COMBINING VAR FORECAST DENSITIES USING FAST FOURIER TRANSFORM

Jakub Ryšánek[*]

1. Introduction

Combining predictions of macroeconomic aggregates and financial variables from various models is a popular technique in applied macro research. One advantage, among others, that speaks for combining different models is that it provides an insurance against selecting inappropriate models as Bache *et al.* (2009) emphasize.

Traditional econometrics may suggest selection of one model that appears to be the best in terms of, say, information criteria or relevant p-values. Information resulting from other models is then *a priori* discarded. In contrast, combining of competing models introduces uncertainty to the model structure, in which outcomes from each single model are taken into account with appropriate weight and no prior information is thus ignored.

First attempts to combine forecasts of different candidate models led to combinations of point forecasts – see Bates and Granger (1969). Logical generalization can be found in literature in the form of combining interval forecasts. Kascha and Ravazzolo (2009) provide a nice overview of the most popular methods in use in this area. Nevertheless, the authors rely on normality of estimated forecasts, for which there exists a good reason with respect to time that would otherwise be necessary to simulate all the candidate models.

Prior model uncertainty is a feature that can be fully exploited from the Bayesian perspective – an approach usually termed *Bayesian model averaging* (BMA), as in Andersson and Karlsson (2007), and Hoeting *et al.* (1999), where each model is *a priori* given a weight in the form of a probability distribution and data is then used to transform prior beliefs into posterior estimates.

In this paper, I propose an estimation of a combined GDP forecast for the Czech economy based on a broad set of vector autoregressive models (VARs). Following Stock and Watson (2004), among others, I choose *ex post* prediction as the main criterion for comparison of competing candidate models. Most Central European economies suffer from short time series available, which suggests that normal approximations of any kind can lead to underestimation of variances of relevant model

outcomes. Monte Carlo methods are therefore fully utilized instead, which makes the analysis computationally demanding. On the other hand, the aim of this paper is to show that simulated predictive densities can be combined in light of convolution theory, specifically via the discrete Fourier transform, for which a fast algorithm was developed (*fast Fourier transform* – in short, FFT). This would not make the analysis appealing since FFT itself is not trivial to compute especially if large input vectors need to be combined repeatedly. However, the main finding of this paper is that FFT works well even if predictive densities of candidate models are poorly simulated with only, say, 100 iterations. While such improper simulation would not lead to meaningful results under usual circumstances, it is shown further that FFT can help here significantly in terms of reduction in simulation time while keeping quality of results very high.

The FFT algorithm is often used in technical applications, e.g., in digital signal processing. It is common to use filtering techniques in macroeconomics to reveal business cycle information from various time series. However, to my best knowledge, I have not encountered any application of FFT in combining predictive densities yet.

Even though this paper is intended to be based on the Bayesian way of thinking, it does not follow the pure Bayesian model averaging approach since the frequentist approach is taken as well during certain stages.

The paper is organized as follows: section 2 comprises the set of competing models that enter the analysis; section 3 uncovers underlying specific features of the Czech economy that affect the modeling approach taken here and data used; section 4 describes the technique of selecting a representative model subset since simulation of all the models considered would not be achievable in reasonable time; section 5 shows how candidate models can be evaluated with respect to their predictive powers, and the combination of candidate models via fast Fourier transform is described; section 6 summarizes the results of the Czech GDP prediction, and section 7 concludes.

2. Suite of candidate models

As stated above, the focus of this paper is on combination of predictions from the broadly defined class of VAR models. From a technical point of view, I consider in turn traditional VARs (see, e.g., Hamilton, 1994; Enders, 2004) and Bayesian VARs – BVARs (see, e.g., Lütkepohl, 2007; Canova, 2007) of the form

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + w_t,$$

where y_t is a vector of k endogenous variables, w_t is a k-dimensional white noise process for which the following facts hold: $E(w_t) = 0$, $E(w_t w_t`) = \Sigma w$ and $E(w_t w_s`) = 0$ for s ≠ t.

In the case of BVARs, the scheme of prior distributions in the Minnesota style as proposed by Doan et al. (1984) or Litterman (1986) fits models where variables are in levels. Regression parameters have an *a priori* zero mean, only autoregressive parameters of order 1 in each equation have a mean equal to one. Time series are thus coped with as independent random walks as opposed to this paper, where variables are

modeled as stationary processes and therefore, depending on the type of the imposed stationarity, I suggest a prior zero mean for all coefficients in the difference stationary processes and AR(1) coefficients equal to 0.8 for the trend stationary processes.

Following Lütkepohl (2007), the covariance matrix of BVAR parameters is diagonal[1] with diagonal elements

$$v_{ij,l} = \begin{cases} (\lambda / l)^2 & i = j, \\ (\lambda\theta\sigma_i/l\sigma_j)^2 & i \neq j, \end{cases}$$

where $v_{ij,l}$ is the prior variance of the i-th coefficient in the j-th equation, l stands for lag, λ is the parameter of tightness, θ is the decay parameter from the interval (0;1) and σ_i is the i-th diagonal element of the covariance matrix of residuals – the ratio of σ_i /σ_j normalizes the different scales of variables i and j.

Since only the forecasting performance of models matters, models were estimated in a reduced form without the need to identify structural hyperparameters.

3. Czech specific features and data used

As was mentioned in the introduction, the suggested method of combining predictive densities is demonstrated on a prediction of the Czech GDP. The Czech Republic is a small open economy that adopted inflation targeting in 1998, when it abandoned the fixed exchange rate regime. Increasing foreign demand has proved to be the key factor in the sound development of the Czech economy in the last 2 decades. The financial turmoil experienced in the end of 2008 hit the economy fully in 2009 through a decrease in foreign demand. The vulnerability of the Czech economy nowadays also emanates from the exchange rate that is hard to predict. Relevant domestic time series exist for the past 15 years. With this in mind, I constructed a series of several models that have performed well in modeling closed economies, and a rather larger set of models that allow transmission of shocks from abroad – the link between the domestic economy and EU members is captured via export/import prices and the exchange rate (Euro). The models considered contain 2 to 4 equations; a simple AR process for the GDP is also assumed. The maximum lag length was set in turn to 1 – 4 quarters. The domestic time series hardly allow longer lag length. Variables considered are, e.g., domestic/ foreign GDP, inter-bank market interest rates, inflation measures (Czech CPI, European HICP), dynamics of newly issued credit, and unemployment rate.

Since GDP is a complex indicator of an economy's output that comprises hetero-geneous components, isolated blocks of GDP were studied in separate sub-models for private and government consumption (C),[2] investment (I), exports (X), and imports (M). The high ratios of C/GDP and I/GDP hold approximately constant throughout the

---

1   A diagonal form of the covariance matrix of regression parameters corresponds with a VAR model
    represented in a compact form.

2   I have neglected consumption of non-profit organizations.

course of time in the Czech Republic just like in large economies.[3] The foreign trade variables (X and M) exhibit a permanently increasing proportion of the GDP.

Since the analysis is focused on GDP prediction, all the GDP components and their forecast densities can be transformed to the GDP with the help of an *error correction model* (ECM) based on the following intuition. Having said that, e.g., aggregate consumption retains a constant proportion in the GDP, it is possible to regress the GDP on the consumption in a linear model of the form

$$GPD_t = (\alpha_0) + \alpha_1 C_t + u_t, \tag{1}$$

where the coefficient α_1 represents the long-run proportion of consumption in output. The constant term α_0 can be included for technical reasons, if needed, and u_t is an *i.i.d.* error term of the usual properties. We can call equation (1) a cointegration regression since we assume that this relationship stays valid in the long run, whereas short-run dynamics can be described by the following ECM (the variables GDP_t and C_t are in differences this time):

$$\Delta GDP_t = \beta_0 + \beta_1 \Delta C_t - \gamma u_{t-1} + v_t, \tag{2}$$

where β_0, β_1 and γ are regression coefficients, u_{t-1} stands for the lagged deviation from the long-run equilibrium defined in (1) and v_t, again, is an *i.i.d.* error term of the usual properties.

The whole idea of transforming GDP components to GDP remains the same in the case of the foreign trade variables, whose proportion to the GDP increases over time. Only the cointegration regression is now derived so as to embed the time-varying long-run equilibrium. If the ratio of exports on output demonstrates a linear trend, the same can be deduced about the inverse ratio, or we can expect

$$\frac{GDP_t}{X_t} \approx \alpha_2 + \alpha_3 t,$$

where α_2 and α_3 are coefficients, t denotes time. By equating both sides and multiplying by X_t, we arrive at the final cointegration regression

$$GPD_t = \alpha_2 X_t + \alpha_3 t X_t + v_t, \tag{3}$$

where, for the sake of completeness, we add an *i.i.d.* error term of the usual properties, w_t. Since the ECM itself tracks short-run dynamics of the model around the long-run trend, it is not necessary to implement any changes in the ECM equation itself as long as the cointegration is handled by (3) instead of (1). Thus,

$$\Delta GDP_t = \beta_2 + \beta_3 \Delta X_t - \beta w_{t-1} + v_t,$$

---

3   The I/GDP ratio shows a higher volatility compared to the C/GDP ratio.

where interpretation of the regression coefficients and the error term coincides with those in equation (2).

Stationarity of modeled variables was achieved via differencing or with the help of a *band-pass filter* as suggested by Baxter and King (1995) and further improved by Christiano and Fitzgerald (1999). Parameters of the CF filter were set so as to extract business cycle information from time series that corresponds to a spectrum of frequencies between 6 and 32 quarters as Canova (2007) proposes.

The number of all the considered models is quite high. All the model alternatives can be evaluated as:

VAR versus BVAR	2 variants
model suite	88 variants[4]
maximum lag length	4 variants
stationarity type	2 variants
	2 x 88 x 4 x 2 = 1,408 models.

It is possible to encounter even larger pools of models in the literature: Bache *et al.* (2009) has 1,777 models; Sala-i-Martin *et al.* (2004), who deal with a single equation model, consider a huge amount of possible regressors resulting in $2^{67} \approx 1.48 \times 10^{20}$ model variants. They try to overcome this unbelievable number by using a stochastic algorithm from the class of MCMC (*Markov chain Monte Carlo*) methods.

My ambition here was to combine whole predictive densities, not just point forecasts. To simulate all the 1,408 models[5] would be a tedious task. Hence, while exploiting the power of MCMC methods, I tried to create a representative subset of candidate models. The selected models then served as objects of simulation in the subsequent analysis. The procedure of model selection is described in the following section.

4. Selecting a representative model subset

Sufficient simulation that would ensure an estimation of the moment characteristics of the predicted endogenous variable (GDP) with high preciseness is time demanding. In the case of the *residual bootstrap*, which I applied to VARs in a reduced form, Efron (1979) suggests several thousand simulation iterations. In the case of BVARs, prior coefficient normality results in multivariate normal posterior estimates due to the principle of conjugacy. It is possible to use the intuitive *Gibbs sampler* to generate draws from relevant distributions (see Lancaster, 2005). This technique, however, incorporates the property of a *Markov chain* and generated draws are correlated. Taking, say, every tenth draw can minimize the degree of correlation. Sadly, the simulation time increases ten times (!) in this way. All these facts indicate that in order to keep the simulation time manageable for the combined model, it is necessary to reduce the model space.

---

4   14 models for a closed economy, 74 models for an open economy.

5   Some of the 1,408 models were estimated as unstable – at least one unit root was identified in the coefficient matrix of the transformed VAR(1) model. Such models were not coped with in the further analysis.

Canova (2007) and Lancaster (2005) offer a heuristic from the class of MCMC methods – the *Metropolis algorithm*. This technique overlaps to a great extent with the *simulated annealing* algorithm. All we need is to specify a criterion that will be minimized and a simple rule stating how the model space will be explored.

For that purpose, I chose a data span 1996Q1 – 2007Q4 as a training period from which ex post predictions for all the available models were computed for the following 4 quarters. For the sake of simplicity, only point forecasts were produced. The model space can be encoded into a vector of zeros and ones of a length of 1,408, i.e.,

$$M = (\underbrace{0,0,1,0,1,1,\ldots\ldots\ldots\ldots}_{1408})^T,$$

where   1 = 'partial model will be included in final combined model',

0 = 'partial model will be discarded'.

Slight changes in the vector *M* (random substitution of 1 for 0 and vice versa) serve as a means of exploring the space of competing models.

The criterion function has three components: (*i*) the measure of mean absolute deviations of the selected models from reality, (*ii*) the measure of symmetry of the selected model predictions around reality, and (*iii*) the number of sub-models approved for the combined prediction. Minimization of component (*iii*) is crucial since the key goal of this analysis is to cut down the number of models while not violating (*i*) and (*ii*) too much. The idea behind the construction of all the components is briefly described further on.

Measure (*i*), the mean absolute deviation (MAD) of *ex post* predictions from reality, is evaluated only over models that are considered in the given iteration (for which $M_j = 1$):

$$\frac{\sum_{j \in M} \left| \hat{y}_j^h - y_j^h \right|}{n},$$

where *j* is an index of the partial model, *h* stands for *h*-step ahead forecast, $\hat{y}_j^h$ and y_j^h denote model prediction and reality, respectively. For the whole testing period, this indicator can be aggregated for all *h*-step ahead forecasts (running from 1 to 4) with a discount factor ρ which dampens the effect of deviations of the models and reality in longer horizons:

$$\sum_{h=1}^{4} \left[\frac{\sum_{j \in M} \left| \hat{y}_j^h - y_j^h \right|}{n} \right] \cdot \rho^h \quad \ldots \rho \in \left(0;1 \right\rangle. \tag{4}$$

77

Besides MAD, it is possible and perhaps desirable to penalize models that constantly overshoot reality from models that fluctuate around reality in the prediction horizon. This idea is captured in measure (*ii*), where a model that overestimates the reality in two quarters and underestimates the reality in the other two quarters is considered more valuable than a model that under/overshoots four times. An aggregation over the selected models yields

$$\sum_{j \in M} \left| x_j - 2 \right|, \qquad (5)$$

where x_j is the number of overshoots of the *j*-th model.

Furthermore, component (*iii*) penalizes large model subsets against small model subsets, e.g., by

$$\exp\left(\frac{M^T E}{E^T E} \right), \qquad (6)$$

where E is a vector of ones of a length of 1,408.

An aggregate criterion function, δ, is then an additive combination of (4), (5) and (6) and a vector of weights with the property $w_i \in (0;1), \sum w_i = 1$:

$$\delta = \left\{ \sum_{h=1}^{4} \left[\frac{\sum_{j \in M} \left| \hat{y}_j^h - y_j^h \right|}{n} \right] \cdot \rho^h \right\} \cdot w_1 + \sum_{j \in M} \left| x_j - 2 \right| \cdot w_2 + e^{\frac{M'E}{E'E}} (1 - w_1 - w_2). \qquad (7)$$

Now this function may look intimidating, but on top of that, it is necessary to normalize each component, C, according to $(C - C_{min})/(C_{max} - C_{min})$ in order to get meaningful results.

The Metropolis algorithm as suggested by Canova (2007) and Lancaster (2005), tailored to this optimization problem, can be summarized in following steps:

1. Initialize the vector M with ones (all models selected), δ = sufficiently high value, iteration index $i = 0$, maximum of iterations $i_{max} = 10,000$;

2. Generate free parameters from their prior distributions, if any such parameters exist – I used loose prior definition of $\rho \sim$ Uniform(0.8,1);

3. Randomly select, say, 10 candidate models (i.e., 10 indices in M) => if a model was not previously chosen, choose it next time, if it was chosen, ignore it next time;[6]

---

6   At ten randomly selected models, we switch 0 for 1 and vice versa.

4. Evaluate δ_i;

5. If $\delta_i < \delta_{i-1}$, then accept the suggested M, otherwise accept the M with a probability of $\{1; \delta_{i-1}/\delta_i\}$

6. If $i = i_{max}$, the algorithm stops, else proceed with step 2.

With this routine, I managed to cut down the number of selected models from 1,408 to 159. The initial values were set as follows: $i_{max} = 10{,}000$; $w_1 = 0.05$; $w_2 = 0.8$. The posterior mode for ρ came out as 0.93. An overview of the selected models can be found in the Appendix at the end of this paper.

5. Combining predictive densities

The set of 159 representative candidate models in the preceding analysis was selected with respect to their relative successfulness *in ex* post point predictions. These models can now be simulated over moving time windows to provide a series of both *ex post* and *ex ante* predictions, including relevant confidence bands. *Residual bootstrapping*[7] was fully utilized in the case of traditional reduced-form VAR models, *Gibbs sampling* helped in the case of BVARs.

Gerard and Nimark (2008) propose an expert linear opinion pool which can combine the outputs of various models into a final distribution of the relevant variable (GDP):

$$p_t^c(gdp_{t+h}) = \sum_{m=1}^{159} p_{t,m}(gdp_{t+h})w_{m,h},$$

where $p_t^c(gdp_{t+h})$ is a combined probability density function *(pdf)* of the GDP evaluated at the time t on h periods ahead, $p_{t,m}(gdp_{t+h})$ is the *pdf* of the GDP implied by the m-th model, and $w_{m,h}$ is a weight of the m-th model. The weights of particular models can be derived in a Bayesian style[8] using

$$w_{m,h} = \frac{p_{t,m}(gdp_{t+h})p(M_m)}{\sum_{i=1}^{159} p_{t,i}(gdp_{t+h})p(M_i)},$$

where $p(M_m)$ is the prior belief in the m-th model.

---

7 It is for this time-consuming bootstrap and Gibbs sampler that the number of the candidate models had to be reduced to 159.

8 Pure Bayesian analysis requires input marginal *pdf* of predicted variable, $p_t^c(gdp_{t+h})$, to be also estimated in a Bayesian manner taking into account the uncertainty of regression parameters, which is not met in the case of VAR, where the frequentist bootstrap was used to derive the relevant moment characteristics. The view taken here therefore does not belong to genuine *Bayesian Model Averaging* (BMA).

5.1 Combining predictions using Fourier transform

The foregoing combinations via an expert linear opinion pool work well provided that all partial models yield sufficiently accurate outcomes. Insufficient simulation usually causes moment characteristics under consideration to be biased. One could simply conclude that there is a trade-off between the quality of model outcomes on the one hand and the necessary simulation time on the other, both bound in the same direction. I would like to argue, though, that the use of *discrete Fourier transform* (DFT) provides a convenient framework in which poorly simulated models can be combined without severe loss of the quality of results. As a consequence, a significant portion of time can be saved during the simulation.

The underlying idea is as follows. Results from different sub-models in the form of histograms are combined as in a series of convolutions. We can exploit the fact that a convolution of two random variables can be easily derived once we transform both variables into a frequency domain – here, convolution becomes an easy task of element-by-element multiplication.

In the case of combining two variables, the whole procedure can be summarized in these steps: (1) we apply the DFT algorithm to each variable in order to convert the underlying histograms into frequency domain counterparts, (2) transformed variables get multiplied element by element, (3) we apply the inverse DFT to convert the result back into the original domain, (4) the average distribution of both variables is achieved by a horizontal adjustment – e.g., we have simulated histograms of variables X and Y at hand and need to know the distribution of (X+Y)/2 (i.e., distribution of the average). The first three steps of the algorithm only evaluate the distribution of (X+Y), thus, each position of the resulting histogram needs to be moved horizontally in order to achieve a true (X+Y)/2 distribution.

Asmar (2005) shows the analytical formula for the DFT procedure:

$$F(x_k) = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn}, \ k = 0, ..., N-1,$$

where i is an imaginary unit, x is an original vector corresponding to a histogram of the underlying stochastic proce ss. An efficient algorithm called *fast Fourier transform* (FFT) was developed in the 1960s to carry out both DFT and inverse DFT with impressive speed – see Asmar (2005), page 455.[9]

As mentioned above, I claim that the use of FFT when combining predictive densities of various models can significantly reduce computational time since even poorly simulated candidate models provide enough information for the FFT algorithm to sketch the combined density with either negligible loss in the quality of the results or no loss at all.

---

9   I have exploited the power of Matlab functions 'fft' and 'ifft' to compute FFT and inverse FFT, respectively.

This fact is empirically proved in Figures 1 and 2, where I compare the performance of traditional simulation and FFT upgrade in turns for exhaustive simulation (2,000 iterations) and poor simulation (100 iterations), respectively. For illustrative purposes here, I assume only computation of the convolution X+Y instead of the weighted sum w_1X+w_2Y. Section 5.2 deals with this relaxation.

Figure 1
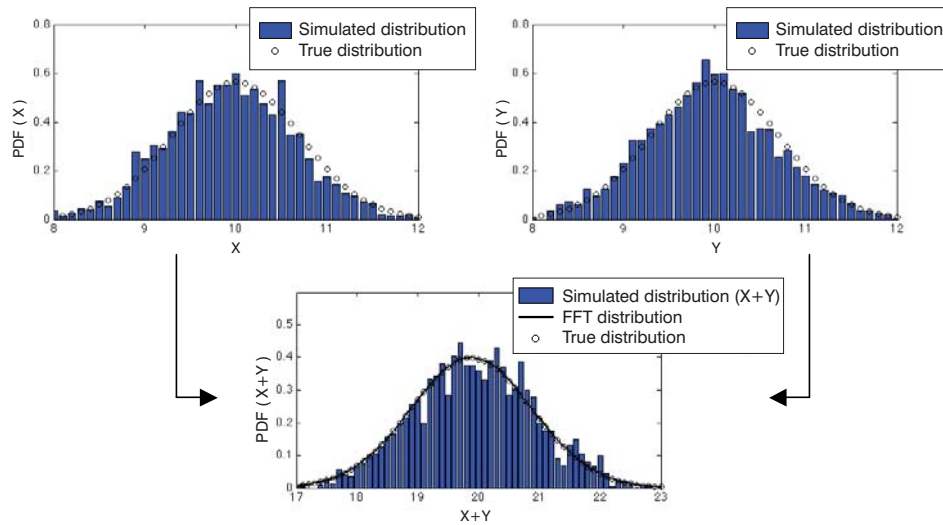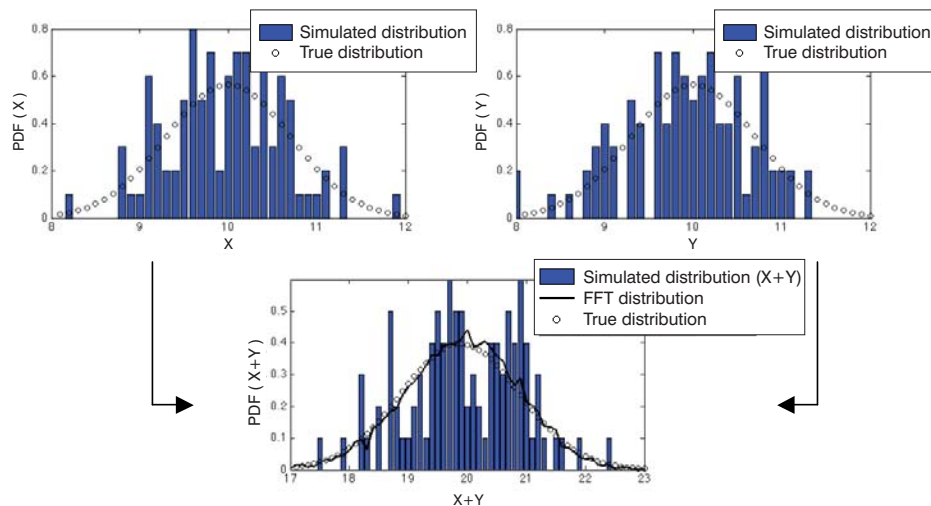
Simulation experiment – 2,000 iterations

Figure 2

Simulation experiment – 100 iterations

Both Figure 1 and Figure 2 illustrate the same kind of experiment. The only difference is in the number of simulation iterations that were used to produce the appropriate graphs. The random variables X and Y are generated from normal distribution[10] with a mean of 10 and a variance of 0.5. X and Y are drawn from their distributions independently. The question at hand could be to ask what the distribution of the random variable X+Y is. Note that the independence assumption is crucial; otherwise the use of the FFT algorithm would not be possible. This corresponds with real applications, where predictive densities of partial models are also taken as independent and combined afterwards. In traditional simulation, we generate X and Y independently, add them together to create X+Y and by doing this repeatedly, we get the distribution of X+Y.[11] In this trivial example, it is easy to deduce analytically that the distribution of X+Y will again be normal with a mean of 20 and a variance equal to 1, which emanates from the property of independence of X and Y. It is, however, interesting to see how this distribution is approximated with traditional simulation and with the FFT upgrade in the spirit of this section for 2,000 and 100 simulation iterations, respectively.

In Figure 1, both partial histograms for X and Y relatively well approximate the bubble charts that represent true Gaussian curves. This causes the variable (X+Y) to also be approximated to a very high degree of preciseness. Note that convolution via FFT is useless in this case since the same result can be achieved without it.

The situation in Figure 2 is very different. As the partial histograms for X and Y approximate true Gaussian processes relatively poorly, the simulated distribution of (X+Y) is also of poor quality – it even looks bimodal (!). On the other hand, the convolution via FFT takes the same inputs for X and Y but the result represented by a solid line closely traces the true Gaussian curve for the process of (X+Y). While the number of simulation iterations in this case is insufficient for constructing a simulated distribution of (X+Y), it is a fair enough amount to be used to convolve X and Y with only a minor loss in quality.

The reason why the FFT upgrade performs better is straightforward – it takes the full amount of information on the input, i.e., the whole underlying histograms of X and Y, while traditional simulation only works with partial sums, i.e., generated draws of X and Y.

Since the underlying histograms of X and Y are only discrete approximations of continuous densities, the results of the above analysis may vary if the way histograms are constructed varies; however, the effect of binning was not studied here.

---

10 The illustrative example with normally distributed variables was selected intentionally so that the reader can easily compare the simulation results with the analytic solution. However, implications from this example can be generalized and hold across any distributions considered.

11 It is possible to convolve X and Y in the original domain, x, according to

$$f(x) * g(x) = \sum_m f(m) g(x - m),$$

which gives the same result as a convolution via FFT. Nevertheless, this approach is much slower compared to FFT – even slower than plain simulation in many cases, depending on variable binning.

5.2 Time effectiveness of the FFT approach

The only question remaining is to state when it makes sense to proceed with FFT after a simulation and when it does not generate any value added. According to Figure 1 and Figure 2, simulating 100 times and applying FFT afterwards makes perfect sense because it is improbable (in real applications, not in this simple one) that FFT would take more time than carrying out the remaining 1,900 iterations, provided that we know in advance that 2,000 iterations make FFT redundant (Figure 1 situation).

As a real example, I consider a pair of VAR models with a maximum lag length set to 3. Not to speak of variables involved, both of these models are in turns processed through a bootstrapping sequence in order to provide simulated predictive densities of the variable of interest. Such a simulation batch is repeated for a variable number of simulation iterations, and the computational times necessary to accomplish each simulation cycle are stored. At the end, the simulation-based approach and the one with FFT extension are compared in terms of computational time and degree of precision (measured by mean absolute deviation of both simulation variants from true distribution). Results of this analysis are presented in Figure 3.

Figure 3
(A) – Deviation of the simulated model from the true model as a function
   of algorithmic iterations;
(B) – Associated computational time. Source: own calculation
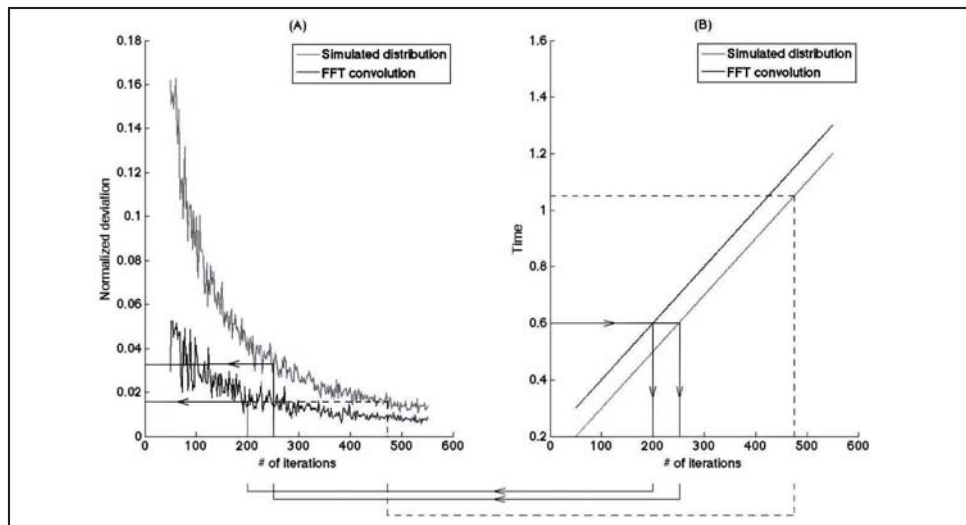
Graph (A) shows how the overall preciseness of both simulation approaches depends on the number of iterations conducted during bootstrapping (running from 50, going by one up to 550 iterations). It can be seen that whenever the FFT approach is performed after a regular simulation batch, the deviation of the results from reality is more or less than twice as small (which doubles the accuracy), especially in the case of a lower number of executed iterations.

Of course, additional use of the FFT algorithm is costly in terms of time – Graph (B) shows how long it takes to carry out a certain number of simulation iterations, regardless of whether the time units on the vertical axis are in seconds, hours, or days. It took me 0.004 seconds on average to carry out one estimation cycle needed in the residual bootstrap with the VAR models under consideration. Therefore, the dependence of the computational time on the number of iterations appears to be linear (0.004 being the slope of the line). Processing the result of bootstrapping through FFT takes always the same amount of time, no matter how accurate bootstrapped predictive distributions are taken as an input – consequently, the line expressing the time dependency in the case of FFT in Graph (B) has the same slope, it is only vertically shifted upwards.

While linear relationships were intuitively suggested in the case of Graph (B), the same cannot be deduced from Graph (A). Not only can a rather exponential pattern be observed but also the curves in Graph (A) are not smooth, which emanates from the fact that bootstrapping incorporates a stochastic behavior and therefore high-frequency fluctuations around otherwise clearly visible trends are present.

To close the discussion of this example, if, for instance, the time allowed for the computation is 0.6 units of time (follow solid lines with arrows in Figure 3), then either 200 iterations followed by FFT or, 250 iterations of plain simulation can be carried out. Furthermore, these numbers imply the accuracy of the results – following the ticks on the vertical axis in Graph (A), the efficiency of the FFT approach is twice as high compared to plain simulation.

Going in the reverse direction and fixing the accuracy of the results (follow the dashed line this time), some 480 iterations of plain simulation are necessary to get the same result as with 200 iterations followed by FFT. Tracing the dashed line back to Graph (B) implies that in order to achieve equal effectiveness of both simulation approaches, the time needed for plain simulation is almost double compared to the FFT approach.

5.3 Is computation of combined densities desirable?

A natural caveat that could arise is why to care about the whole predictive densities when a combination of any densities should, due to the central limit theorem, result in an asymptotic normal distribution.[12] The bootstrapping sequence of a well-specified model usually results in symmetric predictive densities, with the exception of models that incorporate zero bounds on certain variables, e.g., interest and unemployment rates. In general, if predictive densities or parameter estimates are not symmetric, something must have gone wrong. Thus, predictive densities should be in some sense normally distributed – but what matters is the standard deviation, which is commonly better captured by simulation experiments while normal approximations tend to underestimate standard deviations. Simulation certainly generates value added.

---

12 I would like to thank an anonymous referee for pointing this out.

5.4 FFT drawbacks

For the FFT algorithm to work efficiently, it is necessary to provide input vectors of a length that is an integer power of two (2, 4, 8, 16, 32, 64, …). This can be overcome by padding the input vectors with zeros when needed. Another disadvantage is that FFT can convolve two vectors according to $w_1 X + w_2 Y$ only for integer weights. If for example $w = (1/3; 2/3)$ then FFT must be instructed to take the weights so that the result for $1 \cdot X + 2 \cdot Y$ is computed and then a horizontal adjustment is needed in order to handle the fact that the weights do not sum to one any more. Also, it can happen easily that a convolution of vectors can lie far beyond the length of the input vectors – in that case, the convolved distribution spills over the whole domain without any warning (so-called wrap-around). One must arrange for sufficient length of input vectors in order to avoid this case.

6. Results

Following the procedure described above, I constructed a series of *ex post* predictions for the Czech GDP based on a combination of 159 representative models running from 2007 and ending in the second quarter of 2009, from where an *ex ante* combined prediction was computed. Figure 4 shows the medians of relevant combined distributions for 1-step to 6-step ahead predictions from each forecast horizon considered. Figure 5 shows the *ex ante* prediction including confidence bands.
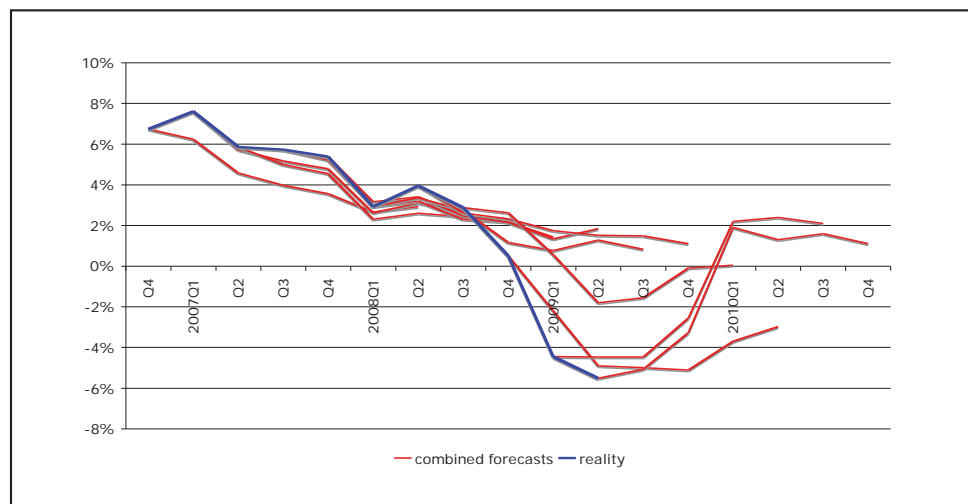
Figure 4
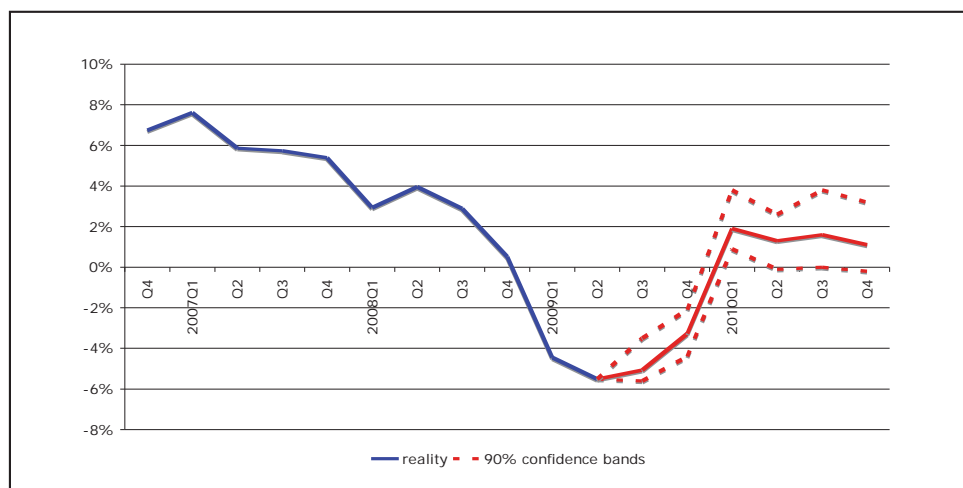**Prediction of YoY GDP growth**

Figure 5
Prediction of YoY GDP growth

7. Conclusion

Combining forecast densities that come from a broad class of models was demonstrated on two auxiliary theoretical examples and then applied to a prediction of the Czech GDP. Combining densities in a Bayesian style is computationally intensive in practice. Therefore, a reasonable reduction in model space was suggested in order to keep simulation procedures manageable. A suboptimal model suite was selected with the help of the Metropolis algorithm. Poorly simulated models were then combined with fast Fourier transform via an expert linear opinion pool.

The FFT algorithm proved to help significantly reduce the computational time with hardly any loss in the quality of final results – on the contrary, the theoretical part of this paper offers empirical evidence that FFT exceeds brute force simulation in all aspects. An example with two VAR models, demonstrated in this paper, illustrated both the low computational demands of FFT and its high degree of efficiency compared to the plain simulation technique. The combining of two experimental VAR models was carried out in half the time that would otherwise be necessary to meet an equal quality of results. Also, the convenient features of FFT are even more impressive if one works with multidimensional models – more equations and longer lag lengths in VAR models imply many parameters to be estimated and, thus, any bootstrapping-type technique slows down the computation severely, which brings the FFT approach to an even brighter light.

Speaking of which, it took almost 12 hours of computational time for the suggested combined GDP model to get processed through brute force simulation in a Bayesian manner. When FFT was applied in relevant phases of the estimation, the computational time was cut down tremendously to only 2 hours 45 minutes without any visible loss being encountered.

The combined *ex ante* prediction of the Czech GDP suggests that the Czech economy has touched its trough in the third quarter of 2009. The relatively high GDP growth in 2010Q1, implied by the combined model, results mainly from the sharp GDP decline in the same quarter of the preceding year. I would not interpret this pattern of predicted GDP as a W-shaped recovery.

VAR models usually serve as a convenient framework to trace out responses in endogenous variables on artificially generated shocks (impulse-response analysis). GDP in combined models has non-constant variability across models and, what is more challenging, many models considered in the analysis here contain isolated GDP components instead of GDP. Combined impulse-response functions would need to compare variables with somehow normalized variability and take into account the way non-stationarity was handled or perhaps the cross-correlation between GDP and modeled variables. This sounds like a tedious task but on the other hand, some puzzles, e.g., the price puzzle, could perhaps be solved in this way.

Appendix

The Metropolis algorithm helped to cut down the number of models considered from 1,408 to 159. Table 1 provides an overview of what model types are included in the representative model set.

Table 1

Suite of representative models

		# of selected models	Total models
Maximum lag length	1	39	159
	2	42	
	3	46	
	4	32	
Estimation	VAR	89	159
	BVAR	70	
Stationarity	Differenced	57	159
	Band-pass filtered	102	
# of equations	1	18	159
	2	50	
	3	72	
	4	19	
Endogenous variable of interest**	GDP (C+I+G+NX)	4	159
	Turnover in international trade (X+M)	12	
	Export (X)	39	
	Import (M)	27	
	Investment (I)	6	
	Consumption (C+G)	22	
	GDP (BP filtered)	49	

**) Recall that variables other than GDP werw transformed to GDP through aforementioned coinegrating ralationships.

References

ANDERSSON, M.; KARLSSON, S. 2007. Bayesian Forecast Combination for VAR Models [Working Paper No. 216]. Sveriges Riksbank, 2007.

ASMAR, N. H. 2005. *Partial Differential Equations with Fourier Series and Boundary Value Problems*. 2nd ed. Upper Saddle River, NJ : Prentice Hall, 2005.

BACHE, I.; JORE, A.; MITCHELL, J.; VAHEY, S. 2009. Combining VAR and DSGE Forecast Densities. Norges Bank, 2009.

BATES, J.; GRANGER, W. 1969. The Combination of Forecasts. *Operational Research Quarterly.* 1969, vol. 20, s. 451–468.

BAXTER, M.; KING, R. 1995. Measuring Business Cycles Approximate Band-Pass Filters for Economic Time Series [Working Paper No. 212]. NBER, 1995.

CANOVA, F. 2007. *Methods for Applied Macroeconomic Research*. Princeton, NJ : Princeton University Press, 2007.

CHRISTIANO, L.; FITZGERALD, T. 1999. The Band Pass Filter [Working Paper No. W7257]. NBER, 1999.

DOAN, T.; LITTERMAN, R.; SIMS, C. 1984. Forecasting and Conditional Projection Using Realistic Prior Distribution. *Econometric Reviews*. 1984, vol. 3, no. 1, s. 1–100.

ENDERS, W. 2004. *Applied Econometric Time Series*. Hoboken, NJ : Wiley&Sons, 2004.

EFRON, B. 1979. Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*. 1979, vol. 7, s. 1–26.

GERARD, H.; NIMARK, K. 2008. Combining Multivariate Density Forecasts Using Predictive Criteria [RBA Research Discussion Papers]. Reserve Bank of Australia, 2008.

HAMILTON, J. 1994. *Time Series Analysis*. Princeton, NJ : Princeton University Press, 1994.

HOETING, J.; MADIGAN, D.; RAFTERY, A.; VOLINSKY, C. 1999. Bayesian Model Averaging: A Tutorial. *Statistical Science*. 1999, vol. 14, no. 4, s. 382–417.

KASCHA, C.; RAVAZZOLO, F. 2009. Combining Inflation Density Forecasts. Norges Bank, 2009.

LANCASTER, T. 2008. *An Introduction to Modern Bayesian Econometrics*. Oxford : Blackwell Publishing, 2008.

LITTERMAN, R. 1986. Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business and Economic Statistics*. 1986, vol. 4, no. 1, s. 25–38.

LÜTKEPOHL, H. 2007. *New Introduction to Multiple Time Series Analysis*. Berlin : Springer-Verlag, 2007.

SALA-I-MARTIN, X.; DOPPELHOFER, G.; MILLER, R. 2004. Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review*. 2004, vol. 94, no. 4, s. 813–835.

STOCK, J.; WATSON, M. 2004. Combination Forecasts of Output Growth in a Seven-Country Data Set. *Journal of Forecasting*. 2004, vol. 23, s. 405–430.

COMBINING VAR FORECAST DENSITIES USING FAST FOURIER TRANSFORM

Abstract: In this paper, I propose the use of fast Fourier transform (FFT) as a convenient tool for combining forecast densities of vector autoregressive models in a hybrid Bayesian manner. While a vast amount of papers comprises combinations based on normal approximations, Monte Carlo methods were fully utilized here, which made the analysis computationally demanding. For the sake of minimization of computational time, the FFT algorithm was used to combine the densities of poorly simulated partial models. As a result, a minor loss of quality in the final combined model was allowed, in contrast with the reduction in the necessary simulation time. However, it turns out in the end that the FFT-based approach exceeds 'brute-force' simulation in all aspects. The suggested method is demonstrated on an ex ante prediction of the Czech GDP and on a pair of artificial examples.

Keywords: Bayesian model averaging, fast Fourier transform, Markov chain Monte Carlo, vector autoregressions

JEL Classification: C11, C30, C53, E17