The Econometrics Journal

Econometrics Journal (2015), volume **18**, pp. C22–C41.
doi: 10.1111/ectj.12038

Economic theory and forecasting: lessons from the literature

RAFFAELLA GIACOMINI[†,‡]

[†]*Department of Economics, University College London, 30 Gordon Street, London WC1H 0AX, UK.*

[‡]*Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK.*
E-mail: r.giacomini@ucl.ac.uk

First version received: February 2014; final version accepted: October 2014

Summary Does economic theory help in forecasting key macroeconomic variables? This article aims to provide some insight into the question by drawing lessons from the literature. The definition of 'economic theory' includes a broad range of examples, such as accounting identities, disaggregation and spatial restrictions when forecasting aggregate variables, cointegration and forecasting with dynamic stochastic general equilibrium (DSGE) models. We group the lessons into three themes. For the first, we discuss the importance of using the correct econometric tools when answering the question. For the second, we present examples of theory-based forecasting that have not proven useful, such as theory-driven variable selection and some popular DSGE models. For the third set of lessons, we discuss types of theoretical restrictions that have shown some usefulness in forecasting, such as accounting identities, disaggregation and spatial restrictions, and cointegrating relationships. We conclude by suggesting that economic theory might help in overcoming the widespread instability that affects the forecasting performance of econometric models by guiding the search for stable relationships that could be usefully exploited for forecasting.

Keywords: *Aggregation, DSGE models, Instability, Out-of-sample forecasting, Parameter restrictions.*

1. INTRODUCTION

Does economic theory help one make more accurate forecasts than those based on atheoretical econometric models? Should one forecast with models derived from first principles or focus on econometric methods that deliver accurate forecasts in practice but do not 'tell a story'? Perhaps one should adopt hybrid approaches that combine elements of both? These are fundamental questions that have inspired some prominent debates in the profession and have affected the practice of forecasting in both academic and policy circles for many years. An early example is the discussion in the 1980s surrounding the collapse of the Cowles commission theoretical models that were popular at the time, which was partly motivated by the inability of these models to forecast the stagflation of the late 1970s. After a period in which the profession focused on reduced-form models in forecasting, a recent trend in the literature has been a return to theoretical models such as the large-scale estimated dynamic stochastic general equilibrium (DSGE) models – see, e.g. Smets and Wouters (2007) – which have now become part of the suite of forecasting models at many central banks and policy institutions around the world. Partly in response to the inability of these first-generation estimated DSGE models to predict the 2007 crisis, we can observe a tendency in the literature towards increasingly richer and larger models, which account for features that were ignored by the earlier models, such as financial frictions and a foreign sector. This trend could raise a concern that history might be repeating itself and, at the very least, makes this a good time to look back and attempt to take stock of what we have learned regarding whether our knowledge about economic theory makes us better forecasters than a computer churning data or a monkey flipping a coin.

For understandable reasons, this being a vast and controversial topic, the review is far from comprehensive and it has a strong focus on econometric methodology. I also discuss some empirical findings but my statements should be viewed as a summary of a few key contributions, rather than a serious meta-analysis of the literature.

There are some natural questions that arise for the reader at this point. What do you mean by 'theory'? Which type of forecasting are you talking about? Surely there are many ways to incorporate theory into forecasting?

First, a rigorous discussion of which type of economic theory one should expect to be useful for forecasting is, on the one hand, challenged by the fact that often what we call 'theory' is in fact based on empirically observed facts. For instance, the introduction of 'frictions' or 'rigidities' into economic models could, at least in part, be ascribed to the need for the model to replicate the persistence observed in time series data. Because a correct specification of the dynamics of a variable is tantamount to accurate forecasting, any mechanism that generates persistence has the potential to improve the forecasting properties of the model, regardless of the plausibility of its theoretical foundations. On the other hand, there are modelling choices that have nothing to do with empirical fit, but are just convenient mathematical representations that facilitate closed-form solutions to the model. For example, it is not pre-ordained that there should be a level, slope and curvature factor in the term structure of interest rates, let alone that we should work with affine factor dynamics. In this paper, I instead use the term 'theory' rather loosely and discuss a broad range of examples, from the use of simple national accounting identities to estimation of full-fledged DSGE models. Intermediate examples include imposing spatial restrictions when forecasting variables that are aggregated over regions; using partial equilibrium restrictions such as Taylor rules for inflation, Purchasing Power Parity (PPP) restrictions for exchange rates and Euler equations; cointegrating restrictions and no-arbitrage restrictions in dynamic term structure models.

Second, I restrict attention to the problem of forecasting key macroeconomic and financial variables such as measures of real activity, inflation, interest rates and exchange rates using historical data at medium and low frequencies (monthly, quarterly, annually). I do not consider financial forecasting using high-frequency data.

Third, I discuss various ways to incorporate theory into forecasting. These include letting theory guide the choice of predictors in the context of univariate models; imposing theoretical restrictions on reduced-form models; using prior distributions based on a theoretical model; combining theoretical and reduced-form models; forecasting with estimated DSGE models.

I structure the review as a series of lessons drawn from the literature, discuss a few questions that I believe would benefit from further investigation, and conclude by suggesting some new directions in which economic theory might benefit forecasting. The lessons from the literature fall into three main themes.

The first theme concerns the importance of using the appropriate econometric methodology when assessing the usefulness of economic theory for forecasting. One, perhaps obvious, lesson here is that forecasting is a decision problem and therefore any discussion about the usefulness of economic theory should take into account the uses to which the forecast is made and what the forecaster's loss function is. For example, it would be interesting to understand whether one of the touted advantages of forecasting with DSGE models, their ability to 'tell a story', can be formalized within a decision-theoretic framework. A second lesson is that theoretical restrictions do not have to be correct in order to be useful for forecasting, or, conversely, that even valid restrictions may not deliver forecast accuracy improvements. The correct tool for assessing the usefulness of economic restrictions is thus one that moves away from hypothesis testing and towards a framework that can capture these trade-offs. One key technical device for accomplishing this is the use of an asymptotic framework with non-vanishing estimation uncertainty, which has been a central theme of some of my research contributions. Finally, the notion of usefulness is perhaps best expressed in relative terms: compared to what is theory useful? This relative focus makes the conclusion sensitive to a number of user-defined choices: the choice of the benchmark; the size of the model; the choice of the evaluation sample (because the relative performance might be time-varying); whether and how to pre-filter the data (because these might differentially affect the theoretical and the benchmark model). This points to the importance of conducting sensitivity analysis when answering the question.

The second set of lessons is that some popular examples of theory-based forecasting have not proven successful, at least not incontrovertibly. They include the use of theory-driven variable selection in univariate forecasting and multivariate forecasting based on estimated DSGE models such as Smets and Wouters (2007). With regards to the latter, the picture that emerges from the literature is that DSGE models are generally outperformed by simple econometric models and by survey forecasts, particularly when one takes into account the methodological issues highlighted above and performs some sensitivity analysis. One possible reason for this finding is that a forecast based on the model of Smets and Wouters (2007) embeds a number of restrictions and assumptions that can potentially affect its accuracy, such as restrictions on the cross-sectional dependence among the variables, restrictions on the dynamics, shrinkage implicitly imposed through the choice of priors and treatment of possible trends and cointegrating relationships. More research is needed to understand the relative contribution of each of these restrictions to the forecast performance of DSGE models.

For the third set of lessons, we consider what the literature has shown to be useful in forecasting and ask the following two questions. Does economic theory have anything to do with it? Can this inform future theoretical modelling efforts?

We first argue that different forces might be at play at short and long forecast horizons. At short forecast horizons, what has proven useful is the ability to extract information from 'big data' in a way that avoids the curse of dimensionality. This includes the use of factor models, Bayesian shrinkage, model combination and survey expectations, which could be viewed as a summary measure of the information available to survey participants at the time of forecasting. At long forecast horizons, there is some evidence that what can help is a careful modelling of trends as well as the imposition of cointegrating restrictions. Save for cointegration, which could be viewed as a type of theory-based restriction based on the notion of long-run equilibrium, none of the other features seems to be grounded in economic theory. However, these findings have already started to affect theoretical modelling efforts, and a number of 'hybrid' DSGE models have appeared in the literature, which incorporate external information such as factors extracted from large data sets or survey forecasts. Save for the criticism about the dubious 'structural' foundations or such approaches, this merging of economic theory and econometric practice might be a useful direction in which to take these models. At the very least, it would be useful to assess the empirical performance of these hybrid methods in a systematic way.

We then suggest that it might be useful to separate what economic theory implies for the cross-sectional dependence among variables from what it says about their dynamics. Although the literature has shown that theory might suggest useful restrictions that affect the cross-sectional dependence (such as the use of accounting identities or, more in general, spatial or sectoral disaggregation, possibly disciplined by spatial dependence restrictions), the jury is still out on whether economic theory has anything more useful to say about the dynamic behaviour of variables than reduced-form modelling approaches. This is not to say that it should not. In fact, the final suggestion of the paper is that economic theory might help in the fight against one of the biggest enemies of forecasting: structural instability. I suggest that this battle might be fought on two fronts. The first involves a serious investigation of the propagation mechanism embedded in the current generation of general equilibrium models. This echoes the call in a series of recent papers by Faust (2012) and Faust and Gupta (2012) for an assessment of the plausibility of these propagation mechanisms for reality, to which I would add an investigation of their stability over time. The second front is to use economic theory as a guide in finding structural relationships that are, by their very definition, stable over time. These could be simple parameter restrictions or moment restrictions involving future observables, such as Euler equations. I end this review by discussing how Bayesian methods and exponential tilting can be used to incorporate these two classes of restrictions into forecasting models. Thus, they can be helpful tools for any forecaster interested in finding a more optimistic answer to the question in the first line of this introduction than the one that emerges from the rest of the review.

2. LESSON 1: ECONOMETRIC METHODOLOGY MATTERS

This section highlights the importance of utilizing the appropriate econometric methodology when assessing the usefulness of economic theory for forecasting.

2.1. The decision problem and the choice of loss function

The first lesson from the literature is the, perhaps obvious, point that forecasting is a decision problem and therefore the usefulness of economic theory depends on the underlying preferences and the constraints of the decision maker. The typical framework for forecast evaluation, for example, considers a variable of interest Y_{t+h} and the h-step-ahead forecast based on the information available at time t, f_t, and evaluates the accuracy of f_t by the expected loss $E[L(Y_{t+h}, f_t)]$.

Granger and Machina (2006) show that decision-based loss functions should obey certain properties and that the commonly used quadratic loss has restrictive and unrealistic implications for the underlying decision problem. In spite of this, the forecasting literature has for the largest part continued to focus on the quadratic loss.

Carriero and Giacomini (2011) show that the choice of loss function matters when investigating the usefulness of no-arbitrage restrictions for forecasting the term structure of interest rates. They measure the usefulness of the theoretical restrictions as the optimal weight λ in a combination of the forecast subject to theoretical restrictions f_t^{R} and the unrestricted forecast f_t^{U} and construct an out-of-sample test of the null hypothesis that the restrictions are not useful. Optimality is defined with respect to a general loss function. Their empirical results show that the choice of a standard quadratic loss leads to the conclusion that the theoretical restrictions are no more useful than atheoretical restrictions, which similarly reduce the dimensionality of the system, such as the random walk assumption. The no-arbitrage restrictions are instead useful when one considers an economically meaningful loss, which is related to the profit realized by investing in a bond portfolio with weights based on the theoretical model. One take-home lesson is that assessing the usefulness of parameter restrictions in large-dimensional systems (such as no-arbitrage term structure models) using a quadratic loss may put too much emphasis on the variance reduction accomplished by imposing any type of restriction. It also raises the question whether economic theory offers a superior opportunity for dimension reduction than other popular methods that are not necessarily based on theory.

The choice of an appropriate loss function is less straightforward when moving away from portfolio allocation decisions and considering, for example, the decision problem of a forecaster at a central bank. One issue to consider in this context is that, typically, central banks use the same model for both policy analysis and forecasting, which introduces a trade-off between the model's forecasting performance and its theoretical coherence (Pagan, 2003). This trade-off is echoed in the literature investigating the forecasting performance of DSGE models, which are often praised for their ability to 'tell a story' (Edge et al., 2008), in spite of being, in most instances, outperformed by survey forecasts (Del Negro and Schorfheide, 2012) or econometric models (Gürkaynak et al., 2013). A natural question to ask is: can this trade-off be formalized in a decision-theoretic context? I am not aware of any contribution in the literature that has attempted to do so. Without such a formalization one is thus left with several open questions, and two in particular merit further discussion. The first question is whether a better understanding can be gained about the nature of the 'story' that central bankers want to hear, and whether parts of it can be told without resorting to full-fledged DSGE models. For example, it might be possible to tell a story behind a forecast using simple partial-equilibrium restrictions (e.g. the Phillips curve) or disaggregation arguments (e.g. when a forecast of an aggregate variable can be justified by particular developments in a region or sector of the economy). If the 'story' is the transmission mechanism specified by a DSGE model, a second natural question that emerges is: what if competing mechanisms result in similar forecasting performance? In other words, how should one choose among different theoretical models, knowing that they are just approximations of the truth? A related discussion (in the different context of policy-making) is contained in a series of thought-provoking articles by Faust (2012) and Faust and Gupta (2012). Faust advocates following other disciplines, such as toxicology, in developing a framework for integrating econometric evidence (which does not say much about transmission mechanisms), general equilibrium models (which specify transmission mechanisms but have important omissions and coarse approximations) and – what he argues is currently lacking in the literature – a set of tools for assessing the relevance of the transmission mechanism for reality. An answer to Faust's call would undoubtedly also increase our understanding of the usefulness of DSGE models for forecasting.

2.2. *Relative versus absolute evaluation*

The likely misspecification of any theoretical model means that absolute evaluation of forecast performance (e.g. assessing the optimality of the forecast for a given loss function) is not likely to be informative and that the question would be best answered by using relative evaluation methods, which compare the performance of a theory-based forecast to that of a benchmark. Naturally, the choice of benchmark will greatly affect the conclusions of the exercise. Consider, for example, the case of a linearized DSGE model, which, under general conditions (see, e.g., the review by Giacomini, 2013), can be represented as a state-space model:

$$\begin{aligned} X_t &= \widetilde{A} X_{t-1} + \widetilde{B} \varepsilon_t \\ Y_t &= \widetilde{C} X_{t-1} + \widetilde{D} \varepsilon_t . \end{aligned} \tag{2.1}$$

Here, X_t represents the state variables of the model, Y_t the observable variables and ε_t the structural shocks. The matrices $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$ and $\widetilde{D}$ are subject to cross-equation restrictions implied by the DSGE model. Let $\theta = (\text{vec}(A)', \text{vec}(B)', \text{vec}(C)', \text{vec}(D)')$ denote the unrestricted parameters of (2.1) and let $\widetilde{\theta}$ indicate an estimator of θ subject to the DSGE restrictions. If the forecasting performance is evaluated using a quadratic loss, then the accuracy of a forecast will depend on the accuracy of the estimator on which it depends and it is possible that in finite samples the bias–variance trade-off is such that a benchmark forecast based on an alternative restricted estimator $\widehat{\theta}$ could yield more accurate forecasts than those based on $\widetilde{\theta}$, even if the model (2.1) were correctly specified. Comparing the accuracy of the restricted estimator $\widetilde{\theta}$ to that of an unrestricted estimator of θ will also be sensitive to the dimension of θ relative to the sample size, as in large-dimensional systems it is possible that any kind of 'shrinkage' will result in superior forecasting ability, regardless of its theoretical foundations – as Carriero and Giacomini (2011) show for the case of no-arbitrage affine term structure models.

2.3. The right evaluation framework

From the perspective of a forecaster interested in using economic theory for real-time forecasting, the right question to ask is not whether the theoretical restrictions are valid, but, rather, whether imposing them can result in accurate forecasts, for a particular loss function. Moreover, if we accept the arguments in Section 2.2, the question is best framed as a relative rather than absolute evaluation problem. This means that hypothesis testing, whether in-sample or out-of-sample, would not necessarily deliver the right answer; in fact, as illustrated by Inoue and Kilian (2005) and Hansen and Timmermann (2013), out-of-sample hypothesis tests of the type considered by West (1996) and Clark and McCracken (2001) could result in substantial power loss relative to their in-sample counterparts. In this section, we instead illustrate the evaluation framework of Giacomini and White (2006), which considers out-of-sample relative evaluation tests that can capture the possibility that valid theoretical restrictions may not be useful for forecasting, or, conversely, that theoretical restrictions that are only approximately valid could deliver forecast accuracy gains. The key difference between the testing framework of Giacomini and White (2006) and that of West (1996) and Clark and McCracken (2001) is that in the former the estimation uncertainty contained in the forecasts is preserved asymptotically, whereas in the latter the estimation uncertainty disappears as the sample size grows. In practical terms, this means that the hypothesis considered by Giacomini and White (2006) is formulated in terms of in-sample parameter estimates, whereas the hypothesis of West (1996) depends on population parameters. A test for the usefulness of theoretical restrictions in the out-of-sample framework of Giacomini and White (2006) involves the following steps. At time t, one considers two h-step-ahead forecasts for the vector of observables Y_{t+h}: one that imposes the theoretical restrictions,

$f^{\text{theory}}_{t,h}(\widetilde{\theta})$ and one based on a benchmark model, $f^{\text{benchmark}}_{t,h}(\widehat{\theta})$. The benchmark forecast could, for example, be based on the unrestricted model (i.e. a VAR for Y_t) or be obtained by imposing a competing set of restrictions (i.e. random walk restrictions). The out-of-sample procedure involves splitting the sample into an in-sample portion of size m and an out-of-sample portion of size $n \equiv T - h - m + 1$. The forecasts are made for time periods $t = m, \ldots, T$ and depend on parameters estimated using in-sample data. The sample split m is arbitrarily chosen by the user. The two main 'out-of-sample schemes' one typically considers are the 'recursive' scheme, where the in-sample data at time t include observations indexed $1, \ldots, t$, and the 'rolling window' scheme, where the models are estimated using the most recent m data points, so that the in-sample data at time t include observations indexed $t - m + 1, \ldots, t$. The two sequences of competing forecasts are thus $\{f^{\text{theory}}_{t,h,m}(\widetilde{\theta}_m)\}_{t=m}^{T-h}$ and $\{f^{\text{benchmark}}_{t,h,m}(\widehat{\theta}_m)\}_{t=m}^{T-h}$, where the dependence of the parameters on the in-sample size m is made explicit. The goal is to compare the performance of the two forecasts by testing the hypothesis that they yield the same out-of-sample loss in expectation. This is typically carried out separately for each component of Y_t. For a quadratic loss, for example, one would test the hypothesis

$$H_0 : E[\Delta L_{t+h}] \equiv E[(Y_{i,t+h} - f^{\text{theory}}_{t,h,m}(\widetilde{\theta}_m))^2 - (Y_{i,t+h} - f^{\text{benchmark}}_{t,h,m}(\widehat{\theta}_m))^2] = 0,$$

using a simple t-test, which depends on a heteroscedasticity- and autocorrelation-consistent (HAC) estimator of the asymptotic variance of ΔL_{t+h} (typically with truncation lag $h - 1$). The key assumption that we introduce to preserve the parameter estimation uncertainty asymptotically is that the in-sample size m is finite, which is compatible with a rolling window scheme but rules out the adoption of a recursive scheme.

2.4. The need for sensitivity analysis

The discussion in the previous subsections makes it clear that there are a number of arbitrary choices that one needs to make when assessing the usefulness of theoretical restrictions for forecasting. Take as an example the state-space model in (2.1), which is the form in which one can express the linearized model of Smets and Wouters (2007). Because one of the reasons for the popularity of this model is the finding by Smets and Wouters (2007) that a version of the model estimated using Bayesian methods has comparable forecasting performance to that of atheoretical benchmarks, it seems important to assess the sensitivity of this conclusion to the choices that underlie the analysis. There are several such choices. The first is to decide which variables are observable and thus enter the vector Y_t. This choice is arbitrary because the model only determines what the state variables X_t are, and it is up to the econometrician to choose the elements of Y_t in order to be able to write a likelihood. Second, one typically needs to decide how to deal with stochastic singularity (which occurs when there are more observables than shocks), whether by increasing the number of shocks, eliminating variables from Y_t or adding measurement error to the equation for Y_t. Third, the choice of benchmark will affect the outcome of the comparison. An unrestricted VAR probably suffers from the curse of dimensionality, a random walk imposes too tight restrictions and a Bayesian VAR (with the Minnesota prior) falls somewhere in the middle. Still, it is unclear why one should focus on multivariate models for Y_t and not adopt more flexible modelling approaches for each component, particularly considering that the forecast accuracy is usually assessed separately for each element of Y_t. This point is clearly illustrated by the results in Gürkaynak et al. (2013), who show how the relative accuracy of the DSGE model of Smets and Wouters (2007) and econometric models depends on the choice of benchmark. They further question the use of a BVAR as a natural benchmark in this context, and suggest that a more appropriate benchmark is an unrestricted VAR for output, inflation and interest rates. Fourth, the choice of priors and their hyperparameters for both the DSGE and the BVAR will affect the relative accuracy of the forecasts, unless the forecasts are drawn from the posterior predictive distributions; see Del Negro and Schorfheide (2012) for a review of how to do this for the DSGE model. Fifth, the data pre-processing and, in particular, the detrending method can have dramatic effects on the comparison between a DSGE model relative to a BVAR, as suggested by Sims (2003) and empirically documented (in the context of in-sample fit) by Giacomini and Rossi (2015). Finally, the choice of the sample matters, as the relative performance of the models could be varying over time. This is shown by Giacomini and Rossi (2009, 2010, 2015), who propose tests for assessing the time-varying relative performance of two models in terms of in-sample fit or out-of-sample forecast accuracy, respectively. This work shows that standard out-of-sample comparisons based on comparing average performance over a particular out-of-sample period could give misleading conclusions when the relative performance of two models is unstable; for example, indicating that two models performed equally well when in reality there was a reversal in relative performance at some point during the sample, which led to the relative performances cancelling each other out. The application of their methods to the comparison between the DSGE model of Smets and Wouters (2007) and a BVAR shows that indeed the relative performance of the two models has varied considerably over time, with the BVAR outperforming the DSGE model over the last part of the sample they considered.

3. LESSON 2: WHAT DOES NOT WORK

In this section, I argue that there is not sufficient evidence in the literature about the usefulness of two types of theory-based forecasts: those based on estimated DSGE models, such as the popular model of Smets and Wouters (2007), and those based on univariate models with theory-driven variable selection.

3.1. DSGE models

The number of arbitrary choices one makes in the econometric methodology when assessing the forecasting performance of DSGE models, as well as the possibly conflicting conclusions that different choices may lead to, are reflected in the mixed evidence presented by the literature investigating the forecasting performance of DSGE models empirically. This literature has been growing at a fast pace since the publication of the influential article by Smets and Wouters (2007). Recent contributions include Adolfson et al. (2007), Edge and Gürkaynak (2010), Edge et al. (2010), and the literature reviews by Del Negro and Schorfheide (2012) and Gürkaynak et al. (2013). The various articles consider different versions of DSGE models, different data, i.e. European for Adolfson et al. (2007), versus US data for all others, revised data versus real-time data, the latter considered by Edge and Gürkaynak (2010) and Gürkaynak et al. (2013), and different data periods, for example, Gürkaynak et al. (2013) showed how the forecasting performance varied over time. The two recent reviews by Del Negro and Schorfheide (2012) and Gürkaynak et al. (2013), in particular, nicely complement each other because they both consider the model of Smets and Wouter (2007) but the former mainly focuses on the comparison with survey forecasts and the latter on the comparison with benchmark econometric models. The survey by Gürkaynak et al. (2013), in particular, is remarkable in its addressing many of the concerns about the effects of the econometric methodology on the results of the exercise that I have discussed in Lesson 1. The main conclusion from both studies is that econometric models and surveys are generally more accurate than the DSGE model for forecasting inflation and interest rates at all horizons and for output growth at short horizons. The DSGE model has only a marginal advantage over both surveys and econometric models when forecasting output growth at long horizons (about two years). Moreover, the performance of the DSGE has deteriorated in the years after the publication of the article (i.e. the true out-of-sample period), perhaps prompting a concern that some of the modelling choices behind the DSGE model of Smets and Wouters (2007), such as the choice of priors, reflected the information contained in the data available at the time the article was written. The empirical literature thus paints a somewhat negative picture about the usefulness of DSGE models as forecasting tools, particularly given their poor performance in recent years. Moreover, even in the only case in which the DSGE model outperforms surveys and econometric models – forecasting output growth at long horizons – it is possible that the finding is sensitive to the choice of benchmark. Regarding the comparison with surveys, it is unsurprising that a model outperforms surveys at long horizons, as it is well known that survey forecasts tend to be accurate for short-horizon forecasting but do poorly at long horizons; Deschamps and Ioannidis (2013) argue that this is consistent with strategic behaviour by forecasters who want to appear well informed before the predicted variable is observed. Regarding the comparison with econometric models, the finding by Gürkaynak et al. (2013) that a DSGE model outperforms simple benchmarks such as univariate AR models and unrestricted VAR models for long-horizon output growth forecasting could be linked to the original comment by Sims (2003) to Smets and Wouters (2007) that the way the data are detrended might penalize the econometric model in favour of the DSGE model. It is thus possible that the relative performance of the DSGE model and econometric models could change if one were to choose a benchmark that takes into account possible common trends in the data, or if one were to use a different detrending method.

3.1.1. Hybrid DSGE models. Partly in response to the disappointing empirical performance of DSGE models in recent years, a small body of literature has emerged that tries to adopt more flexible, 'semi-structural' approaches to modelling using theoretical restrictions. Del Negro and Schorfheide (2004) propose a Bayesian approach to relaxing the cross-equation restrictions imposed by the DSGE model, which results in a 'hybrid' model that can be viewed as optimally combining the DSGE model and its unrestricted, reduced-form counterpart. Del Negro et al. (2007) find some evidence that the hybrid approach outperforms either a dogmatic DSGE model or an unrestricted econometric model in terms of forecasting performance over the period 1985–2000. Caldara et al. (2012) formalize what they argue is already a practice at central banks of extending DSGE models to include missing channels, such as financial or housing sectors. They propose augmenting the DSGE model using proxies for the missing channels that are based on auxiliary econometric models. The issue of the sensitivity of the performance of DSGE models to the modelling of trends has also been documented by Canova (2009) and Canova and Ferroni (2011). Moretti and Nicoletti (2010) propose a flexible method for estimating DSGE models with unknown persistence and Pesaran and Smith (2011) suggest using theory to guide the specification of long-run restrictions while leaving the short-run dynamics unrestricted. To my knowledge, the forecasting performance of these flexible approaches has not yet been systematically evaluated.

3.2. Theory-driven variable selection

At the opposite end of the spectrum from DSGE models, one could consider simple, univariate models in which economic theory guides the choice of predictors. Examples are Phillips curve models for forecasting inflation and models imposing PPP conditions for forecasting exchange rates. A review of the vast literature on the topic of variable selection in forecasting models is beyond the scope of this paper, so I just focus here on the cases of inflation and exchange rate forecasting. Faust and Wright (2013) present a comprehensive review of the literature on inflation forecasting and make a convincing argument that survey forecasts are the most accurate predictors of inflation at all horizons and consistently outperform model-based forecasts, including those based on economic theory. Groen et al. (2009) reach a similar conclusion for the UK. In the context of exchange rates forecasting, it has been known, at least since Meese and Rogoff (1983), that the random walk is a benchmark that is hard to beat. Rossi (2013a) offers a recent survey of the literature and seeks to answer the question: does anything forecast exchange rates and, if so, which variables? Her conclusions are that the random walk is still a powerful benchmark and that whether models based on economic theory beat the random walk depends on the forecast horizon, sample period and forecast evaluation method. A general consensus that seems to have emerged is that monetary and PPP fundamentals have no predictive ability, whereas Taylor rules and net foreign assets models have some forecasting ability at short horizons, but the relationships are characterized by widespread instabilities. One of the referees of this paper makes the interesting point that the fact that it is difficult to beat the random walk is perhaps less surprising once we take into account the variability in exchange rates vis-à-vis interest rate spreads, and suggests that perhaps we should view the good performance of the random walk model as a triumph for efficient markets theory rather than a failure of theory.

4. LESSON 3: WHAT WORKS

We now turn our attention to what the literature has shown to be successful at forecasting. The goal is to understand if, on the one hand, economic theory has anything to do with it, and if, on the other hand, the insights gained can be used to inform new theoretical modelling efforts.

4.1. Separating short-term and long-term forecasting

The literature has shown that what works at short-horizon forecasting does not necessarily help forecast at long horizons. The conclusions usually differ for different variables (e.g. real versus nominal), but there are some general tendencies. For example, there is evidence that econometric models are quite successful at forecasting real variables for the current quarter (nowcasting) and for short forecast horizons. One of the key issues at short horizons appears to be the ability to extract information from large data sets in a way that avoids the curse of dimensionality, for example by using dynamic factor models (Stock and Watson, 2006), or Bayesian shrinkage methods (De Mol et al., 2008). Another important issue is the ability to guard against the structural instabilities that are widespread in forecasting performance (e.g. by methods such as intercept corrections or forecast combination); see the review by Rossi (2013b). None of these methods relies on economic theory. When forecasting at long horizons, instead, what seems to matter more is the modelling of trends and the possible imposition of cointegrating relationships, which can be viewed as theoretical restrictions motivated by the existence of an equilibrium among the variables considered. The literature appears to be divided on the usefulness of imposing cointegrating restrictions on a forecasting model; see the survey by Elliott (2006). However, this claim appears to be mostly based on Monte Carlo simulations. Clements and Hendry (1995) find that the usefulness of cointegration depends on whether one is forecasting growth rates or their ratios. I am not aware of the existence of up-to-date comprehensive studies that investigate the question empirically, and it would be useful to gain better understanding about the importance of modelling trends and allowing for cointegration for long-horizon forecasting. For example, going back to the result by Del Negro and Schorfheide (2012) about forecasting with DSGE models, it would be interesting to understand how much of the predictive ability of DSGE models for output growth at long horizons is due to the data pre-processing or the imposition of cointegrating relationships, as opposed to the internal mechanics of the model. Regarding whether these empirical insights could guide new theoretical developments, there is already some evidence that the literature has started moving in this direction. For example, Boivin and Giannoni (2006) propose estimating DSGE models in a way that exploits the information contained in factors extracted from large-dimensional data sets. They show that exploiting this information is important for accurate estimation, in particular with regards to the measurement of inflation.

4.2. Survey forecasts

There is by now ample evidence that survey forecasts or market-based expectations (such as those implied by the prices of traded futures contracts) are among the most accurate predictors of key macroeconomic variables. For inflation, Faust and Wright (2013) show that surveys are a benchmark that is hard to beat at both short and long forecast horizons. For interest rates, Chun (2009) and Altavilla et al. (2013) show that surveys are accurate, but mostly at short horizons. For output growth, D'Agostino and Schnatz (2012) and Del Negro and Schorfheide (2012) show that surveys are reliable for nowcasting and short-horizon forecasting. These conclusions, together with the finding about the effectiveness of data-reduction methods at short horizons discussed in Section 4.1, are consistent with the conjecture that what matters for forecasting (at least at short horizons) is the ability to incorporate into the forecasts any information about the current and future state of the economy in a timely fashion and in a way that effectively deals with large-dimensional data sets. Survey and market participants seem remarkably skilled at doing so. For example, Altavilla et al. (2013) show some evidence that the superior performance of survey forecasts of bond yields can be linked to their ability to capture information about the current state of the real economy as well as forward-looking information, such as that contained in monetary policy announcements. The fact that surveys tend to be outperformed by models at long horizons – see, e.g. Altavilla et al. (2013) and Del Negro and Schorfheide (2012) – is also consistent with the conjecture that survey participants do not necessarily have a deeper knowledge than a hypothetical econometrician of the inter-relationships and dynamic forces that drive the economy, but they are simply better at processing information in real time. An important concern that should be borne in mind when comparing survey forecasts to model-based forecasts is that the point in time when surveys are conducted often does not precisely line up with the timing of the information used in model-generated forecasts. This is, in part, because surveys often are conducted during the month whereas models restrict information sets to information available at the end of a particular month. Unless high-frequency information is used in the model forecasts, it is difficult to compare them with surveys, given the difference in conditioning information. Altavilla et al. (2013) take this concern into serious consideration and use daily yield data to align the information set of the model and that of the survey participants. The usefulness of incorporating survey forecasts into a model has been recognized in the literature, and a number of contributions have shown how doing so results in sizable accuracy gains. For example, Del Negro and Schorfheide (2012) show that the forecasting performance of Smets and Wouters (2007) is improved by incorporating into the model long-run inflation, output and interest rate expectations. Altavilla et al. (2013) use the exponential tilting method described in Section 5.1 to incorporate survey forecasts into a base model. They show that the performance of the dynamic Nelson–Siegel model of Diebold and Li (2006) for the term structure of interest rates is substantially improved by anchoring the yield curve using survey expectations. Moreover, Altavilla et al. (2014) find that market-based expectations implied by federal fund futures can be usefully incorporated into models of bond yield returns in a way that generates profitable investment opportunities in bond markets.

4.3. Separating cross-sectional dependence from dynamics

When it comes to understanding the usefulness of theoretical restrictions for forecasting, it might be helpful to separate the discussion between what theory implies for the cross-sectional dependence among variables and what it says about their dynamic behaviour. This is something that is challenging to accomplish in the context of DSGE models, which imply restrictions for both. Save for the case of trends and cointegrating relationships discussed in previous sections, there is also scant evidence that economic theory offers insights about the dynamics of a time series that can be usefully exploited for forecasting. In this section, I focus instead on the implications that economic theory has for cross-sectional dependence among variables and I discuss two simple forms of theoretical restrictions that have proven somewhat useful: the use of accounting identities (or, more generally, cross-sectional disaggregation) and the imposition of spatial restrictions.

4.3.1. Accounting identities and disaggregation. In a somewhat loose sense, the use of national accounting identities and spatial or sectoral disaggregation could be interpreted as a simple way to exploit theoretical restrictions when forecasting aggregate variables, as it can allow one to capture developments in certain regions or sectors of the economy that affect the forecast of the aggregate variable of interest. The question of whether disaggregation results in improved forecasting ability has been well investigated in the literature. Although the evidence is mixed, the conclusion is that it can. Bermingham and D'Agostino (2014) find disaggregation to be useful for prices in both the US and the Euro area, and Hendry and Hubrich (2006) reach a similar conclusion for the US. Perevalov and Maier (2010) find some improvements from disaggregation when forecasting US GDP. Carson et al. (2011) forecast air travel demand using disaggregated airport-specific data and find benefits from disaggregation as well as from imposing some restrictions on the heterogeneity across airports. The theoretical literature has, for some time, acknowledged the usefulness of disaggregation in modelling and forecasting, and a high level of disaggregation is one of the distinguishing features of the large-scale macroeconomic models (e.g. FRB/US at the Federal Reserve Board) developed by several central banks around the world, including the Federal Reserve Board, the Bank of England, the Swedish Riksbank and the European Central Bank. Even though these models have come under attack – see, e.g. Sims (1980) – they still play a key role in the forecasting and policy analysis carried out by these institutions. In fact, disaggregation is inherently built into the forecasting process, as many central banks have dedicated staff members who are responsible for producing (initial) forecasts for individual sectors or variables. Unfortunately, it seems impossible for an outsider to understand whether this disaggregation is partly responsible for the well-documented accuracy of the forecasts of central banks, such as the Greenbook forecasts (e.g. Romer and Romer, 2000), as the final published forecasts are the result of an iterative procedure that further involves a good dose of judgement. It would be interesting to have access to data that document how the forecast changes through the various iterations, in order to understand the impact of disaggregation, the role of economic theory and the extent of the judgemental corrections. I doubt that such a data set exists and, if it does, that I will be able to obtain it anytime soon. The possible benefits of disaggregation have also been recognized in the context of DSGE modelling, with models such as that of Bouakez et al. (2014), which allow for a large number of heterogeneous sectors and are driven by sector-specific, as well as aggregate shocks. Open-economy DSGE models such as that of Adolfson et al. (2008) could also be viewed in this light. Unfortunately, these models are liable to the same comments and criticisms that we discussed in Section 3.1, and in addition they still lack the comprehensive assessment of their empirical performance that is available for the model of Smets and Wouters (2007), so it is difficult to say if they represent a useful addition to the forecaster's toolkit at this point in time.

4.3.2. Spatial restrictions. When forecasting variables that are aggregated over regions, a further use of theory is the imposition of spatial restrictions that limit the amount of spatial dependence in the system. Disaggregated models can be high-dimensional, and the use of spatial restrictions can be an effective way to overcome the curse of dimensionality. Giacomini and Granger (2004) show how forecasting with disaggregated models that impose spatial correlation restrictions can help to improve the forecast accuracy of the aggregate variable. Hernández-Murillo and Owyang (2006) show that accounting for spatial correlations in regional data can improve forecasts of national employment in the US. Girardin and Kholodilin (2011) also find that allowing for spatial effects when forecasting Chinese output growth results in substantial accuracy improvements. In terms of the usefulness of spatial restrictions for theoretical modelling, I am not aware of existing attempts to do so, but it should not be difficult to embed these types of restrictions into a DSGE model estimated by Bayesian methods, by choosing appropriate priors.

5. FUTURE DIRECTIONS

After the somewhat negative picture painted by the previous sections, I would like to discuss some open questions and possible research directions in which economic theory might have something more promising to say about forecasting. One way to summarize Lessons 2 and 3 is that successful forecasting relies on efficient information extraction from a large cross-section of variables and on accurately modelling the dynamic behaviour of the series of interest. Economic theory might help in disciplining the cross-sectional dependence, but it is unclear that the restrictions imposed by the current generation of DSGE models are able to capture the dynamic behaviour of macroeconomic variables as well as atheoretical econometric models. What seems particularly important for forecasting purposes is to be able to identify the contribution to the dynamics of a model of the internal propagation mechanism as opposed to the (persistent) exogenous shocks, because the two sources of persistence could have very different forecasting implications. A comparison of alternative mechanisms, types of frictions, and the number and type of shocks would also be useful, echoing the previously discussed call by Faust (2012) and Faust and Gupta (2012) for a serious assessment of the empirical plausibility of the transmission mechanisms embedded in current theoretical models. Another key lesson that has emerged from the forecasting literature, and a running theme of this survey, is that structural instabilities are pervasive in economic time series and the forecasting performance of models is itself time-varying, which implies that even forecasting methods with a good historical track record are not guaranteed to perform well in different time periods. Finding forecasting methods that are robust to instabilities is one of the main challenges in the forecasting literature at the moment, and not one that has been successfully resolved yet; see the review by Rossi (2013a). This is an area in which economic theory may provide some useful insights. One possible approach would be to evaluate the internal propagation mechanism of a general equilibrium model not only in terms of its empirical plausibility, but also in terms of its stability over time. Another approach would be to take a step back from general equilibrium models and look for simpler structural relationships that are grounded in economic theory and that, by their very definition, should be stable over time.

We focus next on two general classes of such relationships: those that can be captured by imposing parameter restrictions on existing models, and those that come in the form of moment conditions that involve future observables. Several examples of theoretical restrictions that we have considered in the paper can be viewed as part of the first class, including accounting restrictions, Taylor rules and PPP. One example of the second class is restrictions implied by inter-temporal optimization assumptions, such as expectational Euler equations, which restrict the joint density of future consumption and real interest rates, conditional on current observables. In Section 5.1, I discuss two leading approaches to formulating forecasts that are compatible with these two types of theoretical restrictions.

5.1. Theory-coherent forecasting

5.1.1. Bayesian methods. Bayesian methods are a natural way to impose theoretical restrictions that can be expressed as parameter restrictions. There are several examples in the literature of Bayesian estimation of models using priors based on economic theory. We have already mentioned the approach of Del Negro and Schorfheide (2004), which is an example of formulating priors based on the DSGE model for the purpose of conducting Bayesian inference in reduced-form models. To my knowledge, a serious investigation of whether this approach to estimation can improve the forecasting performance of VARs is still missing from the literature, but it would be interesting to investigate whether such priors can deliver more accurate forecasts than the popular ‘atheoretical’ Minnesota priors typically used in the estimation of Bayesian VARs. Another class of models, where theory-based priors have been successfully used for estimation, is consumer demand models (Montgomery and Rossi, 1999), a case in which theory suggests many relationships that expenditure and price elasticities should fulfil. One appealing feature of incorporating theory-based restrictions using Bayesian methods is the possibility to include a measure of the strength of the beliefs in theory by deciding how dogmatic those priors are. For example, in a model of mutual fund return performance, $r_t = \alpha + \beta r_{mt} + \varepsilon_t$, imposing the no-skill restriction $\alpha = 0$ as a point prior corresponds to ruling out persistent abnormal returns for the fund. This might correspond to an efficient-market restriction. Conversely, viewing α as drawn from, say, a Gaussian distribution with prior standard deviation σ, the higher σ is, the more agnostic the investor is; see Pastor and Stambaugh (2002) and Banegas et al. (2013) for a discussion. Montgomery (2002) similarly discusses an approach to estimating consumer demand that reflects the uncertainty about economic theory, which is based on a hierarchical Bayesian formulation that considers the theoretical restrictions to hold only approximately. Again, it might be interesting to consider similar approaches in the context considered in this review of forecasting macroeconomic variables.

5.1.2. Exponential tilting. We conclude this review by considering economic restrictions that can be expressed as moment conditions involving future observables (e.g. Euler equations). The econometric challenge in deriving forecasts that satisfy this type of restrictions is that moment conditions do not directly provide a conditional density that can be used for forecasting. In all but the simple cases, these conditions cannot easily be converted into parameter restrictions that are amenable to a Bayesian analysis. Giacomini and Ragusa (2014) consider a way to overcome this challenge and propose a general method for producing theory-coherent forecasts when economic theory only provides limited information about the joint conditional density of the variables of interest. The method starts from a base density forecast for the variables of interest (e.g. one implied by a reduced-form model, which in general does not satisfy the theoretical restrictions) and then uses exponential tilting to force the forecast to satisfy a set of moment conditions. The method formalizes previously considered approaches – in particular, Robertson et al. (2005) – in the context of a classical inferential context with out-of-sample evaluation and parameter estimation uncertainty. Giacomini and Ragusa (2014) also show that a tilted density forecast, which incorporates moment conditions that are true in population but may depend on estimated parameters, is more accurate than the initial density forecast when the accuracy is measured by the logarithmic scoring rule of Amisano and Giacomini (2007). The method works as follows. Let $f_{t,h}(\cdot)$ be a h-step-ahead density forecast for the vector of interest Y_{t+h} made at time t and implied by a base model. Consider the problem of incorporating into the forecast a set of k theory-based moment conditions:

$$E_t[g(Y_{t+h}, \theta_0)] = 0. \tag{5.1}$$

The parameter θ_0 could be calibrated or estimated using in-sample data, in which case one would substitute θ_0 with the in-sample estimate $\widehat{\theta}_t$ in the procedure described below. The tilting procedure yields a new density forecast $\widetilde{f}_{t,h}$, which is the (unique) density that, out of all the densities that satisfy the moment conditions, is the closest to $f_{t,h}$ according to the Kullback–Leibler information criterion, i.e., it solves

$$\min_{h_{t,h}\in\mathcal{H}} \int \log \frac{h_{t,h}(y)}{f_{t,h}(y)} h_{t,h}(y)\, dy \tag{5.2}$$

$$s.t. \int g(y, \theta_0) h_{t,h}(y)\, dy = 0. \tag{5.3}$$

Under general conditions, one can show that the tilted density is given by

$$\widetilde{f}_{t,h}(y) = f_{t,h}(y) \exp(\eta_{t,h}(\theta_0) + \tau'_{t,h}(\theta_0) g(y, \theta_0)), \tag{5.4}$$

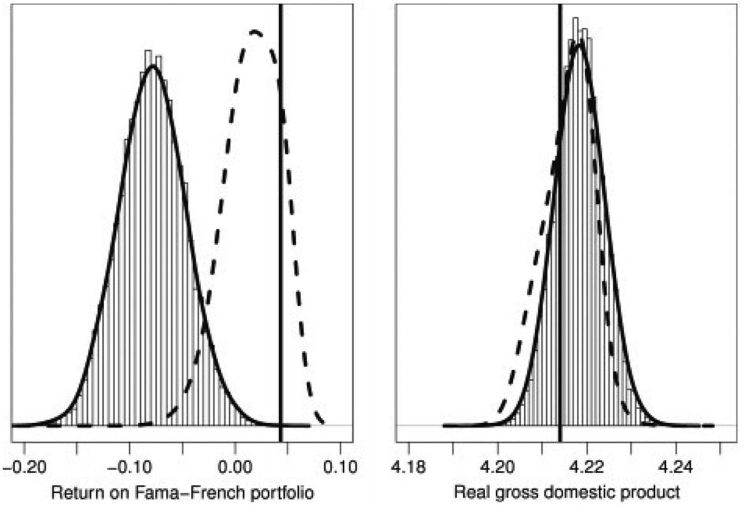

Figure 1. Base and tilted density forecasts.

where $\eta_{t,h}(\theta_0)$ and $\tau_{t,h}(\theta_0)$ are defined as

$$\tau_{t,h}(\theta_0) = \arg\min_{\tau} \int f_{t,h}(y) \exp(\tau' g(y, \theta_0))\, dy \tag{5.5}$$

$$\eta_{t,h}(\theta_0) = \log\Big(\int f_{t,h}(y) \exp(\tau_{t,h}(\theta_0)' g(y, \theta_0))\, dy\Big)^{-1}. \tag{5.6}$$

In practice, the tilted density at time t can be obtained numerically as follows. First, generate D draws $\{y^d\}_{d=1}^{D}$ from $f_{t,h}(y)$. Second, solve $\tau_{t,h} = \arg\min_\tau (1/D) \sum_{d=1}^{D} f_{t,h}(y^d)$ $\exp(\tau' g(y^d, \theta_0))$. Finally, obtain $\eta_{t,h}$ as $\eta_{t,h} = \log((1/D) \sum_{d=1}^{D} f_{t,h}(y^d) \exp(\tau_{t,h}' g(y^d, \theta_0)))^{-1}$. In the paper, we apply the method to base forecasts from a Bayesian VAR model for 22 variables including real consumption (C_t), real returns on the Fama–French portfolio R_t and real GDP; see Giacomini and Ragusa (2014) for data description.[1] We incorporate the Euler condition $E_t[\beta(C_{t+1}/C_t)^{-\alpha} R_{t+1} - 1] = 0$, with calibrated parameters $\beta = 0.999$ and $\alpha = 0.6$. We find that imposing the Euler equation restrictions improved the out-of-sample accuracy of the model's density forecasts over the period 1980–2009 and it also resulted in point forecast accuracy gains for asset returns. As an illustration of the effect of tilting at one particular date in the sample, Figure 1 shows the base and tilted density forecasts for the real GDP and for the return R_t on a Fama–French portfolio at one particular point in time (1988:Q1): the histogram is the one-step-ahead density forecast implied by a BVAR with 22 variables, which include real GDP, non-durables and services real consumption, the federal funds rate and the return on the Fama–French portfolio. In each graph, the dashed line is the tilted density forecast that incorporates the Euler equation; the solid vertical line is the realization of the variable. Note that returns are directly restricted by the Euler equation, whereas GDP does not enter the Euler equation but

[1] R_t is the value-weighted CRSP stocks on NYSE+Nasdaq+Amex minus the inflation rate.

is nonetheless affected by the tilting because of the dependence among all the variables in the system. The figure shows that, for both variables, incorporating the Euler equation restrictions changes the shape of the densities and it shifts them towards the variable realizations, thus improving the accuracy of both forecasts.

The application is just an illustration of the opportunities that the method offers for investigating the effects of imposing theoretical restrictions on an existing forecast. The advantages of the method are that it offers flexibility in choosing the base models and the number of theoretical moment conditions, it allows one to impose moment conditions that only involve a subset of the variables of interest (as will typically be the case) and to capture their effect on all the variables in the base model, and it allows one to incorporate nonlinear restrictions in a straightforward manner.

6. CONCLUSION

This paper presents a selective review of the literature in an attempt to understand whether and how economic theory can help in forecasting. The literature can be grouped into three main lessons. The first lesson is that the econometric methodology used in answering the question matters and one should conduct a serious sensitivity analysis to assess the impact of the many arbitrary choices that typically underlie the analysis. The second lesson is that two types of theory-based forecasting that have not been proven successful are theory-guided variable selection in univariate models and forecasting using estimated DSGE models, such as the popular model by Smets and Wouters (2007). Both approaches are generally outperformed by simple reduced-form models and by survey forecasts. Even for the case in which the DSGE model seems to produce accurate forecasts (for output growth at long horizons), it is unclear how much of the result could be driven by the model's implicit assumptions about trends and cointegrating relationships. As a response to the often-heard argument that favours DSGE models because of their story-telling ability, we suggest that the argument lacks a formal decision-theoretic underpinning, which for example would make it difficult to select among competing stories characterized by similar empirical performance. The third lesson is that there are some methods that deliver accurate forecasts at short horizons, but they are not generally grounded in economic theory. I am referring in particular to the use of survey-based forecasts or methods that extract information from large data sets while controlling the dimension of the system, such as factor models, Bayesian shrinkage or model combination. At long forecast horizons, instead, there is some evidence that simple theory-based cointegrating restrictions can help. Another conclusion is that economic theory seems to have more to say about the cross-sectional dependence among variables (e.g. via the use of accounting identities, disaggregation and spatial dependence restrictions) than about their dynamic behaviour. I conclude on an optimistic note by suggesting that economic theory might be a useful guide in finding forecasting methods that are robust to structural instability, a problem that is widespread in economic data and that impairs the forecasting ability of atheoretical models.

ACKNOWLEDGEMENTS

I thank the editor, Richard Smith, and an anonymous referee for useful comments and suggestions, and I gratefully acknowledge financial support from the Economic and Social Research Council (ESRC) through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001.

REFERENCES

Adolfson, M., J. Linde and M. Villani (2007). Forecasting performance of an open economy Dynamic Stochastic General Equilibrium model. *Econometric Reviews 26*, 289–328.

Adolfson, M., S. Laseen, J. Lindé and M. Villani (2008). Evaluating an estimated new Keynesian small open economy model. *Journal of Economic Dynamics and Control 32*, 2690–721.

Altavilla, C., R. Giacomini and G. Ragusa (2013). Anchoring the yield curve using survey expectations. CWP 52/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London.

Altavilla, C., R. Costantini and R. Giacomini (2014). Bond returns and market expectations. *Journal of Financial Econometrics 12*, 708–29.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics 25*, 177–90.

Banegas, A., B. Gillen, A. Timmermann and R. Wermers (2013). The cross section of conditional mutual fund performance in European stock markets? *Journal of Financial Economics 108*, 699–726.

Bermingham, C. and A. D'Agostino (2014). Understanding and forecasting aggregate and disaggregate price dynamics. *Empirical Economics 46*, 765–88.

Boivin, J. and M. Giannoni (2006). DSGE models in a data-rich environment. Working Paper 12772, National Bureau of Economic Research.

Bouakez, H., E. Cardia and F. Ruge-Murcia (2014). Sectoral price rigidity and aggregate dynamics. *European Economic Review 65*, 1–22.

Caldara, D., R. Harrison and A. Lipinska (2012). Practical tools for policy analysis in DSGE models with missing channels. Finance and Economics Discussion Series 2012-72, Board of Governors of the Federal Reserve System (US).

Canova, F. (2009). How much structure in empirical models? In T. Mills and K. Patterson (Eds.), *Handbook of Applied Econometrics*, *Volume 2*, 68–97. London: Palgrave Macmillan.

Canova, F. and F. Ferroni (2011). Multiple filtering devices for the estimation of cyclical DSGE models. *Quantitative Economics 2*, 73–98.

Carriero, A. and R. Giacomini (2011). How useful are no-arbitrage restrictions for forecasting the term structure of interest rates? *Journal of Econometrics 164*, 21–34.

Carson, R. T., T. Cenesizoglu and R. Parker (2011). Forecasting (aggregate) demand for US commercial air travel. *International Journal of Forecasting 27*, 923–41.

Chun, A. L. (2009). Forecasting interest rates and inflation: blue chip clairvoyants or econometrics? EFA 2009 Bergen Meetings Paper. Available at http://ssrn.com/abstract=946667.

Clark, T. and M. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics 105*, 85–110.

Clements, M. P. and D. F. Hendry (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics 10*, 127–46.

D'Agostino, A. and B. Schnatz (2012). Survey-based nowcasting of US growth: a real-time forecast comparison over more than 40 years. ECB Working paper 1455, European Central Bank.

Del Negro, M. and F. Schorfheide (2012). DSGE model-based forecasting. Staff report 554, Federal Reserve Bank of New York.

Del Negro, M. and F. Schorfheide (2004). Priors from general equilibrium models for VARs. *International Economic Review 45*, 643–73.

Del Negro, M., F. Schorfheide, F. Smets and R. Wouters (2007). On the fit of new Keynesian models. *Journal of Business and Economic Statistics 25*, 143–62.

De Mol, C., D. Giannone and L. Reichlin (2008). Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*, 318–28.

Deschamps, B. and C. Ioannidis (2013). Can rational stubbornness explain forecast biases? *Journal of Economic Behavior and Organization 92*, 141–51.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–64.

Edge, R. M. and R. S. Gürkaynak (2010). How useful are estimated DSGE model forecasts for central bankers? *Brookings Papers on Economic Activity 41*, 209–59.

Edge, R. M., M. T. Kiley and J. P. Laforte (2008). Natural rate measures in an estimated DSGE model of the US economy. *Journal of Economic Dynamics and Control 32*, 2512–35.

Edge, R.M., M. T. Kiley and 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*, 720–54.

Elliott, G. (2006). Forecasting with trending data. In G. Elliott, C. W. J Granger and A. Timmerman (Eds.), *Handbook of Economic Forecasting*, *Volume 1*, 555–604. Amsterdam: Elsevier.

Faust, J. (2012). DSGE models: I smell a rat (and it smells good). *International Journal of Central Banking 8* (1), 53–64.

Faust, J. and A. Gupta (2012). Posterior predictive analysis for evaluating DSGE models. Working Paper 17906, National Bureau of Economic Research.

Faust, J. and J. H. Wright (2013). Forecasting inflation. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, *Volume 2A*, 2–56. Amsterdam: Elsevier.

Giacomini, R. (2013). The relationship between DSGE and VAR models. *Advances in Econometrics 32*, 1–25.

Giacomini, R. and C. W. J. Granger (2004). Aggregation of space–time processes. *Journal of Econometrics 118*, 7–26.

Giacomini, R. and G. Ragusa (2014). Theory-coherent forecasting. *Journal of Econometrics 182*, 145–55.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies 76*, 669–705.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*, 595–620.

Giacomini, R. and B. Rossi (2015). Model comparisons in unstable environments. CWP 13/12, Centre for Microdata Methods and Practice, Institute for Fiscal Studies and University College London. Forthcoming in *International Economic Review*.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–78.

Girardin, E. and K. A. Kholodilin (2011). How helpful are spatial effects in forecasting the growth of Chinese provinces? *Journal of Forecasting 30*, 622–43.

Granger, C. W. J. and M. J. Machina (2006). Forecasting and decision theory. In G. Elliott, C. W. J. Granger and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, *Volume 1*, 82–98. Amsterdam: Elsevier.

Groen, J. J. J., G. Kapetanios and S. Price (2009). A real time evaluation of Bank of England forecasts of inflation and growth. *International Journal of Forecasting 25*, 74–80.

Gürkaynak, R., B. Kısacıkoğlu and B. Rossi (2013). Do DSGE models forecast more accurately out-of-sample than VAR models? *Advances in Econometrics 32*, 27–79.

Hansen, P. R. and A. Timmermann (2013). Equivalence between out-of-sample forecast comparisons and Wald statistics. CREATES Working paper 2012-45, School of Economics and Management, University of Aarhus.

Hendry, D. F. and K. Hubrich (2006). Forecasting economic aggregates by disaggregates. ECB Working paper 0589, European Central Bank.

Hernández-Murillo, R. and M. T. Owyang (2006). The information content of regional employment data for forecasting aggregate conditions. *Economics Letters 90*, 335–39.

Inoue, A. and L. Kilian (2005). In-sample or out-of-sample tests of predictability: which one should we use? *Econometric Reviews 23*, 371–402.

Meese, R. A. and K. S. Rogoff (1983). Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics 14*, 3–24.

Montgomery, A. (2002). Reflecting uncertainty about economic theory when estimating consumer demand. *Advances in Econometrics 16*, 257–94.

Montgomery, A. L. and P. E. Rossi (1999). Estimating price elasticities with theory-based priors. *Journal of Marketing Research 36*, 413–23.

Moretti, G. and G. Nicoletti (2010). Estimating DSGE models with unknown data persistence. Working paper 750, Bank of Italy.

Pagan, A. (2003). Report on modelling and forecasting at the Bank of England. *Bank of England Quarterly Bulletin 43*, 1–29.

Pastor, L. and R. F. Stambaugh (2002). Investing in equity mutual funds. *Journal of Financial Economics 63*, 351–80.

Perevalov, N. and P. Maier (2010). On the advantages of disaggregated data: insights from forecasting the U.S. economy in a data-rich environment. Working Paper 10-10, Bank of Canada.

Pesaran, M. H. and R. P. Smith (2011). Beyond the DSGE straitjacket. Discussion Paper 5661, Institute for the Study of Labor (IZA).

Robertson, J. C., E. W. Tallman and C. H. Whiteman, (2005). Forecasting using relative entropy. *Journal of Money, Credit and Banking 37*, 383–401.

Romer, C. D. and D. H. Romer (2000). Federal Reserve information and the behavior of interest rates. *American Economic Review 90*, 429–57.

Rossi, B. (2013a). Exchange rate predictability. *Journal of Economic Literature*, *51*, 1063–119.

Rossi, B. (2013b). Advances in forecasting under instability. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, *Volume 2B*, 1203–324. Amsterdam: Elsevier.

Sims, C. (1980). Macroeconomics and reality. *Econometrica 48*, 1–48.

Sims, C. (2003). Comments on Smets and Wouters. Working paper, Princeton University. Available at http://sims.princeton.edu/yftp/Ottawa/SWcommentSlides.pdf.

Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review 97*, 586–606.

Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. In G. Elliott, C. W. J. Granger and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, *Volume 1*, 515–54. Amsterdam: Elsevier.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*, 1067–84.