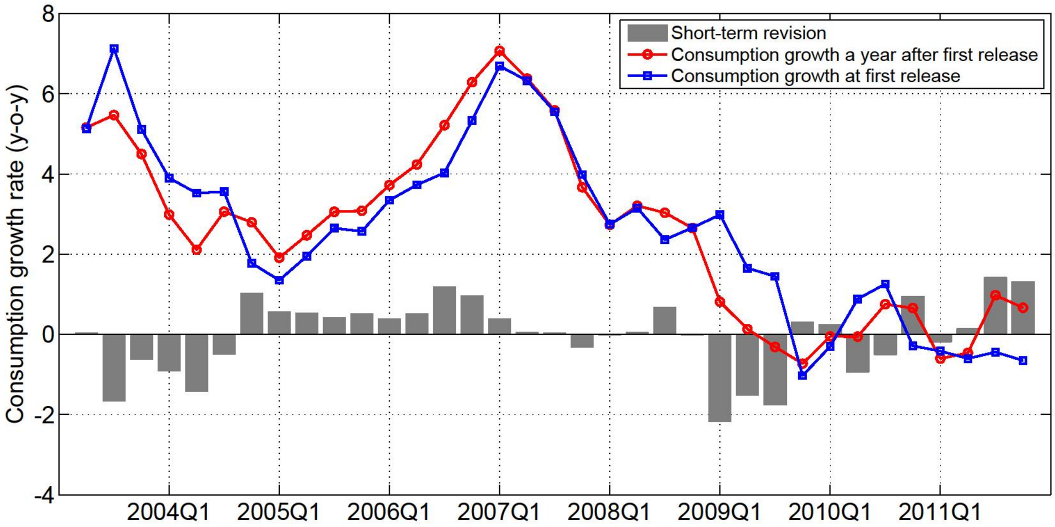
(a) Short-Term Revisions



(b) Long-Term Revisions

Figure 6.9: Revisions to Consumption (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

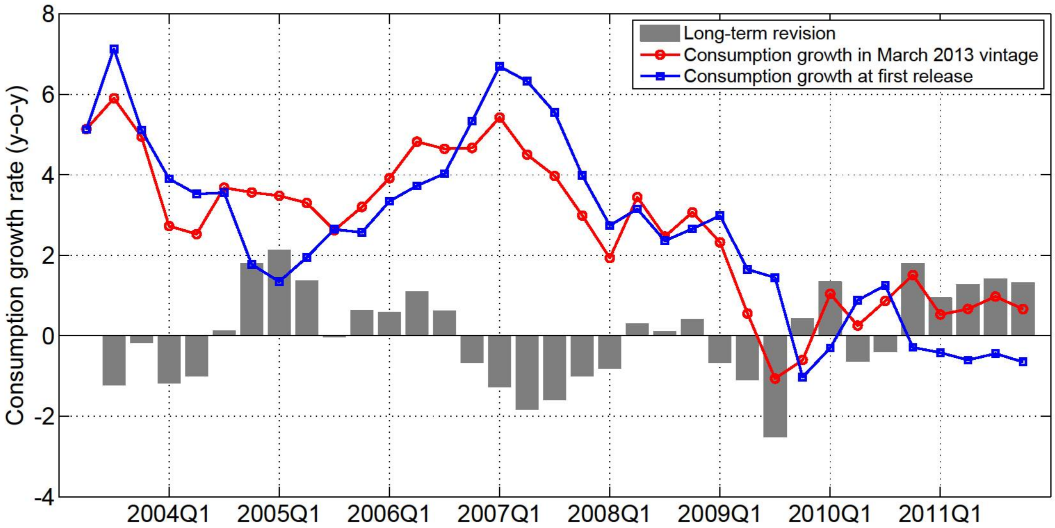
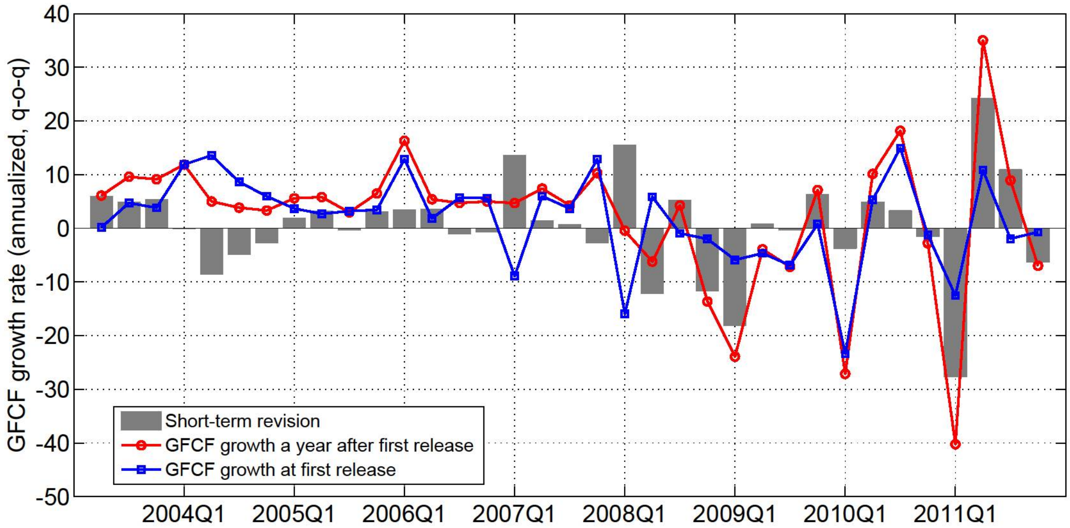


Figure 6.10: Revisions to Gross Fixed Capital Formation (Annualized quarterly growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

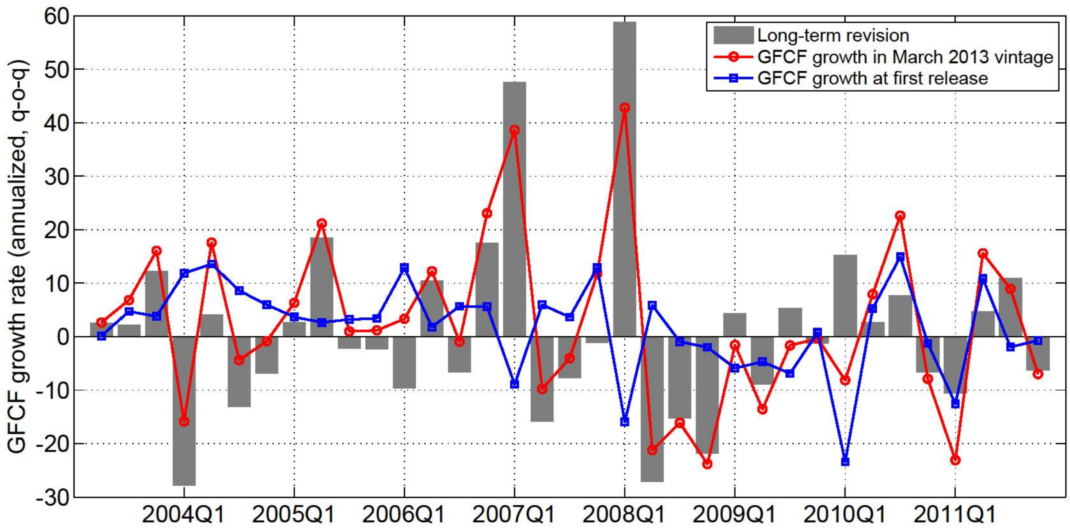
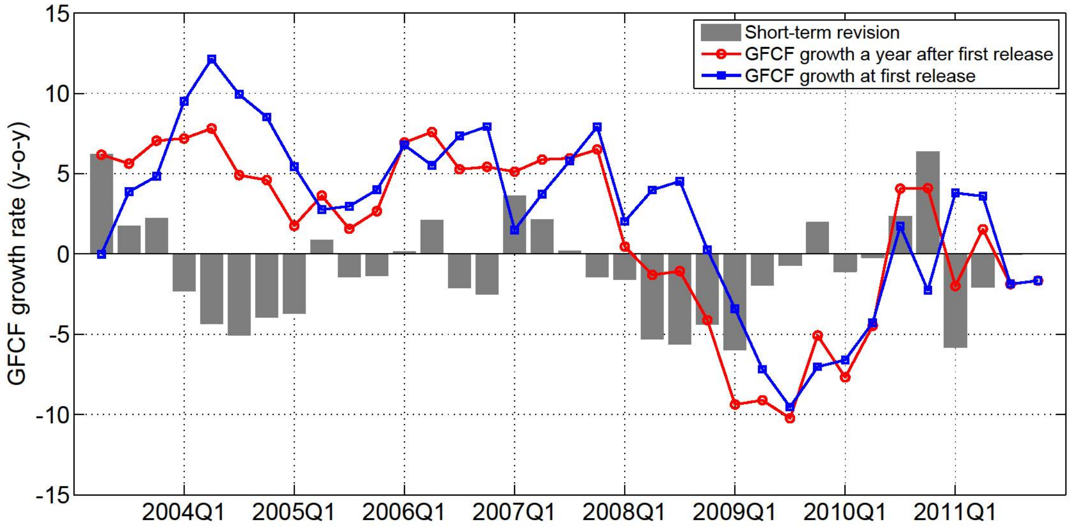


Figure 6.11: Revisions to Gross Fixed Capital Formation (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

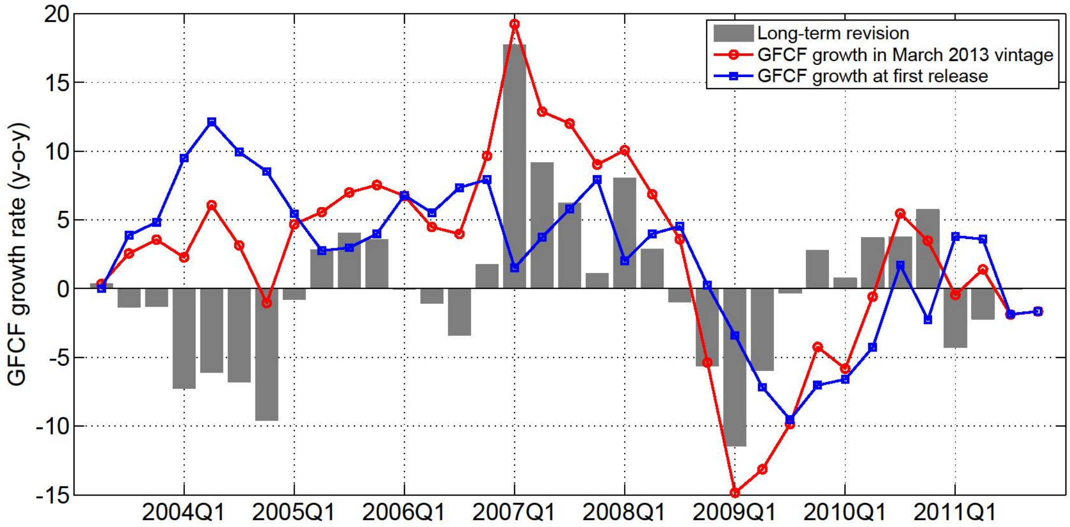
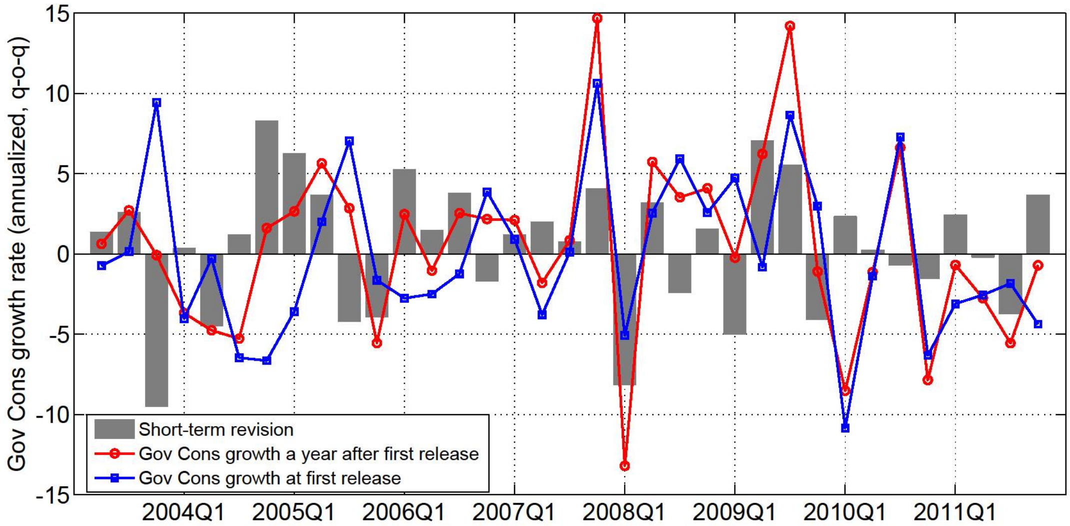


Figure 6.12: Revisions to Government Consumption (Annualized quarterly growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

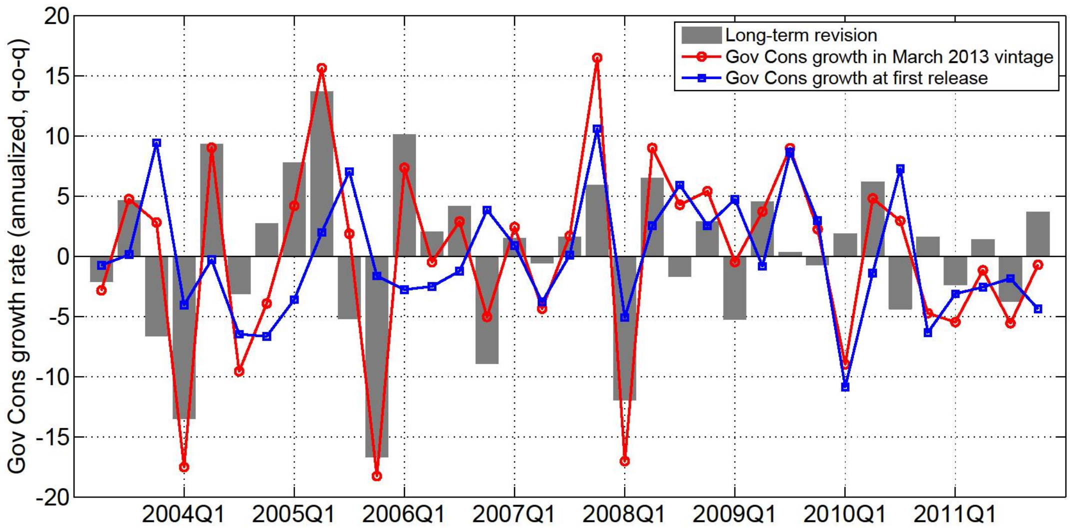
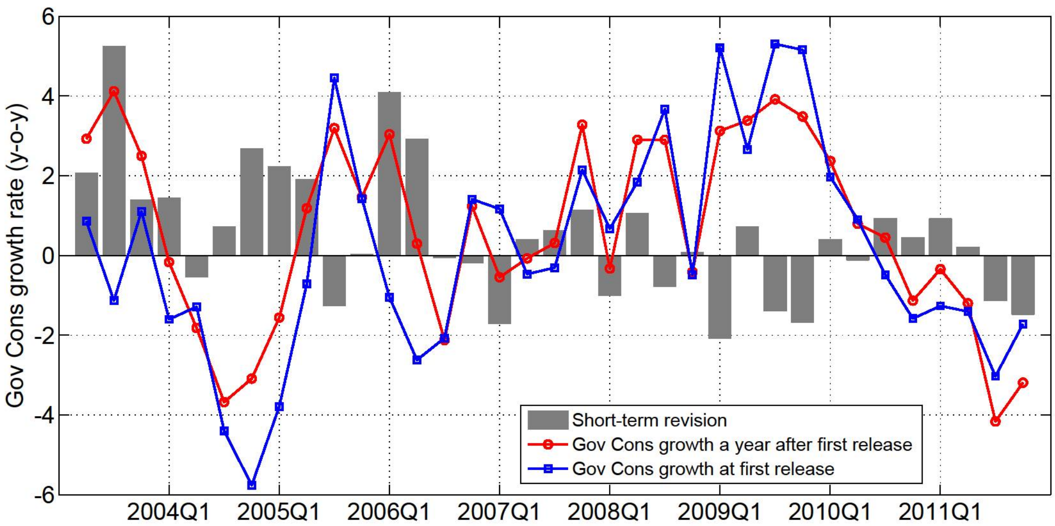


Figure 6.13: Revisions to Government Consumption (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

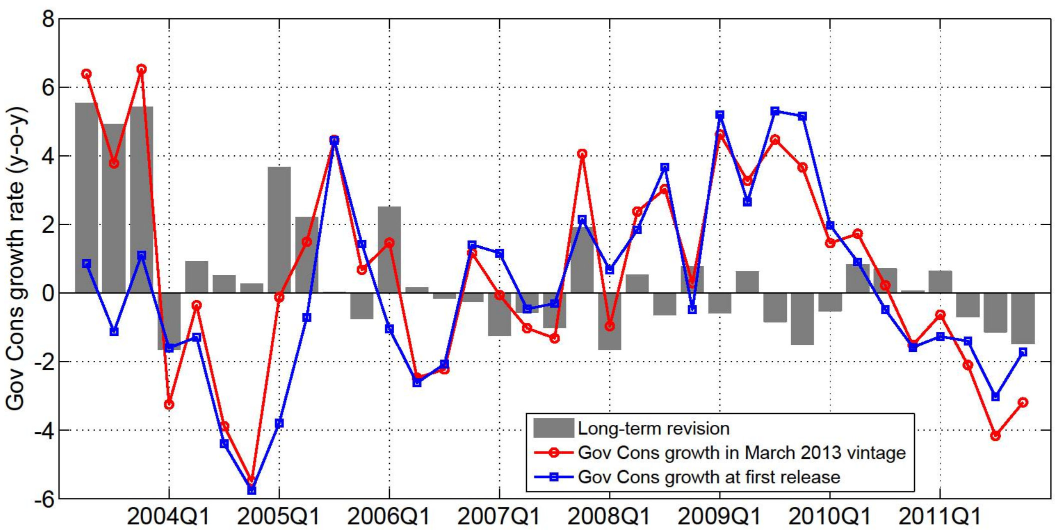
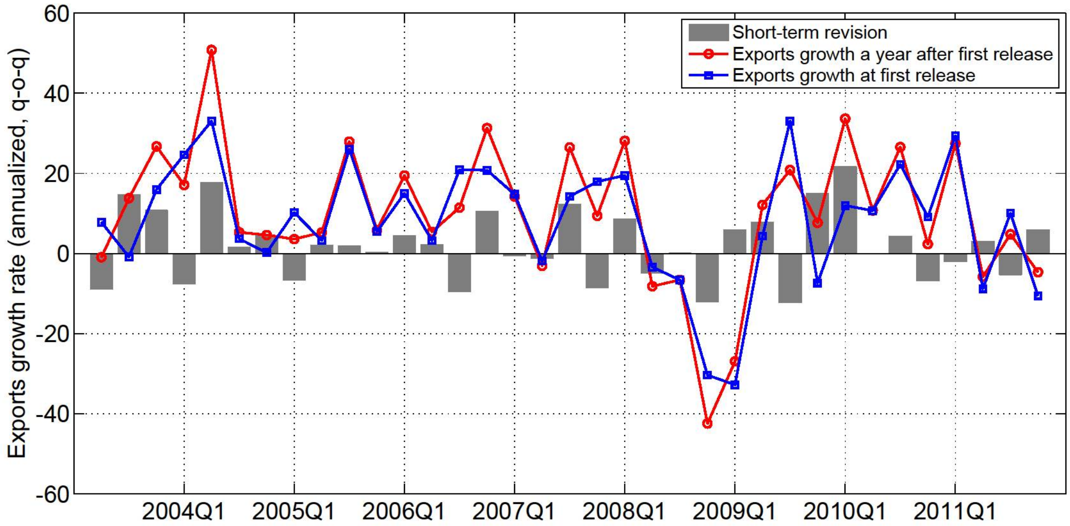


Figure 6.14: Revisions to Exports (Annualized quarterly growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

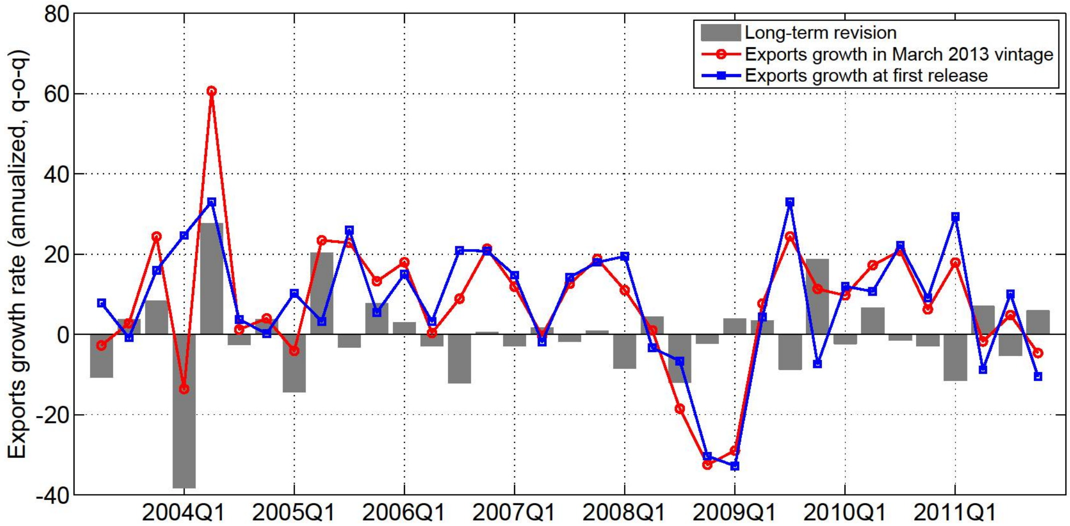
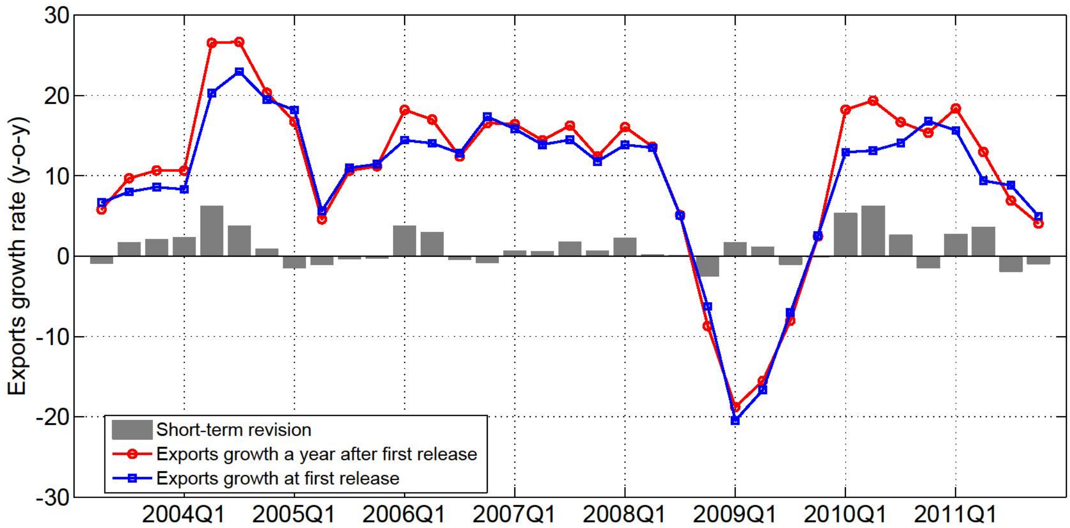


Figure 6.15: Revisions to Exports (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

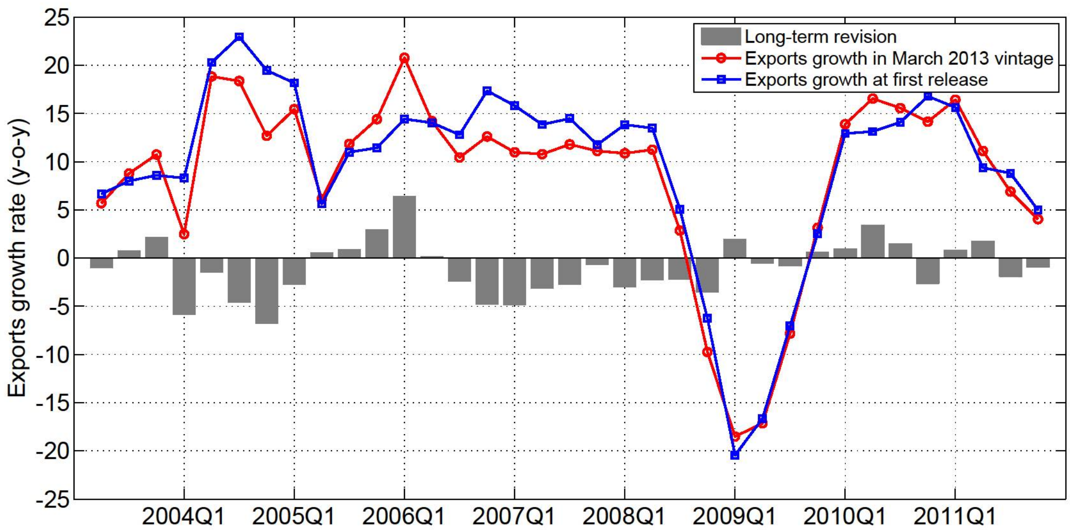
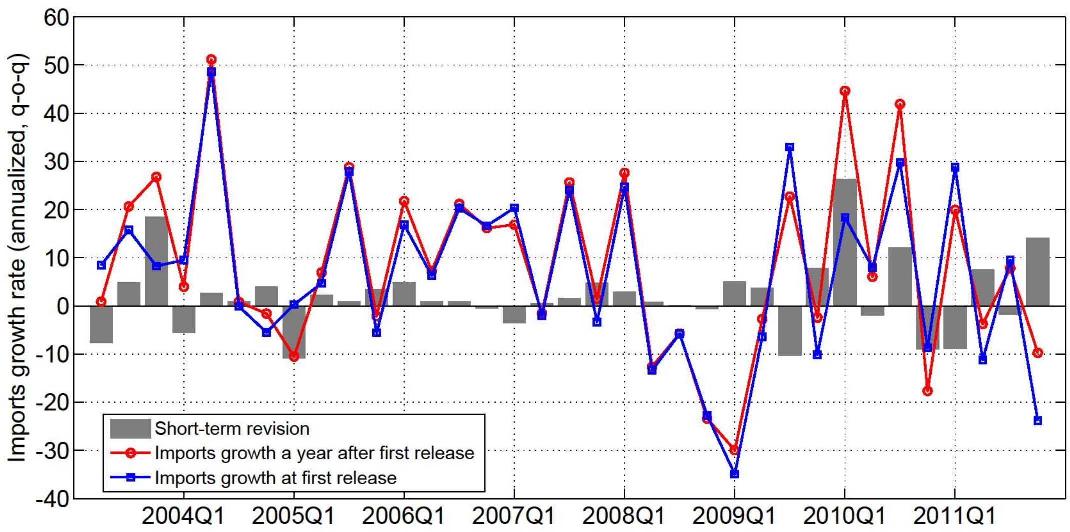


Figure 6.16: Revisions to Imports (Annualized quarterly growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions

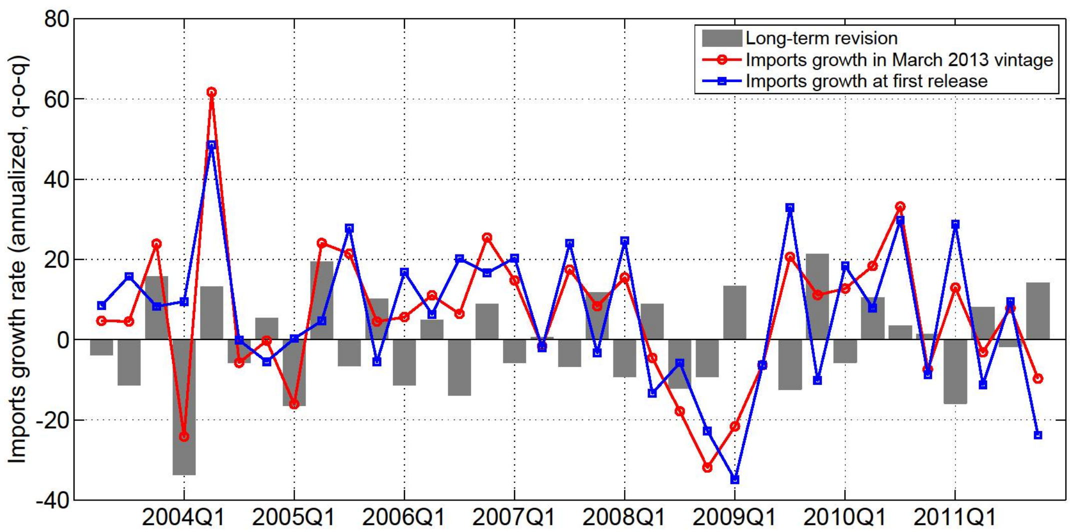
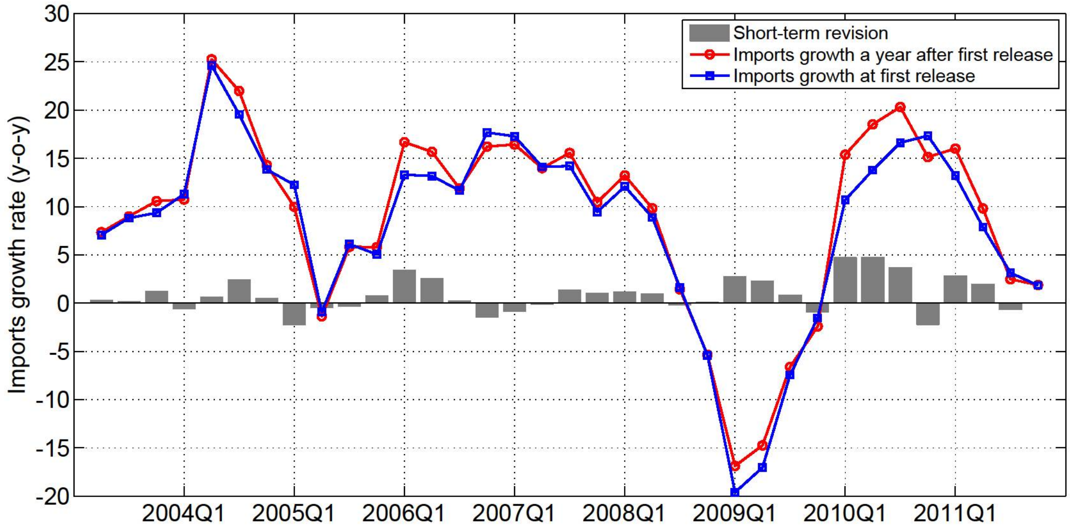
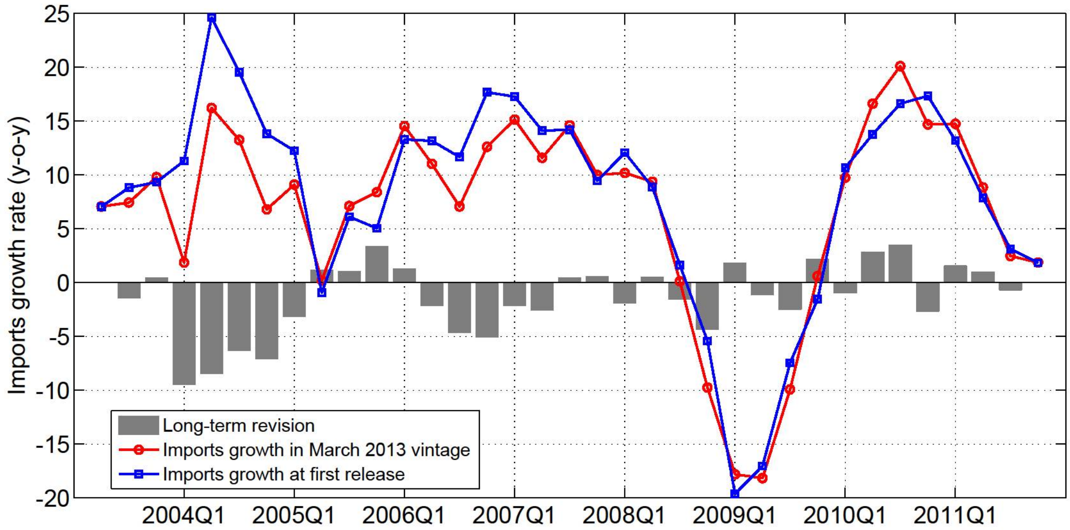


Figure 6.17: Revisions to Imports (Annual growth rates)

(a) Short-Term Revisions



(b) Long-Term Revisions



6.B Historical Vintages

Figure 6.18: Growth rates of various Nominal GDP vintages

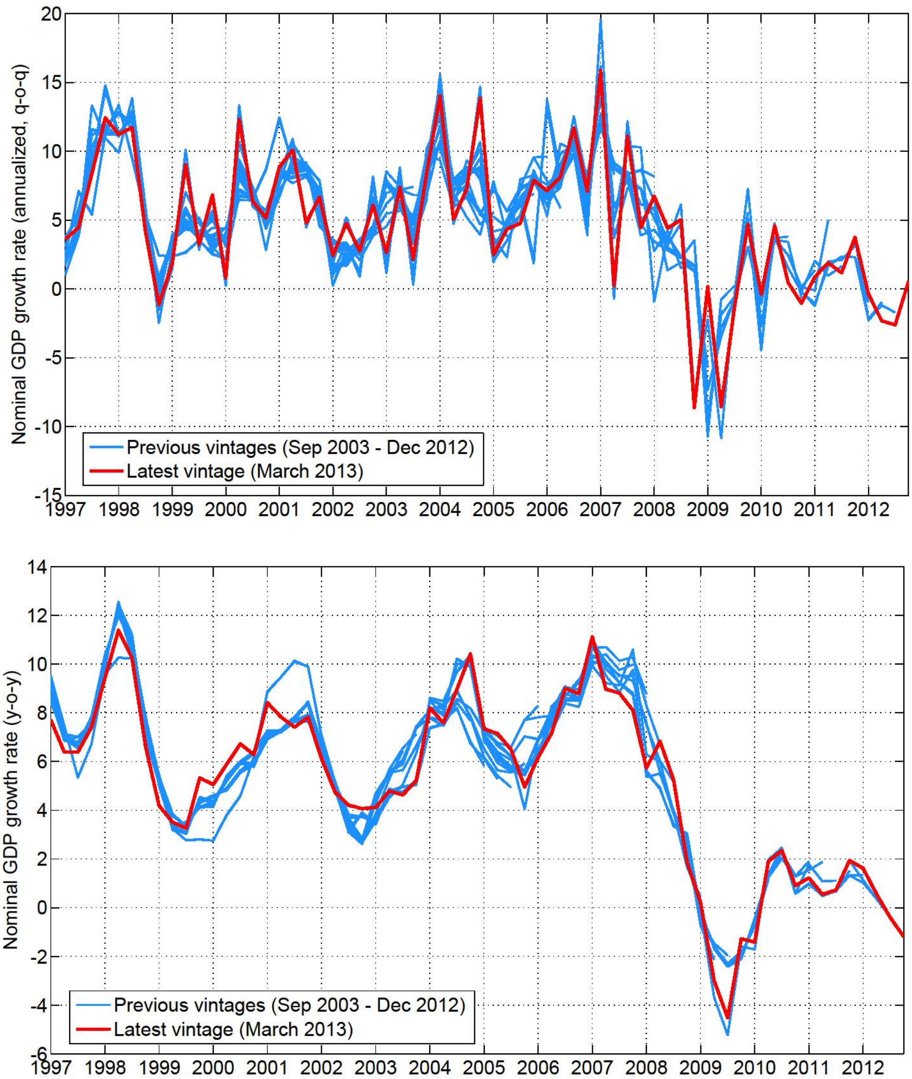


Figure 6.19: Growth rates of various Real GDP vintages

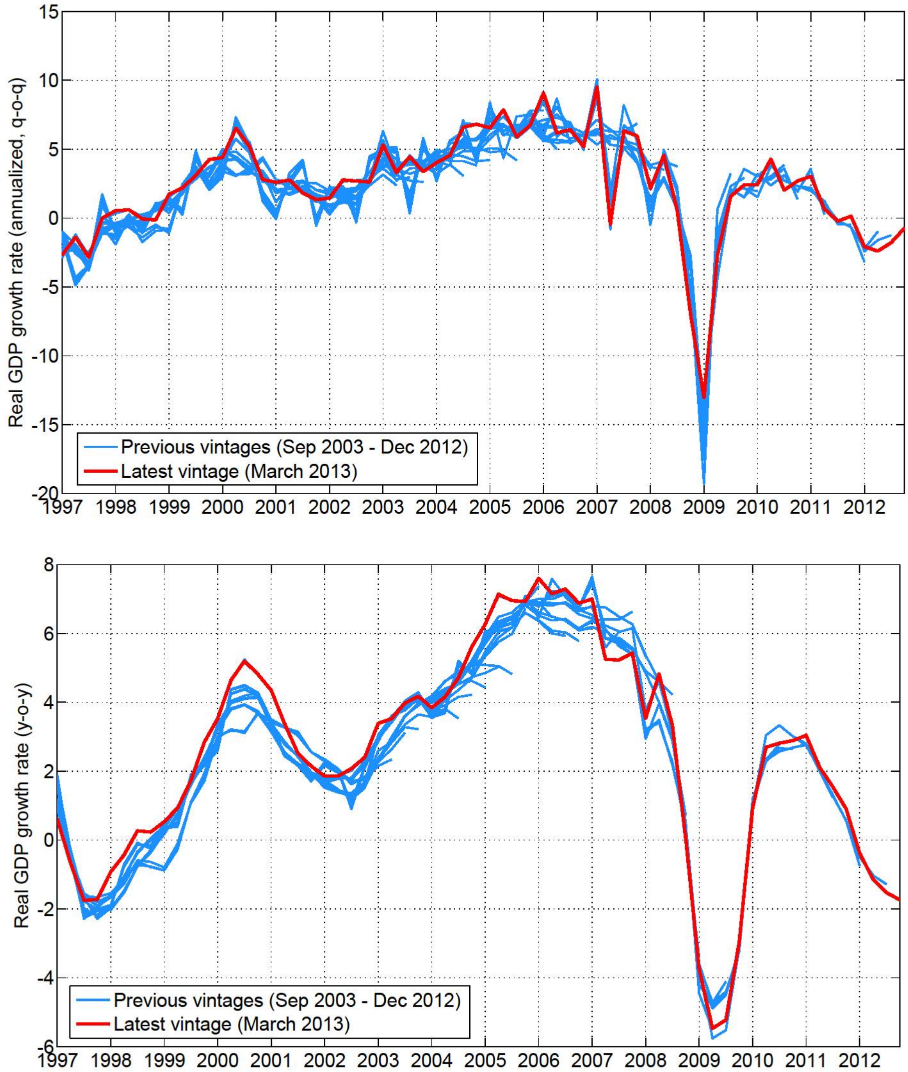


Figure 6.20: Growth rates of various GDP Deflator vintages

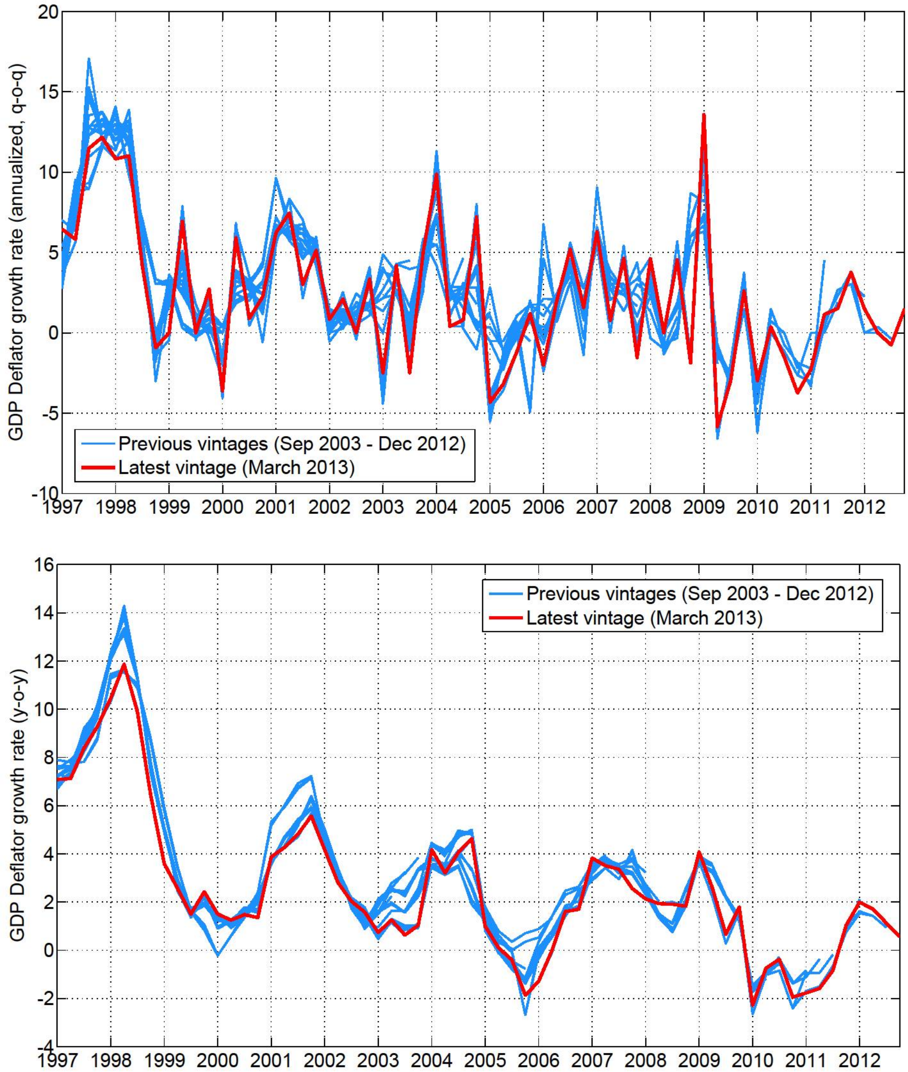


Figure 6.21: Growth rates of various Consumption vintages

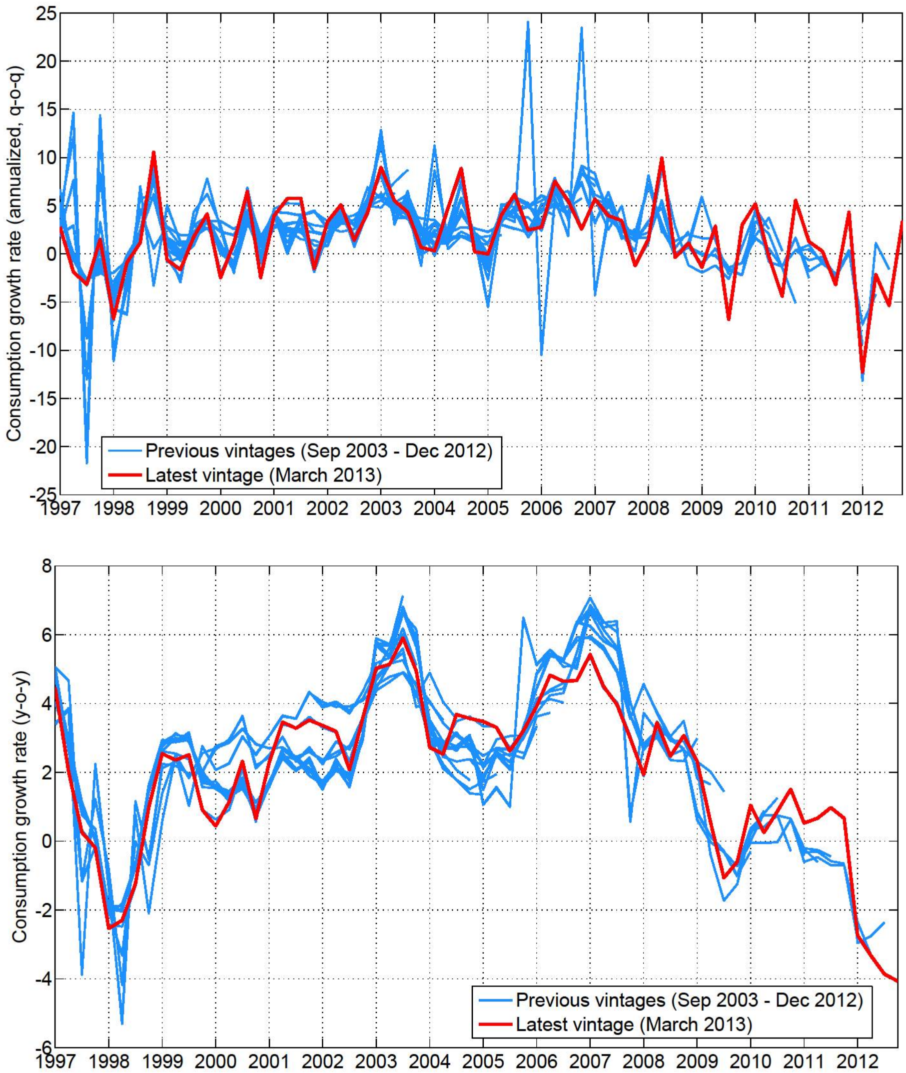


Figure 6.22: Growth rates of various GFCF vintages

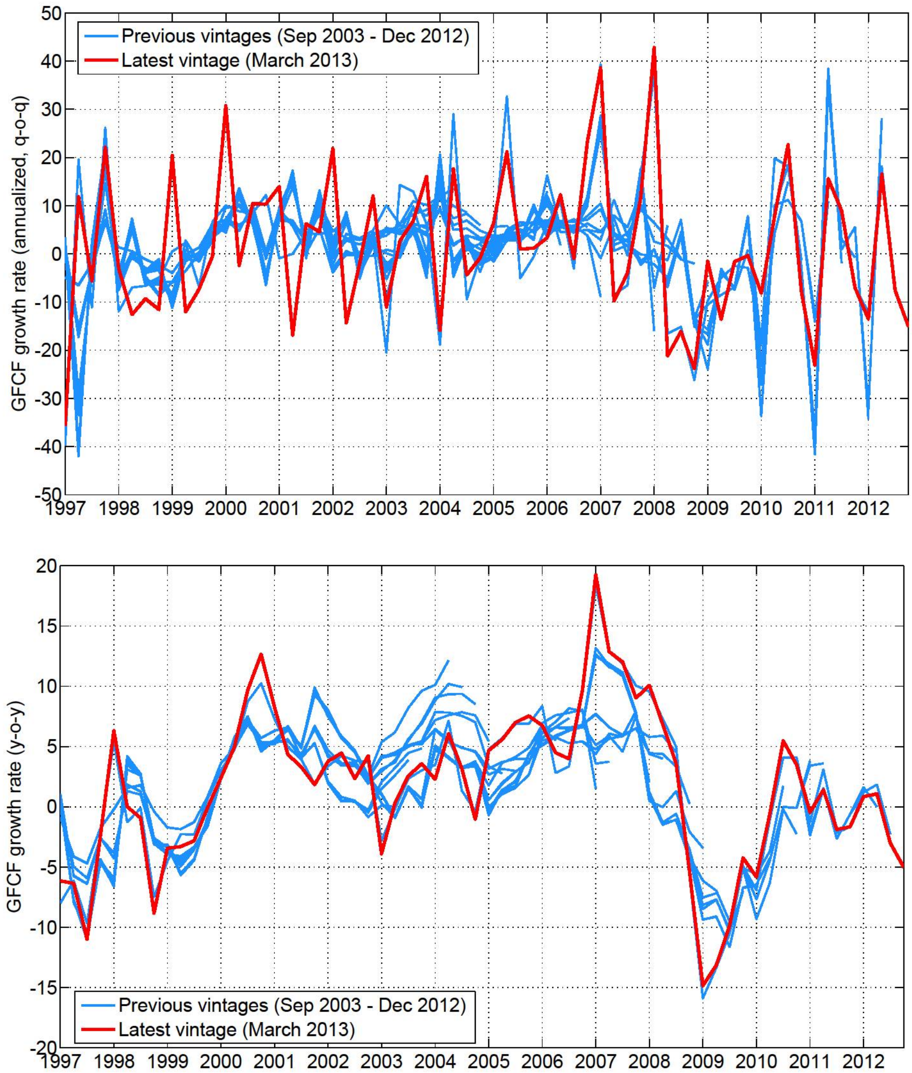


Figure 6.23: Growth rates of various Gov. Consumption vintages

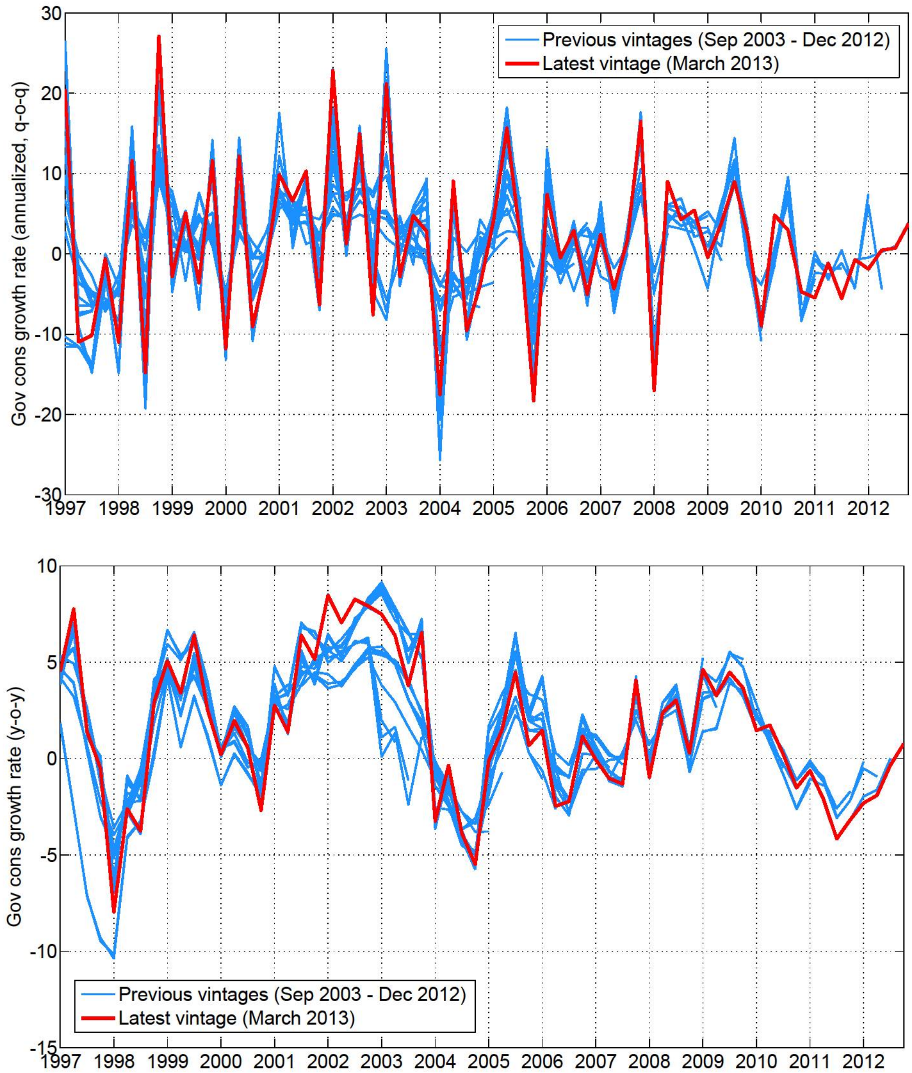


Figure 6.24: Growth rates of various Exports vintages

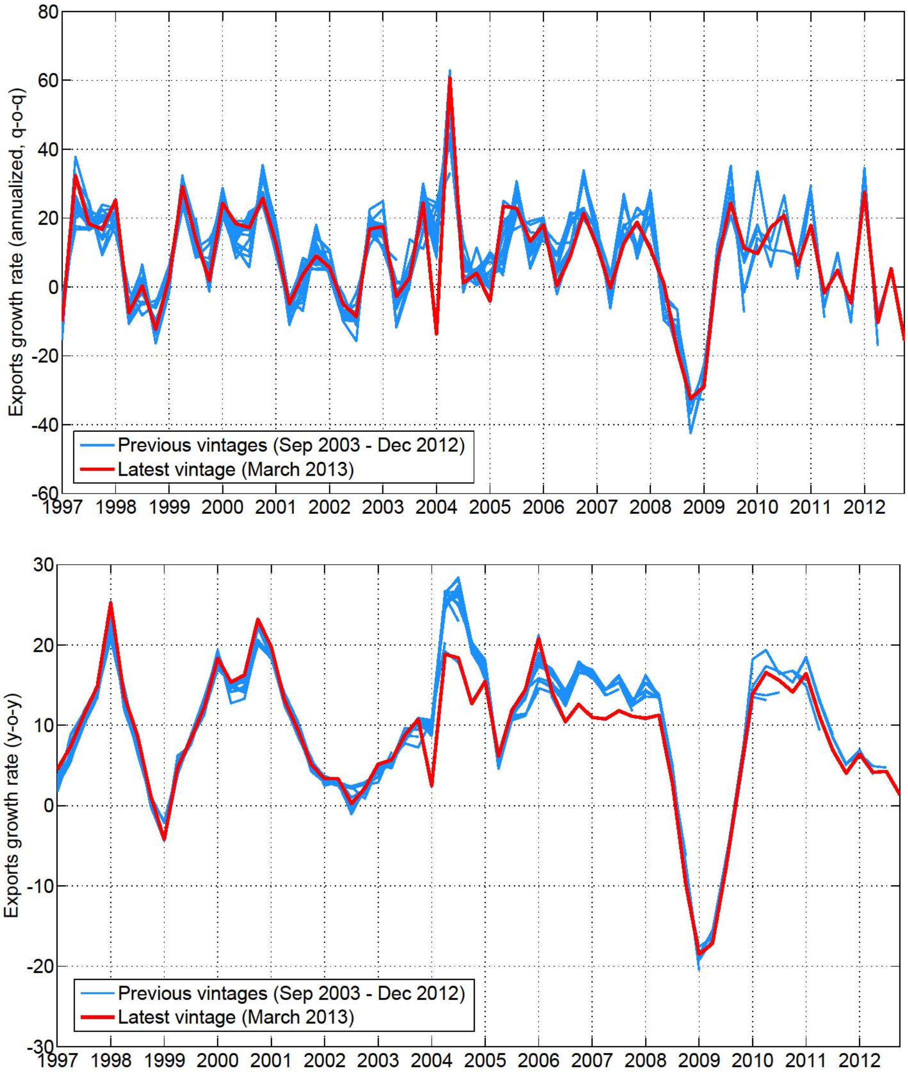
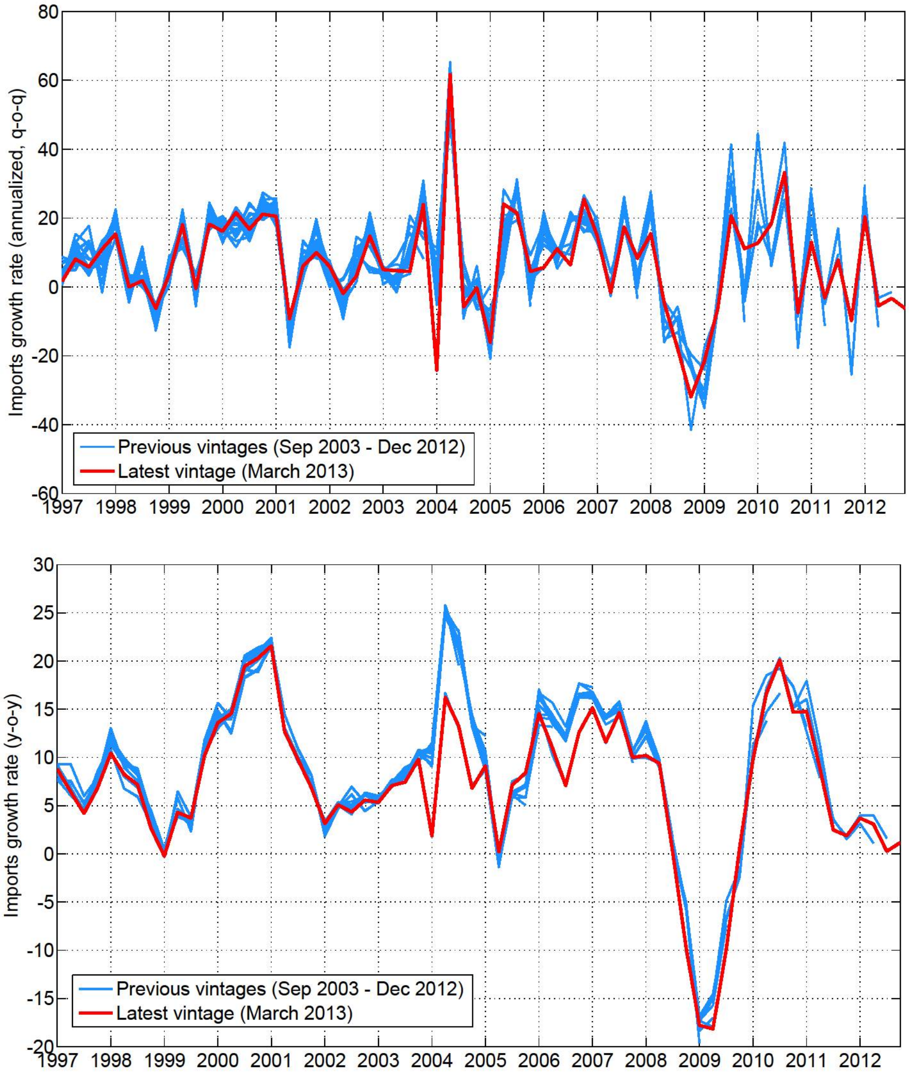


Figure 6.25: Growth rates of various Imports vintages



6.C Preliminary Estimates of Real GDP

Figure 6.26: Scatter plot for Preliminary Estimates of Real GDP (q-o-q) (2007Q4-2012Q2)

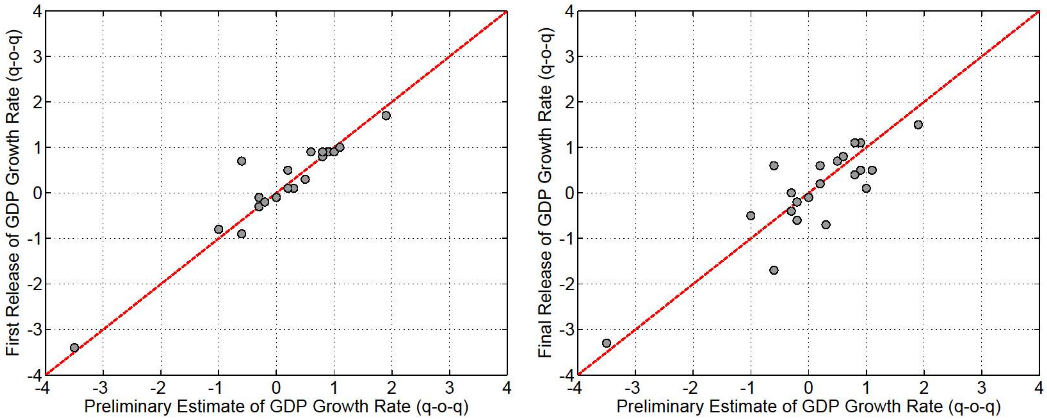
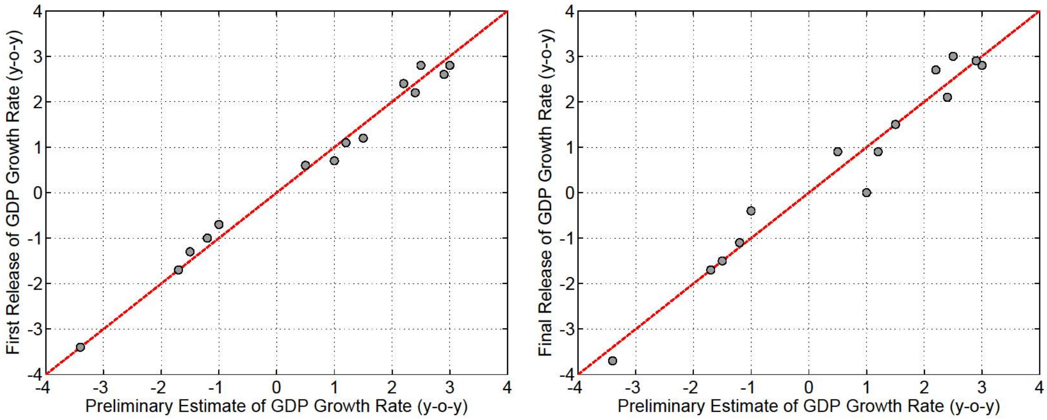


Figure 6.27: Scatter plot for Preliminary Estimates of Real GDP (y-o-y) (2007Q4-2012Q2)



6.D Illustration of the severity of revisions

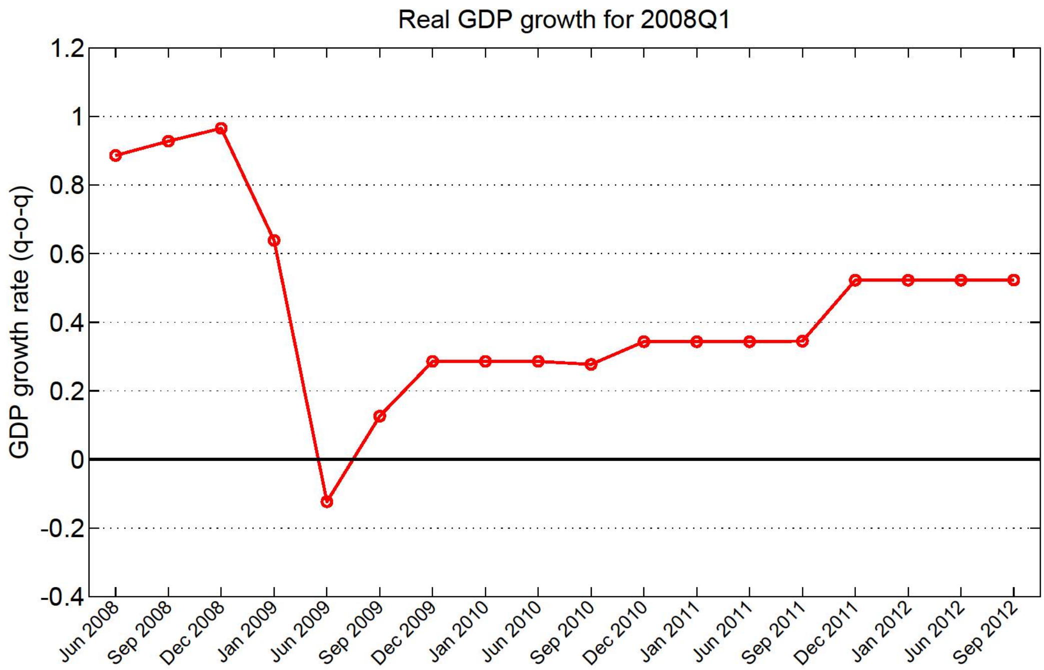


Figure 6.28: Real GDP growth for 2008Q1 (vintages: Jun 2008 - Sep 2012)

Appendix A

Response to Referees

I am grateful to all the referees for their comments and useful suggestions in their referee reports and believe these helped us improve the dissertation. The main comments by the referees are typeset in roman; my response is in italics.

A.1 PhDr. Jaromir Baxa, Ph.D.

... All papers are original, nicely written and all have interesting implications either for monetary policy or for empirical modelling. Let me also mention that four of five papers have been published already and some of them in top journals including Journal of Money, Credit and Banking and International Journal of Central Banking. The last paper, the essay on habit formation, is currently at revise and resubmit stage. Besides papers included in this dissertation, Marek Rusnák has an impressive record of publications. Hence, there's not much space to make objections against papers included in the dissertation and I don't feel there is anything that shall be rewritten for the final defense. ...

I would like to thank Jaromir Baxa for his kind words and very thoughtful referee report.

1. The authors suggest that recursive identification is inferior to alternatives and show that identification via sign restrictions does lead to price puzzle only rarely. From my point of view, this result from meta-analysis is not surpris-

ing. Uhlig (2005) and many others rule out price puzzle by assuming that the response of inflation or price level on interest rate is negative. On the other hand, some authors do not make such restrictive assumptions. Castelnuovo and Sourico (2006) impose sign restrictions on interest rate and output but not on price level and so they allow to speak the data about inflation freely. By this, the authors show that identification via sign-restrictions is not enough to get rid of price puzzle within the U.S. data prior the Volcker disinflation (I think it's Figure 11 in WP version of their paper). I'm unable to assess the proportion of studies where sign restrictions are imposed on the response of price level/inflation but I find important to add that the way how sign restrictions are incorporated seems to matter a lot.

This is a very good point. Indeed, the way how sign restriction are incorporated might matter a lot: it might be important to distinguish between cases where sign restrictions are imposed on the price level response and cases where the response of price level is not restricted. We have now acknowledge this in the text when describing the sign restrictions approach. Unfortunately, controlling for the two cases is not feasible as this would lead to very small number of cases in which dummy variables would be non-zero. For example, for the shortest horizon, the sign-restriction dummy is non-zero in only 31 cases out of 208.

2. Many studies, including Balke and Emery (1994), show that price puzzle appears when monetary policy was passive thus they favour the "monetary regime matters" hypothesis for the reason why price puzzle occurs. However, isn't the result for the post 1979 period driven mainly by the Volcker disinflation rather than being a proper description of the whole period? This seems to be implied by a recent contribution by Coibion (2012; not included in meta- analysis and not mentioned among papers excluded from the meta-analysis intentionally). The author replicates the narrative approach by Romer and Romer (2004). In Romer and Romer paper, no price puzzle appears although the response of price level is almost zero in the first periods. But as Coibion (2012) shows,

when the 1979-1983 period with Volcker disinflation is excluded from the sample, the price-puzzle reappears. Similar implications of the Volcker disinflation can be derived from Barakchian and Crowe (2013) who compare alternative identification schemes on the U.S. data from 1988 to 2007. If the hypothesis of the importance of the Volcker's disinflation for the shape of impulse responses is true, it implies that nicely-looking impulse responses appear when sharp disinflations are included in the sample against the hypothesis of the importance of estimating the monetary VAR on a sample with just one monetary policy regime (I do admit one can have doubts whether the Greenspan's policy of 2000's is similar to 1990's or not).

We are thankful for this observation. A paper by Coibion (2012) was not included in the sample, as our search for primary studies concluded in September 2010 (and the first paper of the dissertation was published in February 2013). Nevertheless, we included the reference to Coibion (2012) and Barakchian and Crowe (2013) in the revised version of the dissertation.

3. Over the past decades, monetary policy has become more and more forward-looking and focused on stabilization of inflation expectations. Thus, the path of inflation is conditional on policy decisions about expected inflation and demand shocks in particular causing the impression of positive correlation between interest rate hikes and rising inflation. Cloyne and Hürtgen (2014) took these considerations seriously and using narrative approach and accounting for inflation forecasts they were able to obtain impulse responses of the U.K. without price puzzle. It seems to me that the implications of forward-looking nature of monetary policy are behind a number of recent contributions in the field of empirical identification of monetary policy transmission mechanism. The simulations using a NK- DSGE model Wolf (2016) implies that the forward-looking nature of monetary policy can provide rationale for an existence of price puzzle. My own personal takeaway is that neither changes in the policy rate nor the ex-post observed inflation rate (or price level) match the ideal

concepts of shocks in policy instrument and response of a target of monetary policy perfectly. The other shocks include shocks in inflation expectations and shocks in inflation target while expected inflation belongs to candidate targets. This underlying complexity of monetary policy, not apparent in stylized structural models, could be the core of the price puzzle problem. Sign restrictions can help with identification of monetary policy shock, but not completely (see Castelnuovo, 2016, for a discussion showing that magnitudes can be misleading when estimating the model using sign restrictions) and one needs to admit that to some extent the results are driven by the assumptions.

We agree that controlling for forward looking nature of monetary policy is crucial. Indeed, some of the proposed solutions to the price puzzle are in this spirit - the inclusion of commodity prices, and to a lesser extent also factor augmented vector autoregressions that make use of large number of variables. We are thankful for this observation. The reference to the paper by Cloyne and Hürtgen (2016) and Wolf (2016) is made in the revised version of the thesis.

A.2 Univ.-Prof. Dr. Ansgar Belke

Let me first state that I very much enjoyed reading the dissertation which consists of an introduction motivating the dissertation and giving a structure to it and five academic papers. My more detailed remarks on the papers which are rather well-done in technical terms and are in the fields of meta-analysis in macroeconomics and real-time properties and forecasting are the following. The problems are clearly set out and the papers meet their own aims formulated in the introductions - in terms of rigor of the analysis. The empirical models the author employs have the advantage of being quite standard, simple to understand (which of course does not imply ”easy to handle”) and well-structured. Moreover, the papers are written in a logically consistent and quite fluent style. Finally, the author could in some cases be more explicit in their conclusions about what is left for future research.

...

But the author also succeeds to present a sound policy-oriented analysis as far as this issue is concerned. Since the author applies different state-of-the-art views and methods in a careful and convincing way, I would rate the inherent academic merits of the paper as sufficiently high. Hence, seen on the whole, the collected papers represent an innovation.

Seen on the whole thus, I can recognize an original contribution of the author. And I consider the thesis defendable at my home institution without hesitations. What is more, it is immediately clear thatthe results of the thesis allow their publication in a respected economic journal. Hence, I was not at all surprised to see the papers already published or in a revise and resubmit mode in good (Economic Modelling, International Journal of Central Banking and Czech Journal of Economics and Finance) and very good (European Economic Review and Journal of Money, Credit and Banking) journals.

...

The topics dealt with in the dissertation are still under-investigated in the relevant literature and appear to be applicable to several other policy fields. What is more, the author comes up with interesting, highly differentiated and widely usable results.

Yes, I recommend the thesis for defense without substantial changes. And I do not propose to change or enhance anything for the thesis to be defensible at the regular defense. Instead I am proposing some questions to be dealt with in the defenseto clarify some issues:

I would like to thank Asgar Belke for his kind words, valuable comments and suggestions.

1. I assume that the PhD committee has clarified the significance of the candidate's contribution relative to his co-authors of the respective papers.

   While I cannot put the exact number of my contribution, let me state that for all of the chapters in my dissertation, the collection of data, computations and first drafts of all co-authored papers were done by me.

   Clarifying questions:

2. Wouldn't it make sense to differentiate between empirical results published in academic journals and those published in the grey literature (where the significance of the results probably does not play the same important role as in case of publications in academic journals)?

   Yes, this is a very good point. Nonetheless, including working papers and mimeographs in meta-analysis might not help alleviate publication bias: if journals systematically prefer certain results, rational authors will already adopt the same preference in the earlier stages of research as they prepare for journal submission. Limited research on this matter exists - a meta-analysis of 87 meta-analyses by Doucouliagos and Stanley (2012) - and finds no difference in the publication bias between published and unpublished studies. Furthermore, our own robustness check with gathering data from working papers on monetary transmission mechanism confirms this result. On the other hand, in the cases where the sample of published studies is too small to conduct MRA, inclusion of working papers might be the the only way to proceed.

3. Do academic studies on the transmission lags of monetary policy not take into account economic policy uncertainty?

   VAR studies on monetary transmission do not really take into account the economic policy uncertainty. The project http://www.policyuncertainty.com seems promising, but is relatively new and so far available only for the handful of countries. Because these data were not available to authors of primary studies, they did not took it into account. Also the estimation of large vector autoregressions has become feasible only recently, so degrees of freedom considerations played a role too when deciding whether to include additional variables into VAR models. Moreover, most of the uncertainty series them seem rather volatile and at high frequencies, therefore aggregating to frequencies used typically in the VARs (quarterly and monthly) would need to be addressed.

4. Under habit formation an agent's consumption exhibits a form of hysteresis, in that his current consumption depend on his past consumption. The academic

literature addresses, among others, habits and hysteresis in labour supply and habit formation and labor migration. Please explain the hysteresis phenomenon in this context. Why do you not deal with it explicitly in this chapter?

Primary studies estimating habit formation in consumption do not generally cite the hysteresis literature. But indeed one can look at the habit formation through the lens of hysteresis theory. I would like to thank for this observation. As a result, in the chapter on habit formation reference is now made to the relevant literature related to hysteresis.

5. Do you think that election cycles in GDP data revisions are possible and probable, e.g. with an eye on the excessive deficit procedure in Europe? What may be the evidence for the Czech Republic?

   Motivation for governments to influence the numbers might indeed be present. This notwithstanding, one would expect that greater pressure would be on the fiscal variables than on the GDP data. Evidence for the Czech republic suggests that revisions to GDP are well behaved. Unfortunately, the short sample covering only nine years does not lend itself to study election cycles, but is indeed good avenue for the future research.

A.3 Prof. Geoff Pugh, M.Sc., Ph.D.

... I recommend the thesis for defense withoutsubstantial changes. However, while not wanting to make unnecessary work for a very productive economist, I would like to hear his thoughts on some issues arising from the papers considered as a whole. I elaborate below. ...
I would like to thank Geoff Pugh for his kind words, valuable comments and suggestions.

1. In the past, MRA has been most unwelcomed in some quarters in economics. Editors and authors did not like the exposure of publication bias (taking this almost as a personal affront) and its corollary that the typical estimates reported

in the literatures under consideration might, to a greater or lesser extent, be misleading. In the context of economics, a rather conservative discipline, MRA was somewhat radical. In contrast, the MRA practice presented in this thesis seems rather sanitized (e.g., pp.3 and 15): theory is used to identify ”misspecifications”, which enter meta-regression models as moderators; then the MRA is used to estimate an ”average effect” corrected for misspecification - i.e. in line with theory. QED! This seems to be almost tautological - using MRA to validate prior views. By estimating average effects ”conditional on best practice” is it possible that you are also filtering out dissident voices; alternative perspectives? For example, by choosing the variables, the price puzzle disappears (pp.32-33). Are there alternative plausible specifications? What is the whole literature telling us?

Yes, admittedly, by performing best practice, we are filtering out part of the literature, but at the same time, this allows us to perform synthesis as opposed to only declaring that the literature is heterogeneous and assigning drivers to this heterogeneity. Of course, in many cases it is not clear what method or aspect of the study design is superior to the others, in such cases this should be stated and best practice should be presented with proper sensitivity analysis. For example, we do so in the first chapter (page 37) Nevertheless, there are also cases when one can say with reasonable certainty that some methods or design aspects are flawed, and eventually one would like to make use of this information. Performing MRA without stating what key messages can be taken from the literature could be considered unsatisfactory. By estimating average effect conditional on best practice we are attempting to figure out what the whole literature is implying.

I am doubtful of claims to filter out the effects of misspecification to ”create a synthetic study with ideal parameters . . . when best practice is followed” (pp.8-9). This seems to move MRA from critique to affirmation. If we know what best practice is, and theory informs us about the ”ideal parameters”,then why bother with MRA? It looks as if the answer that is wanted - the one consistent with established theory - is used to define best practice; and contradictory evidence is waived away as ”misspecification”.

As mentioned in the previous point, the best practice is arguably subjective. But it is more robust than relying on estimates from a single study, in a sense that it uses more information (e.g. by combining different levels of granularity of data i.e. micro and macro data). Furthermore, sometimes not all the combinations of the best practice can be performed in one study. Generally, one could argue that theory provides us with ideal intervals or direction of effect and thanks to MRA, we can obtain ideal parameters.

One specific point in this regard. Among the solutions to the ”price puzzle”, I was surprised to see that more attention is not paid to the ”cost channel”, even though this is described (p.22). For industrial economists, this might even be the prime suspect: there are major firms whose Boards of Directors routinely consider interest rates as an agenda item.

Yes we acknowledge that we did not devote much space to the cost channel explanation of the price puzzle. On the other hand, by trying to be concise we refrained from providing detailed descriptions of other potential explanations and provided an extensive references instead. In that sense, the current discussion of cost channel could be seen as proportional to other explanations in the chapter.

2. The author seems to have an exaggerated respect for the ”top journals”. Arguably, these are even more prone to publication bias (and other forms of bias?) given the extreme incentives to publish in these journals. This is indicated by at least two references in the author’s own work (the referencing is indeed most thorough): Necker (2014); and Brodeur et al. (2016). Moreover, the lack of systematic differences reported between journals of different standing or even between published and unpublished studies (p.10) would suggest a more enquiring and critical approach to assumptions about ”quality”. How do

we control for quality in selecting either the population or the sample? This is still unresolved in MRA (to my knowledge). Some recognition of this and discussion of the options would be useful. For example, on (p.11 we read: "When we were uncertain about the magnitude of shock used in the primary study, we contacted the authors." On the one hand, I am impressed by the author's thoroughness. On the other, I wonder whether this kind of information - informative about the often shockingly bad reporting practices in economics - may be more informative about the "quality" of research than where it is published.

It was by no means our aim to disregard other than to journals. It is true that, we do not consider working papers, but on the other hand we add to the sample also papers from low quality journals. In several cases we provide statistics based on a subsample of estimates from the top journals for motivation, but the main results are always based on the whole population of published studies (i.e. not only those from top journals).

3. I would like to see more recognition and some discussion of other unresolved issues in MRA

   The discussion of selected unresolved issues is now a part of the introduction (paragraphs 3-8 on pages 1-4).

   a) No universally preferred approach to estimating MRA models. I would like more discussion of alternative approaches to MRA models.

   Variable selection is a problem that is not unique in MRA models. It is an issue which every empirical study which does not use experimental data must face. Obviously, the variables used in regressions should be based on theory, but meta-analysis regression is special in that it is sometimes hard to motivate by theory and to discriminate between variables (e.g. publication characteristics). Ideally one would like to control for many factors, but to achieve efficient estimation one needs to select variables among many potential ones. There are many approaches to solve this ranging from general-to-specific to model averaging. Our preference is using Bayesian model averaging. Of course, the choice of the *exact method is case dependent. The discussion of alternative approaches is now part of the introduction.*

b) I would like to see a similarly critical attitude towards BMA. BMA seems to have no concerns in averaging across models that may or may not be satisfactory. Is fit the only consideration?

Of course BMA is not without its issues, but we believe that the advantages of using model averaging overweight the disadvantages. We now discuss issues that arise when one uses BMA in the introduction.

c) BMA and judgement i.e. BMA and best practice

As regards to the relationship between model averaging and best practice, it is crucial to obtain most robust estimates given data, this in turn allows for reliable estimates of the best practice. In principle, one could consider averaging the best practice estimate by assuming all sensible combinations. However, while in meta-regression we often do not have string prior about what might systematically drive variation in estimates or have many different options, in best practice there often is an indication from the literature about the appropriateness of using certain data or methods. In any case, this could be an interesting avenue for future research.

d) Should publication variables be considered as ”K” moderators (i.e. as intercept shift terms, affecting publication bias but no - directly - representative effect)

Ideally, yes. But more often than not it is not easy to draw a clear line between regressors affecting only bias and not the representative effect a priori. Further issues like multicollinearity, problems with sources of variation, complicate such distinction.

e) Justification for pooling econometric estimates with values outputted from DSGE models. Can the latter legitimately be regarded as estimates?

We only take estimates from estimated models, i.e. calibrated models are dis-

carded. We agree that in case of Bayesian DSGE models, greater care should be taken. It is well known that in case when likelihood is not very informative, the prior information dominates, thus leading to identification problems.