Short-term Forecasting and Composite Indicators Construction with Help of Dynamic Factor Models Handling Mixed Frequencies Data with Ragged Edges

Marcin Łupiński

National Bank of Poland, Department of Statistics, Swietokrzyska 19/21 Warsaw, University of Warsaw, Faculty of Economic Sciences, Dluga 44/50 Warsaw

---

Abstract

This paper aims at presenting practical applications of latent variable extraction method based on second generation dynamic factor models, which use modified Kalman filter and Maximum Likehood Method and can be applied for time series with mixed frequencies (mainly monthly and quarterly) and unbalanced beginning and the end of the data sample (ragged edges). These applications embrace nowcasting, short-term forecasting of Polish GDP and construction of composite coincident indicator of economic activity in Poland. Presented approach adopts the idea of short-term forecasting used by Camacho and Perez-Quirioz in Banco de Espana and concept of Arouba, Diebold and Scotti index compiled in the FRB of Philadelphia. According to the author's knowledge, it is the first such adaptation for Central and Eastern Europe country. Quality of the nowcasts/forecasts obtained with these models are compared with standard methods used for short-term forecasting with series of statistical tests in the pseudo real-time forecasting exercise. Moreover described method is applied for construction of composite coincident indicator of economic activity in Polish economy. This newly-created coincident indicator is compared with first generation coincident indicator, based on standard dynamic factor model (Stock and Watson) approach, which has been computed by the author for Polish economy since 2006.

Keywords: short-term forecasting, coincident indicators, factor models, mixed frequencies, ragged edges
JEL Classification: C22, C53

Email address: `marcin.lupinski@nbp.pl` (Marcin Łupiński)

Contents

1. Introduction

Since construction of the first set of composite economic indicators by Burns and Mitchell in the first half of 20th century these indicators have proved crucial in determining and explaining current and future economic situation of particular economies and international organizations. Decision makers and agents observe them to have clear synthetic picture of the state of the economy and to derive its future perspectives. Broad group of the most popular coincident indicators prepared by international institutions such as The Conference Board or OECD is based on heuristic methods examining common properties of time series with help of simple measures like cross-correlation or coherence. Simultaneously there is constantly growing number of institutions (CEPR, Banco de Espana) which use dynamic factor econometric framework proposed by Stock and Watson [8](S&W, 1989) or its generalized version prepared by Forni, Hallin, Lippi and Reichlin [7](FHLR, 2001). S&W coincident indicator is estimated in the time domain on a small set of selected time series, on the other hand FHLR model operates on the broad cross-section of economic data (up to few thousands of series, which gives possibility of exploitation of the model asymptotic features) and use spectral representation during estimation process. Both of them are based on strict mathematical structure of the model, which removes bias of arbitrary judgment of economic conditions by experts, the main disadvantage of heuristic methods. However similarly to expert approach S&W and FHLR frameworks assume that time series, which are coincident indicators' components, should have the same length and the same frequency. This feature generates serious problems for analysts trying to use constructed coincident indicators for their practical application like nowcasting and forecasting because as they are not able to enclose recent information due to delays of some time series publication (e.g. for Polish GDP this delay lasts up to 60 days after the end of reference period). Moreover, according to the paper of Boivin and Ng [3] (2006) and surveys performed by Lupinski [15] (2006) generalized dynamic factor models do not outperform their smaller counterparts, although they have quite complex structures which are not so easy to understand by analyst, who are not interested with their technical interior.

Dealing with dynamic factor models' drawbacks directed some researchers like Shumway and Stoffer [19](1982) towards some accompanying algorithms (in their case Expectation Maximization /EM/ algorithm) to replace missing observations with their optimal extrapolations. Unfortunately incorporating these algorithms into base econometric state pace models generated often additional problems like those connected with identification of matrices linking latent components with observed variables. Alternative approach was proposed by Mariano and Murasawa [10](2003) who directly incorporated missing data and mixed frequency handling into their econometric framework. They used properties of state space representation and Maximum Likehood Estimation (MLE) with missing observations replaced with draws from selected probability distribution (most often from Gaussian one). Mariano and Murasawa model inherited all advantages of dynamic factor models, beside that it gave transparent way of coincident indicator estimation based on well-established mathematical background. Their framework was applied in practice by Camacho and Perez-Quirios [6] (2009) to build coincident indicator used by them to compute nowcasts and forecasts of Euro Area GDP and validate its power in the pseudo real-time exercise. Similar indicator for American Economy was proposed by Arouba, Diebold and Scotti [1](2009). The small scale dynamic factor model with missing data and mixed frequencies handling proposed by Mariano and Murasawa is applied in this paper for estimation of new coincident indicator of Polish GDP and evaluation of its statistical features. Achieved results are compared with Stock and Watson-style coincident indicator that has been computed by the author for Polish economy since 2006. It is worth to mention that before presented survey was started, complete estimation framework was built in MATLAB, which was used to semi-automatize procedure of coincident indicator building and its nowcasting and forecasting power validation. It is also first survey known to the author, which uses real-time database of Polish time series applied to do historical analysis of nowcasting and forecasting performance with coincident indicators applied.

This article is organized as follows. The next section outlines common mathematical background used by dynamic factor models. In the third section second generation dynamic models (with ragged edges and mixed frequencies handling) are presented. Fourth section is devoted to description of economic data and procedures used in the coincident indicator estimation process. Description of recursive procedure of coincident indicators estimation and validation is enclosed in the fifth section. Sixth section was prepared for presentation of achieved results and their interpretation. Last section concludes the survey.

2. Dynamic factor models - mathematical background

In this section mathematical background of econometric models used in the article is presented. Investigation of these model's structures and mechanisms applied for their estimation is crucial for understanding their econometric characteristics as well as their advantages and drawbacks. For example issues connected with model's identification implicate problems while modeling variables with strong stochastic trend, which are the subject of interest in this article. At the beginning general idea of dynamic models with unobserved components is presented.

Let's assume that there is a time sequence (of length T) of vector's elements (of size n)

$$\tilde{Y} = [\tilde{Y}_1, \tilde{Y}_2, ..., \tilde{Y}_T] \tag{1}$$

which groups set of main macroeconomic variables describing current state of the economy. These variables' observations can be perceived as realizations of multivariate stochastic process Y, which will be modelled with proposed framework. Let's further assume that this stochastic process is influenced by unobserved (latent) variables. The goal of the framework is to estimate values of these variables. However because of their latent character it is necessary to go beyond traditional econometric framework. One of the most popular ways of dealing with this issue is to write dependencies determined by the model in the state space representations. This representation consists of two blocks of equations:

- Measurement block describing relations between unobserved components and sequences of stochastic processes' realizations (elements of the matrix Y)
- Transition block, which determines dynamics of latent variables in the form of difference equations' set

Simple notation of dynamic model with r unobserved components and exogenous economic variables (X, of size kxT) written in the state space form is presented below:

$$\begin{aligned} y_t &= Ax_t + Hh_t + w_t \\ h_t &= Fh_{t-1} + v_t \end{aligned} \tag{2}$$

where A, H, F are (nxk), (nxr) and (rxr) matrices respectively, w_t and v_t represent i.i.d errors drawn from Gaussian distributions with covariance matrices R and Q respectively ($E[w_t, v_\tau] = 0$ for each t, τ=1,2,...,T).

For example popular Box and Jenkins's ARIMA framework augmented with unobserved component, used by Stock and Watson [18](1998) to decompose US GDP into permanent and transitory part:

$$\begin{cases} y_t = y_{1,t} + y_{2,t} \\ y_{1,t} = \delta + y_{1,t-1} + w_t \\ y_{2,t} = \phi_1 y_{2,t-1} + \phi_2 y_{2,t-2} + v_t \end{cases} \tag{3}$$

where $w_t \backsim i.i.d\, N(0, \sigma_{w,t}), v_t \backsim i.i.d\, N(0, \sigma_{v,t}),\ E[w_t, v_\tau] = 0$ for t,τ=1,2,...,T, can be noted in the state space form as follows:

$$\begin{array}{l} y_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{2,t-1} \end{bmatrix} \\ \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ v_t \\ 0 \end{bmatrix} \end{array} \tag{4}$$

Let's go back to the general model described with the set of equations (2). The goal of an econometrician is to estimate values of unobserved variables grouped in the vector h_t in all periods t = 1,2,. . . ,T, having at disposal appropriate set of information (data). The estimation procedure can be started once values of all model's parameters (elements of the matrices A, H, F, R, Q) are known. Unfortunately they are not available at the beginning of the modeling process. However an econometrician can determine them using Maximum Likehood Estimation (MLE) method. Description of the MLE's application is presented further into this paper, at this stage assumption is made that all parameters of the model are available. The procedure used for estimating optimal values of unobserved variables in dynamic models with known initial parameters was proposed originally by Kalman in 1960. His recursive procedure yields estimates with lowest Mean Squared Error (MSE).

During each stage of Kalman's procedure (indexed with time subscript t) conditional expectations of unobserved components observations vector are formulated:

$$h_{t|t-1} = E[h_t \mid z_{t-1}] \tag{5}$$

having at disposal information set $z_{t-1} = (y_{t-1}, y_{t-2}, ..., y_1, x_t, x_{t-1}, x_{t-2}, ...,, x_1)$ / x_t is exogenous variable therefore it does not affect values of $h_{t+s}, s = 0, 1, ..., T-t/$. Conditional expectation of h_t are associated with matrix of estimation squared errors:

$$P_{t|t-1} = E[(h_t - h_{t|t-1})(h_t - h_{t|t-1})'] \tag{6}$$

At the beginning of the estimation process no value of h_t is available. Initial unobserved component vector ($h_{0|0}$) is derived from transition block of set (2) after applying unconditional expectation operator to both sides of this expression and assuming that h_t process is stationary which means that its unconditional expectation is constant across time ($E[h_t] = E[h_{t-1}]$)

$$(F - I)E[h_t] = 0 \Rightarrow h_{0|0} = 0 \tag{7}$$

Associated initial value of estimation error is the solution of equation describing h_t's covariance

$$E[h_t h_t'] = E[(Fh_{t-1} + v_{t-1})(h_{t-1}'F' + v_{t-1}')] = FE[h_{t-1}h_{t-1}']F' + E[v_{t-1}v_{t-1}'] \tag{8}$$

with applied stationary covariance matrix condition ($E[h_t h_t'] = E[h_{t-1}h_{t-1}'] = \Sigma$):

$$P_{1|0} = vec(\Sigma) = (I - F \otimes F)^{-1} vec(Q) \tag{9}$$

where $\otimes$ is Kronecker product of two matrices and vec vectorization operator.

After obtaining initial values of h_t and P_t during each iteration of the Kalman's procedure one step ahead conditional prediction of h_t is computed. The following expression:

$$\hat{h}_{t|t-1} = E[h_t \mid z_{t-1}] = F\hat{h}_{t-1|t-1} \tag{10}$$

is used there to evaluate error associated with this prediction. Part of the derivation shown in (8) can be used to gain:

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q \tag{11}$$

Predicted value of unobserved component can be entered into measurement equation with conditional expectation operator applied to its both sides

$$\hat{y}_{t|t-1} = E[y_t \mid z_{t-1}] = Ax_t + H\hat{h}_{t|t-1} \tag{12}$$

to compute y_t's forecast. Observed variables' prediction error ($\xi_{t|t-1} = \tilde{y}_{t|t-1} - Ax_t - H\hat{h}_{t|t-1}$) has in this case the following covariance matrix:

$$\Xi = E[(\tilde{y}_t - \hat{y}_{t|t})(\tilde{y}_t - \hat{y}_{t|t})'] = HP_{t|t-1}H' + R \tag{13}$$

This part of the Kalman's procedure is called prediction phase

The last equation allows to incorporate present information about behavior of observed variables in the forecast of h_t:

$$\begin{aligned} h_{t|t} = E[h_t \mid z_t] &= \hat{h}_{t|t-1} + E[(h_t - \hat{h}_{t|t-1})(\tilde{y}_t - \hat{y}_{t|t-1})']E[(\tilde{y}_t - \hat{y}_{t|t-1})(\tilde{y}_t - \hat{y}_{t|t-1})']^{-1}(\tilde{y}_t - \hat{y}_{t|t-1}) = \\ &= \hat{h}_{t|t-1} + P_{t|t-1}H'(HP_{t|t-1}H' + R)^{-1}H(\tilde{y}_{t|t-1} - Ax_t - H\hat{h}_{t|t-1}) \end{aligned} \tag{14}$$

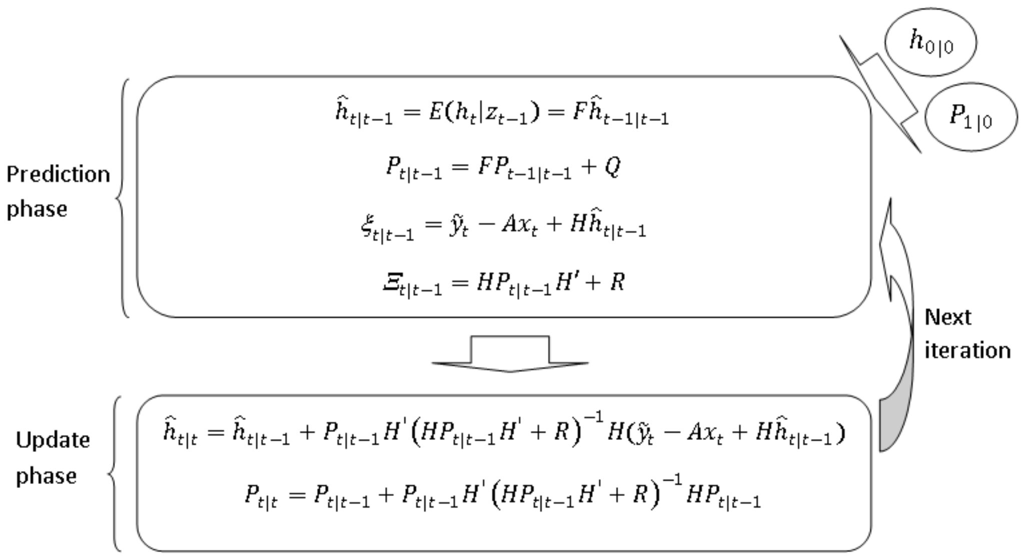
Estimation error's covariance matrix can be updated as well:

$$P_{t|t} = P_{t|t-1} + P_{t|t-1}H'(HP_{t|t-1}H' + R)^{-1}HP_{t|t-1} \tag{15}$$

Expression enclosed in the last two terms ($P_{t|t-1}H'(HP_{t|t-1}H' + R)^{-1}HP_{t|t-1}$) is of special interest because it allows to separate the effect of update process on forecast of h_t and P_t. It is called Kalman gain. Computing it completes update phase of Kalman filter and thus ends one iteration of the whole filtering procedure.

To recapitulate, one particular step of Kalman filter procedure, summarized with six equations, is presented below:

Figure 1.



Kalman filter offers optimal estimation of unobserved variables, taking into account that all structural parameters of the model are identified. As it was mentioned before parameters are unknown at the beginning of the estimation process. Hence MLE framework is used to get their estimates. In the second part of this section MLE framework customized to Kalman filter needs is presented.

Any econometric model specifies joint density function for set of input data ($\tilde{y}$) given m model parameters (θ):

$$p(\tilde{y} \mid \theta) \tag{16}$$

Because this parameters' set is not known, joint density function for parameters given available data (likehood function) is formulated:

$$L(\theta \mid \tilde{y}) \tag{17}$$

and then used to find values for which this function reaches its maximum:

$$\hat{\theta}_{MLE} = argmax(L(\theta \mid \tilde{y})) \tag{18}$$

For time series with independent observations it is easy to identify likehood function:

$$L(\theta \mid \tilde{y}) = \prod_{t=1}^{T} p(\tilde{y}_t \mid \theta) \tag{19}$$

where $p(\tilde{y}_t \mid \theta)$ is marginal density function for observation t.

In the case of dependent observations situation can get complicated as factors of (19) are replaced by conditional densities (conditional on information gathered in sequence $\tilde{y}_{t-1}, \tilde{y}_{t-2}, ..., \tilde{y}_1$)

$$L(\theta \mid \tilde{y}) = \prod_{t=2}^{T} p(\tilde{y}_t \mid \tilde{y}_{t-1}, \theta) p(\tilde{y}_1 \mid \theta) \tag{20}$$

which are often not so easy to identify. However in the case of models for which it can be assumed, that observations of input data are drawn from normal distribution. This problem can be solved by writing likehood function as a multivariate normal distribution and then applying triangular factorization to its covariance matrix. This allows to write it as a function of model prediction errors (ξ_t) and their covariances (Ξ_t) :

$$L(\theta \mid \tilde{y}) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^n det(\Xi_t)}} exp(-\frac{1}{2} \xi_t' \Xi_t^{-1} \xi_t) \tag{21}$$

which can be easily obtained from models estimated with Kalman filter because ξ_t and Ξ_t are computed as "byproducts" of this procedure. In practice natural logarithm of likehood function (which gives the same results because natural log is strictly monotonic function) is very often maximized instead of its original version. Computing logarithm of the function simplifies computations and allows to specify directly asymptotic covariance matrix of the parameters estimator $\hat{\theta}_{MLE}$. This matrix is determined by applying Cramer-Rao inequality, which states that difference between parameters estimator covariance matrix and inversion of information matrix $I(\theta)$ is positive semidefinite.

$$Cov(\theta) - I(\theta)^{-1} \tag{22}$$

Hence the inverse of information matrix is lower bound of covariance matrix. Information matrix is defined as negative expected value of Hessian of loglikehood function at $\hat{\theta}_{MLE}$:

$$I(\theta) = -E\left[\frac{\partial^2 lnL(\theta \mid \tilde{y})}{\partial\theta\partial\theta'}\right] \tag{23}$$

Maximum log-likehood estimator of vector has asymptotic normal distribution:

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta) \rightarrow N(0, \hat{H}^{-1}) \tag{24}$$

where $\hat{H} = lim\frac{1}{T}I(\theta)$, so it is consistent and asymptotically efficient. Hence, to compute covariance matrix of MLE estimator the following formula can be used:

$$\left[-\frac{\partial^2 lnL(\theta \mid \tilde{y})}{\partial\theta\partial\theta'}\right]^{-1} \tag{25}$$

at $\theta = \hat{\theta}_{MLE}$

Very often some additional constraints on chosen elements of likehood function parameters' vector are set. For example when stationary series is modeled with AR(2) process, it is necessary to assume that roots of lag polynomial $1 - L\psi_1 - L^2\psi_2 = 0$ associated with this process lie inside the unit circle, which yields:

$$s_1 = \frac{1}{1 + |\psi_1|},\ s_2 = \frac{1}{1 + |\psi_2|} \tag{26}$$

$$\psi_1' = s_1 + s_2,\ \psi_2' = -s_1 s_2 \tag{27}$$

Imposing this kind of constrains can be applied by computing appropriate continuous function $g(\cdot)$ subject to parameters achieved in the particular step of MLE procedure ($\theta = g(\psi)$). In such case constrained MLE optimization of θ is equivalent to unconstrained optimization of ψ:

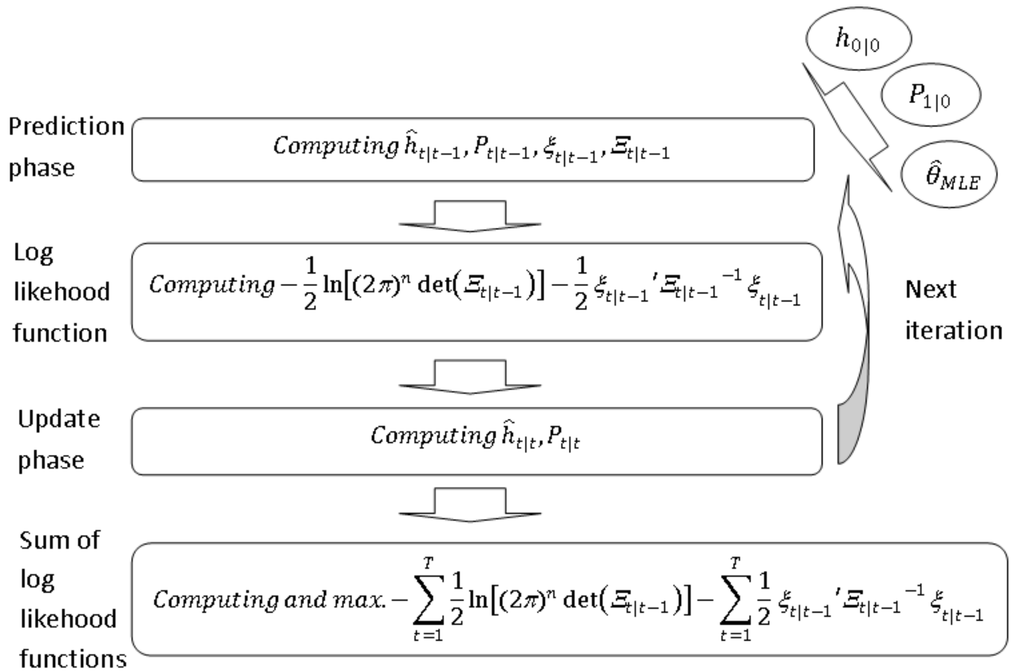
$$lnL(\theta \mid \tilde{y}) = lnL(g(\psi) \mid \tilde{y}) = lnL(\psi|\tilde{y}) \tag{28}$$

Although the formula of likehood function doesn't change, in the case of constrained optimization, computing MLE estimator error covariance matrix for $\hat{\theta} = g(\hat{\psi}_{MLE})$ requires quadratic form of the matrix (25) with gradient of the constraint function $g(\cdot)$:

$$\left[\frac{\partial g(\hat{\psi}_{MLE})}{\partial \hat{\psi}_{MLE}}\right] \left[-\frac{\partial^2 lnL(\hat{\psi}_{MLE} \mid \tilde{y})}{\partial \hat{\psi}_{MLE} \partial \hat{\psi}'_{MLE}}\right]^{-1} \left[\frac{\partial g(\hat{\psi}_{MLE})}{\partial \hat{\psi}_{MLE}}\right]' \tag{29}$$

Combination of model's parameters MLE procedure and Kalman filer estimation of unobserved variable is presented in the following figure.

Figure 2.



3. Applying second generation dynamic factor model for GDP Coincident Indicator construction and GDP nowcast and short-term forecast

Before survey described in this paper was initiated, the National Bank of Poland (NBP) computed a set of conincident indicators based on three schemes: The Conference Board/OECD, Stock and Watson and Forni Hallin Lippi and Reichlin. All above mentioned frameworks offer mechanisms, which allow to capture common dynamics of GDP and other key macroeconomic variables which influence current state of Polish economy. The first one is the simplest - based on heuristics. Second one introduces simple dynamic factor models which offers transparent dependencies between observed series and latent components so they could be easily understood by decision makers and analysts. The third one is the most sophisticated and demands frequency domain methods for estimation and large datasets to exploit its asymptotic features. However, as it was mentioned before, in empirical exercise it doesn't perform better than its simpler competitors. Having this situation in mind the NBP turned its attention to small dynamic factor models with possibility of frequency mixing and handling ragged edges which are relatively easy to estimate and give easily interpretable results.

The NBP's base model prepared in accordance with Stock and Watson approach excludes possibility of mixing frequencies and endogenous handling of input time series' missing observations.

Moreover it was designed to model growth of explanatory variables (due to stationarity requirements) and assumes lack of correlation between common component (interpreted as coincident indicator) and idiosyncratic factors (factors associated with particular input time series). Structure of described model shows the block of equations below:

$$\begin{array}{l} y_t = \delta + \gamma h_t + i_t \\ \Phi(L)h_t = v_{h,t} \\ \Theta(L)i_t = v_{i,t} \\ \begin{bmatrix} v_{h,t} \\ v_{i,t} \end{bmatrix} \sim N\left(0 \quad , \begin{bmatrix} \sigma_h & 0 \\ 0 & \Sigma_i \end{bmatrix} \right) \end{array} \tag{30}$$

where h_t is common latent component, i_t vector of idiosyncratic components, γ vector of coefficients quantifying dependencies between latent variable(s) and input time series, $\Phi(L)$,$\Theta(L)$ lag polynomials of common and idiosyncratic components accordingly, σ_h common component's variance and Σ_i covariance matrix of idiosyncratic component.

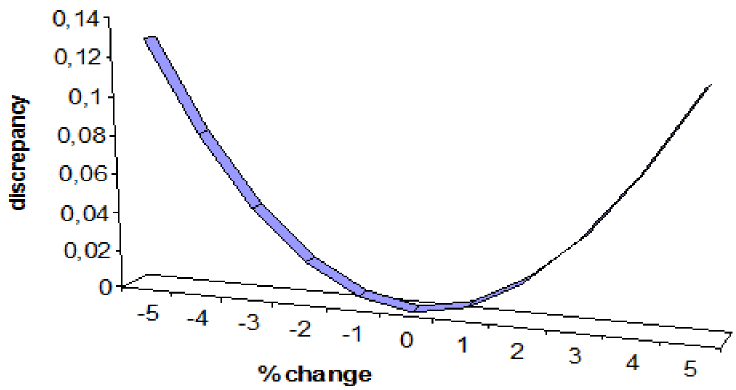
Second generation dynamic model inherits many features of its ascendant, but it allows to introduce mixed frequencies and deal with missing observation within and at the beginning and the end of the sample (ragged edges). First extension is based on a simple approximation of arithmetic average by geometric one and should be only applied to flow data. For monthly and quarterly data (observed month on month and quarter on quarter respectively) dependency between dlog of quarterly and monthly data is derived in the following way:

$$\begin{array}{l} \tilde{Y}_t = 3\frac{\tilde{X}_t + \tilde{X}_{t-1} + \tilde{X}_{t-2}}{3} \backsimeq 3(\tilde{X}_t \tilde{X}_{t-1} \tilde{X}_{t-2})^{1/3} \\ ln(\tilde{Y}_t) = ln(3) + \frac{1}{3}ln(\tilde{X})_t + \frac{1}{3}ln(\tilde{X}_{t-1}) + \frac{1}{3}ln(\tilde{X}_{t-2}) \\ ln(\tilde{Y}_t) - ln(\tilde{Y}_{t-3}) = \frac{1}{3}(ln\tilde{X}_t - \frac{1}{3}ln\tilde{X}_{t-3}) + \frac{1}{3}(ln\tilde{X}_{t-1} - \frac{1}{3}ln\tilde{X}_{t-4}) + ... \\ \tilde{y}_t^{\triangle q} = \frac{1}{3}\tilde{x}_t^{\triangle m} + \frac{2}{3}\tilde{x}_{t-1}^{\triangle m} + \tilde{x}_{t-2}^{\triangle m} + \frac{2}{3}\tilde{x}_{t-3}^{\triangle m} + \frac{1}{3}\tilde{x}_{t-4}^{\triangle m} \end{array} \tag{31}$$

where $\tilde{y}_t^{\triangle q} = \tilde{Y}_t - \tilde{Y}_{t-3}$, $\tilde{x}_t^{\triangle m} = \tilde{X}_t - \tilde{X}_{t-1}$

Error associated with proposed transformation is depicted below:

Figure 3.



Schema of the model with mixed frequencies and all data available in the sample has the structure of:

$$\begin{array}{l} \left[\begin{array}{c} y_{1,t} \\ y_{j,t} \end{array}\right] = \left[\begin{array}{c} \delta_1 \\ \delta_j \end{array}\right] + \left[\begin{array}{c} \gamma_1(\frac{1}{3}h_t + \frac{2}{3}h_{t-1} + h_{t-2} + \frac{2}{3}h_{t-3} + \frac{1}{3}h_{t-4}) \\ \gamma_j h_t \end{array}\right] + \left[\begin{array}{c} \frac{1}{3}i_{1,t} + \frac{2}{3}i_{1,t-1} + i_{1,t-2} + \frac{2}{3}i_{1,t-3} + \frac{1}{3}i_{1,t-4} \\ i_{j,t} \end{array}\right] \\ \Phi(L)h_t = v_{h,t} \\ \Theta(L)i_{j,t} = v_{i,j,t} \\ \left[\begin{array}{c} v_{h,t} \\ v_{i,j,t} \end{array}\right] \sim N\left(0 \quad , \left[\begin{array}{cc} \sigma_h & 0 \\ 0 & \Sigma_i \end{array}\right] \right) \end{array} \tag{32}$$

The simplest state space specification of the model with dynamics of h_t and $i_{j,1}$ represented with AR(1) processes (which in fact turns out to be quite robust during coincident index estimation) has the following form:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix} + \begin{bmatrix} \frac{1}{3}\gamma_1 & \frac{2}{3}\gamma_1 & \gamma_1 & \frac{2}{3}\gamma_1 & \frac{1}{3}\gamma_1 & \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \gamma_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \gamma_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \gamma_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_{t-1} \\ h_{t-2} \\ h_{t-3} \\ h_{t-4} \\ h_{t-5} \\ i_{1,t-1} \\ i_{1,t-2} \\ i_{1,t-3} \\ i_{1,t-4} \\ i_{1,t-5} \\ i_{2,t-1} \\ i_{3,t-1} \\ i_{4,t-1} \\ i_{5,t-1} \end{bmatrix} + \begin{bmatrix} v_{h,t} \\ 0 \\ 0 \\ 0 \\ 0 \\ v_{i,1,t} \\ 0 \\ 0 \\ 0 \\ 0 \\ v_{i,2,t} \\ v_{i,3,t} \\ v_{i,4,t} \\ v_{i,5,t} \end{bmatrix} \tag{33}$$

$$\begin{bmatrix} h_t \\ h_{t-1} \\ h_{t-2} \\ h_{t-3} \\ h_{t-4} \\ i_{1,t} \\ i_{1,t-1} \\ i_{1,t-2} \\ i_{1,t-3} \\ i_{1,t-4} \\ i_{2,t} \\ i_{3,t} \\ i_{4,t} \\ i_{5,t} \end{bmatrix} = \begin{bmatrix} \phi & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_4 \end{bmatrix} \begin{bmatrix} h_{t-1} \\ h_{t-2} \\ h_{t-3} \\ h_{t-4} \\ h_{t-5} \\ i_{1,t-1} \\ i_{1,t-2} \\ i_{1,t-3} \\ i_{1,t-4} \\ i_{1,t-5} \\ i_{2,t-1} \\ i_{3,t-1} \\ i_{4,t-1} \\ i_{5,t-1} \end{bmatrix} \tag{34}$$

For presented model input data set consists of one quarterly indicator ($y_{1,t}$) and four monthly indicators ($y_{j,t}, j = 2, 3, ..., 5$).

Second very useful feature of Mariano and Murasawa model is its embedded mechanism of handling missing observations. Solution of this problem is the same without distinction to gaps within sample and at the beginning or the end of time series' set. Observations which are not available are replaced with draws from Gaussian distribution with parameters equal to original variables estimators:

$$\overline{y}_{j,t} = \begin{cases} \tilde{y}_{j,t} \; if \; \tilde{y}_{j,t} \; is \; available \\ w_{j,t} \; if \; \tilde{y}_{j,t} \; is \; not \; available \end{cases} \tag{35}$$

where $w_{j,t} \backsim N(\mu, \sigma)$, μ, σ are estimators of the first two moments of original series.

Moreover missing data handling requires change of model specification (example for handling missing values for first variable):

$$
\begin{bmatrix} y_{1,t} \\ y_{j,t} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_j \end{bmatrix} + \begin{bmatrix} \gamma_1(\frac{1}{3}h_t + \frac{2}{3}h_{t-1} + h_{t-2} + \frac{2}{3}h_{t-3} + \frac{1}{3}h_{t-4}) \\ \gamma_j h_t \end{bmatrix} + \begin{bmatrix} \frac{1}{3}i_{1,t} + \frac{2}{3}i_{1,t-1} + i_{1,t-2} + \frac{2}{3}i_{1,t-3} + \frac{1}{3}i_{1,t-4} \\ i_{j,t} \end{bmatrix}
$$
$$
if\ \tilde{y}_{j,t} is\ available
$$
$$
\begin{bmatrix} y_{1,t} \\ y_{j,t} \end{bmatrix} = \begin{bmatrix} 0 \\ \delta_j \end{bmatrix} + \begin{bmatrix} 0 \\ \gamma_j h_t \end{bmatrix} + \begin{bmatrix} 0 \\ i_{j,t} \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ 0 \end{bmatrix} if\ \tilde{y}_{j,t} is\ not\ available
$$
$$
w_{1,t} = N(\delta_1, \sigma_1)
$$
$$
\ldots \tag{36}
$$

Estimation of dynamic factor models on input variables' growth is connected with identification issue. According to the last model first moment of $y_{j,t}$ ($E[y_{j,t}]$) consists of two parameters, given particular value of γ_j:

$$
E[y_{j,t}] = \delta_j + k * \gamma_j E[h_t] \tag{37}
$$

where k = 3 for j =1 and k=1 otherwise. Due to this feature $E[y_{j,t}]$ can be generated by unlimited number of δ_j and $E[h_t]$. To solve this problem each particular input variable is demeaned. Hence deviations of the unobserved factors from their means are modeled ($h_t^{wm} = h_t - E[h_t]$). This solution however brings the following question: how should this mean be restored when the factor with trend is needed? The answer can be found in the output of basic Kalman filter. It gives estimate of h_t as a linear projection of concurrent and lagged data entered into the model. The relationships between tem can be noted with lag polynomial:

$$
h_{t|t} = W(L)y_t \tag{38}
$$

$$
h_{t|t}^{wm} = W(L)y_t^{wm} \tag{39}
$$

Deviations from mean can be determined with help of steady state Kalman gain computed in practice as Kalman gain for $t \to T$:

$$
W(L) = (I - (I - KH)FL)^{-1}K \tag{40}
$$

Finally unobserved component forecast can be restored using expectations of (38):

$$
E[h_t] = W(L)E[y_t] \Rightarrow mean(h_t) = W(1)\tilde{y}_t \tag{41}
$$

This solution however has some drawbacks which should be considered while applying the model to real data. Using one estimation of latent component's mean doesn't allow to model trends changing across time (stochastic), which are very common macroeconomic time series.

4. Statistical data and procedures used in models' estimation.

Estimation of new coincident indicator of Polish GDP and validation of its nowcasting and forecasting power in a pseudo real-time exercise required creation of a database with contents allowing simulation of historical inflow of observations of component time series. Time span for series enclosed in this database was set from 1996:M1 to 2010:M5 (date of Polish GDP publication for 2010:Q1). Selection of this database contents was started from determining reference series. For this role GDP in average constant prices from 2000 was chosen. Moreover set of potential coincident indicator's components were gathered (all with monthly and quarterly frequency): hard data such as Building Permits, Export, Employment, Import, Industrial Production Index, Production of Cement/Coal, Registered Cars, Retail Trade Turnover Index and soft indicators from the group of industrial confidence indicators and consumer sentiment indicators. PMI and financial indicators such as 3M Interbank Interest Rate Spread, Broad Money (M3) and Index of Warsaw Stock Exchange (WIG) were included into potential set of indicator components as well. Main source of mentioned time series was OECD MEI database, some data came also from other sources as ECB and Eurostat statistical data warehouses.

All procedures used in this survey were written in Matlab. Stock and Watson and Mariano and Murasawa models' codes were implemented by the author. Chistiano and Fitzgerald filtering procedure was upgraded version of the code prepared by Kowal [14](2005), Bry-Boschan business cycle dating routines were taken from the webpage of Inklaar[12]. The author created also consistent framework which allowed him to semi-automatize substantial parts of pseudo real-time nowcasting, forecasting and dating exercise.

5. Building coincident indicators and evaluating their nowcasting, forecasting and business cycle dating accuracy

In this section the procedures of coincident indicators building and using them for nowcasting/forecasting GDP are presented. Procedures applied for indicators' creation are based on two frameworks described in previous section. The first indicator is computed with help of Stock and Watson (S&W) model and the second one with schema proposed by Mariano and Murasawa (M&M). Then both of them are used in pseudo real-time exercise to check their usefulness for nowcasting and forecasting. Finally possibility of business cycles dating is checked as well.

Procedure of coincident indicator building was initiated with time series selection. Dependencies among transformed time series (first differences of logs) were checked simultaneously in time (with cross-correlation) and frequency domain (with coherence). Beside statistical testing, economic importance of each potential indicator's component was taken into consideration. During this stage four indicators were chosen: Industrial Production Index (IPI), Retail Trade Turnover Index (RTTI), Export (EXP) and Import (IMP). This selection of time series allows to monitor process of GDP creation (IPI) and distribution (EXP, IMP) and take a look at demand side of Polish economy (RTTI). Moreover inclusion of two external trade time series gave possibility of quantifying influence of global economic situation on Polish small open economy.

In the second phase S&W and M&M frameworks were applied for estimation of coincident indexes. To simulate as realistic as possible situation faced by analysts trying to use these indexes in real time for assessment of current and future economic situation, computations were performed on growing datasets. Introduction of particular observations at the ends of each time series were done in consistence with Polish official statistical Data Release Calendar available in ECB Statistical Data Warehouse website (http://sdw.ecb.int). This observation by observation growing dataset was used during final computations of coincident indicators. Earlier, in the phase of model parameters estimation, datasets reflecting situation at the end of each year of pseudo real-time exercise time span were used. Constructing this pseudo real-time data panel was one of the most time and effort consuming part of the survey.

To be consistent with stationary assumption used by both analytical frameworks, all selected time series were checked for presence of unit roots with standard ADF and KPSS test. Results of the estimation were gathered in the next two tables.

Table 1.

ADF Unit Root Test						
		GDP	IPI	RTTI	IMP	EXP
Loglevels	Spec.	T, C, 2L	T, C, 1L	T, C, 2L	T, C, 0L	T, C, 1L
	Statistics	-1.427	-2.028	-0.858	-1.235	-2.007
	p-value	0.842	0.551	0,947	0.882	0.572
First diff.	Spec.	C, 1L	C, 0L	C, 1L	C, 0L	C, 0L
	Statistics	-4.013	-15.381	-13.264	-14.244	-16.237
	p-value	0.0002	0.0000	0.0000	0.0000	0.0000

Legend: T - trend, C-constant, XL - number of lags (X)

Source: own computations

Table 2.

KPSS Test						
		GDP	IPI	RTTI	IMP	EXP
Loglevels	Spec.	T,C,BW4	T,C,BW10	T,C,BW10	T,C,BW10	T,C,BW10
	LM stat.	0.174	0.232	0.201	0.245	0.319
	5% level	0.146	0.146	0.146	0.146	0.146
First diff.	Spec.	T,C,BW4	T,C,BW6	T,C,BW6	T,C,BW4	T,C,BW4
	Statistics	0.232	0.057	0.172	0.153	0.178
	p-value	0.463	0.463	0.463	0.463	0.463

Legend: T - trend, C-constant, BW -bandwidth

Source: own computations

Achieved results proved that proposed transformation assured stationarity of selected time series.

Estimation of coincident indicators on growing datasets leads to the situation when edge of time series panel is ragged. Moreover in the case of M&M model which embraces mixed time series frequencies some observations within time sample are missing as well.

The table below shows part of time series sample available for estimation on 20.04.2010

Table 3.

Date	IPI	RTTI	IMP	EXP	GDP
Oct 09	-0.08	-0.15	5.12	1.64	NA
Nov 09	4.16	0	-3.37	-0.97	NA
Dec 09	-3.21	0.45	-0.66	0.16	1.2
Jan 10	2.75	-1.50	7.65	-0.67	NA
Feb 10	1.49	-1.37	-2.96	0.08	NA
March 10	2.252	TBO	TBO	TBO	TBO
TBO - to be observed, NA - not available					

Source: OECD MEI

As it has been mentioned in the previous section Mariano and Murasawa prepared their model for this type of dataset. Gaps within and at both ends of dataset are replaced with numbers drawn from Gaussian distribution. If any observation is missing in the particular row of input data matrix matrices H and R used in model estimation are tailored as well. In the case of S&W model, which does not include such mechanism, this kind of gaps at the end of quarterly dataset were filled with values computed with Expectation Maximization (EM) algorithm.

After dealing with datasets' adjustment problem, both frameworks were ready for parameters estimation. On this stage likehood function maximization procedure with embedded Kalman filtering was employed. To identify proper structure of the models (determining number of lags in the equations modeling behavior of particular input variables and unobserved components) the set of models was estimated and information criteria (Akaike and Bayesian) were computed for each of them. Then models with lowest information criteria values were chosen for further investigation.

Two sets of parameters obtained for models selected in this way (for S&W and M&M models accordingly) were gathered in the following tables.

Table 4. Estimated parameters of selected S&W model

Name	Value	Standard Error
γ_1	1.58	0.2
γ_2	0.75	0.24
γ_3	2.64	0.46
γ_4	3.33	0.53
ϕ_{11}	-0.30	0.01
ϕ_{12}	0.27	0.16
θ_{11}	0.22	0.02
θ_{12}	-0.01	0.001
θ_{21}	-0.17	0.01
θ_{22}	0.31	0.13
θ_{31}	0.18	0.01
θ_{32}	-0.01	0.001
θ_{41}	0.22	0.01
θ_{42}	0.01	0.001
σ_1	1.06	0.06
σ_2	2.64	0.50
σ_3	7.64	1.66
σ_4	8.62	3.41

Log likehood: 485.31
AIC: -2.58
BIC: -2.54

Source: own computations

Table 5. Estimated parameters of selected M&M model

Name	Value	Standard Error
γ_1	0.59	0.13
γ_2	0.21	0.09
γ_3	0.37	0.08
γ_4	0.63	0.08
γ_5	0.76	0.09
ϕ	-0.35	0.09
θ_1	-0.85	0.07
θ_2	-0.25	0.08
θ_3	-0.33	0.08
θ_4	-0.45	0.08
θ_5	-0.36	0.11
σ_1	0.67	0.17
σ_2	0.91	0.1
σ_3	0.79	0.09
σ_4	0.40	0.1
σ_5	0.31	0.13

Log likehood: 312.69
AIC: -3.22
BIC: -3.19

Source: own computations

Estimated parameters determine intuitive dependencies between unobserved components (interpreted as coincident indicator for the S&W model or direct input to equation for coincident indicator computation for M&M framework) and input variables (positive $\gamma's$) and reflect reliable AR processes modeling common and idiosyncratic components (ϕ,$\theta's$ which lie inside unit circle due to applied transformations). Moreover achieved values of variances connected with each modelled variable can be perceived as intuitive (for S&W model these variances can be bigger than one as input variables were not standardized).

Computed parameters were used in the next phase for estimation of two coincident indicators, assessment of their nowcasting and forecasting power and checking their business cycles dating capability. First of these two indicators was build with the use of S&W framework (therefore it was named CISW), second one with M&M schema (named CIMM accordingly). Described phase was divided into loops. Each loop was performed for dataset available at the end of reference period and consisted of:

- Estimation of coincident indicators
- Computation of nowcasts and forecast
- Computation of nowcast and forecast errors
- Smoothing of coincident indicators
- Application of business cycle dating procedure

Estimation of coincident indicators were done with the use of filtering procedures of S&W and M&M described in the previous section. As a cut off date for each estimation last day of reference month (in the case of M&M model) or quarter was used. For M&M model one additional transformation were required to be applied because originally this framework produces monthly common component (indicator $CIMM^m$). To gain coincident indicator comparable with reference series and S&W version for each period t linear combinations of the model's output were computed:

$$CIMM_t^q = \gamma_1(\frac{1}{3}CIMM_t^m + \frac{2}{3}CIMM_{t-1}^m + CIMM_{t-2}^m + \frac{2}{3}CIMM_{t-3}^m + \frac{1}{3}CIMM_{t-4}^m) \qquad (42)$$

In the two charts below S&W and M&M transformed indicators are shown:

Figure 4.

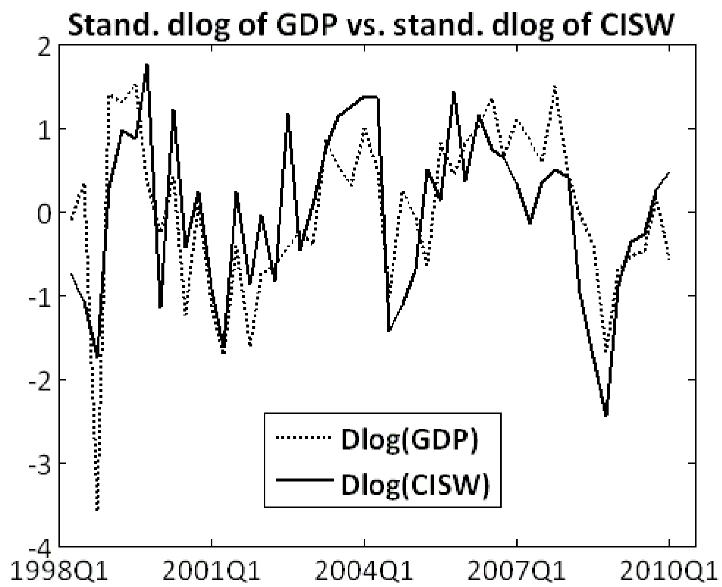
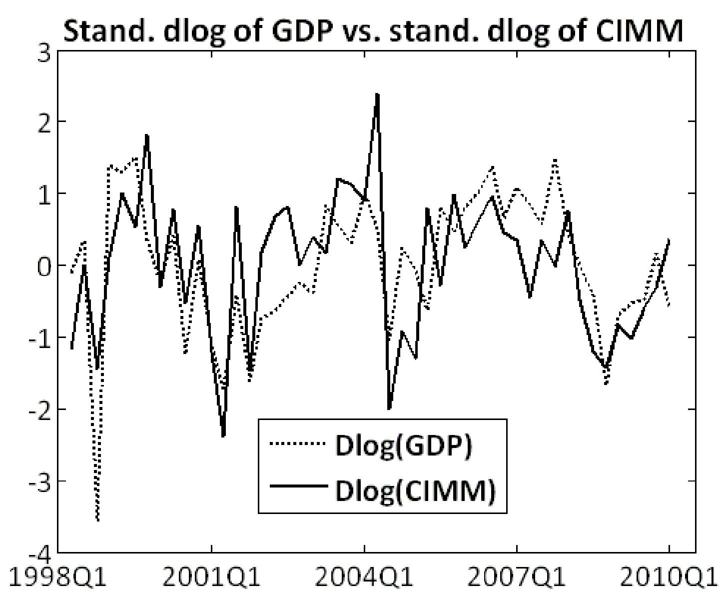


Figure 5.



As depicted above it seems that both coincident indicators share common oscillations with reference series, although they don't register huge decline in GDP growth at the beginning of the sample.

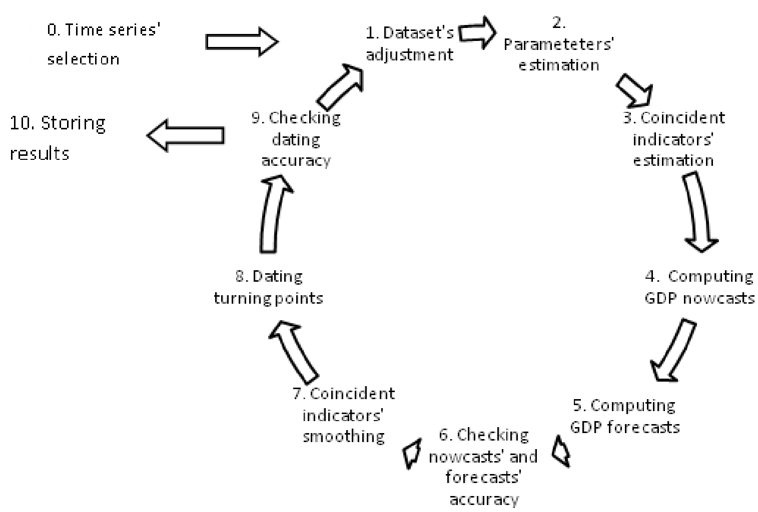
To check their usability as a forecasting variables both indicators were transformed to annual growths /quarter on previous year quarter (QoPYQ)/ and set as an input for the ARX model (autoregressive model with exogenous variables modeled with autoregressive process as well). This framework assumes that forecasted variable and coincident indicator follow an AR process:

$$y_{t+h} = \alpha_0 + \sum_{i=1}^{k} \alpha_i y_{t-i} + \sum_{j=0}^{l} \beta_j . x_{t-j} \tag{43}$$

where y is forecasted variable (GDP in this survey), x is exogenous variable included in the model (CISW and CIMM respectively), h is forecast horizon (h = 0 for nowcast), k is AR process order for forecasted variable and l is AR process order for coincident indicator. Achieved nowcasts and forecast were compared with naive autoregressive model estimated on the same datasets. The standard measures: RMSE, MAE and ME were used as indicators of nowcasting and forecasting power. Results of this exercise are discussed in detail in the next section. In the last phase f the survey computed coincident indicators and reference time series were smoothed with Christiano-Fitzgerald asymmetric filter to highlight their cyclical properties. Finally smoothed indicators and GDP were subject to standard dating algorithm proposed by Bry and Boschan [4] (1971). Similarly to nowcast and forecast results dates of expansions and recessions are reported in the next section.

To recapitulate schema of general coincident indicators building and evaluating procedure is presented:

Figure 6.



6. Evaluation of nowcasting and forecasting power

In this section results of the pseudo real-time nowcasting and forecasting exercise are presented. This exercise was performed for last 5 years of the sample, and the forecast horizon was set up to 4 quarters. Therefore the model was estimated for step-by-step growing sample starting on 1st quarter 2004 (March 2003 for MM model) and ending on the 1st quarter 2009 (March 2009). This span of time guaranteed that characteristics of the coincident indicator's nowcasting and forecasting power were checked during expansion and boom phase and in slowdown and recession period as well. Table 6 presents error statistics computed for nowcasting.

Table 6. GDP nowcast errors

Model	Rank (due to RMSE)	RMSE	MAE	ME
AR	3	1.123	0.961	0.365
CIMM	1	**0.393**	**0.291**	-0.074
CISW	2	0.545	0.424	**-0.029**

Source: own computations

In the nowcast competition CIMM indicator performed better than naive model and its simpler counterpart. In terms of RMSE statistics it was nearly 65% better than naive model and 27% better when compared with S&W indicator. It proves intuitive guess that incorporating fresh monthly data at the end of the sample allowed to react quicker to any changes in the current situation of Polish economy.

During next stage of the survey forecasting power of both indicator was evaluated. Forecasts for each horizon (h=1, 2, 3, 4) were computed separately and then appropriate average statistics were computed.

Table 7 reports gained results:

Table 7. GDP h-periods ahead forecast errors

Model	Rank (due to RMSE)	RMSE	MAE	ME
		h = 1		
AR	3	1.257	1.089	0.413
CIMM	**1**	**0.861**	**0.694**	**0.092**
CISW	2	1.169	1.067	0.528
		h=2		
AR	3	1.944	1.654	0.658
CIMM	**1**	**0.951**	**0.796**	**0.147**
CISW	2	1.207	1.014	0.364
		h=3		
AR	3	2.696	2.301	1.038
CIMM	1	**1.485**	**1.259**	**0.370**
CISW	2	1.608	1.410	0.461
		h=4		
AR	3	2.851	2.560	1.224
CIMM	**1**	**2.033**	**1.805**	**0.896**
CISW	2	2.285	2.153	1.119

Source: own computations

In a manner analogous to nowcasting part of the exercise second generation dynamic factor indicator allows to formulate forecasts which are the best approximation of Polish GDP. For all verified horizons it offers minimal error associated with forecast (with over 50% reduction in comparison with naive model for 2 quarters horizon in terms of RMSE). Despite the fact that its advantage diminishes across the time, even in the perspective of 4 quarters it is still by more than 25% better than the AR model. CISW also allows to reduce forecast error, but it is not so effective as its first generation competitor. Hence, it can be stated that although both coincident indicators are helpful in forecasting process, M&M index brings more useful information about future development of economic environment in Poland.

Three figures below present evolution of nowcast and forecast errors with respect to particular horizon index.

Figure 7.

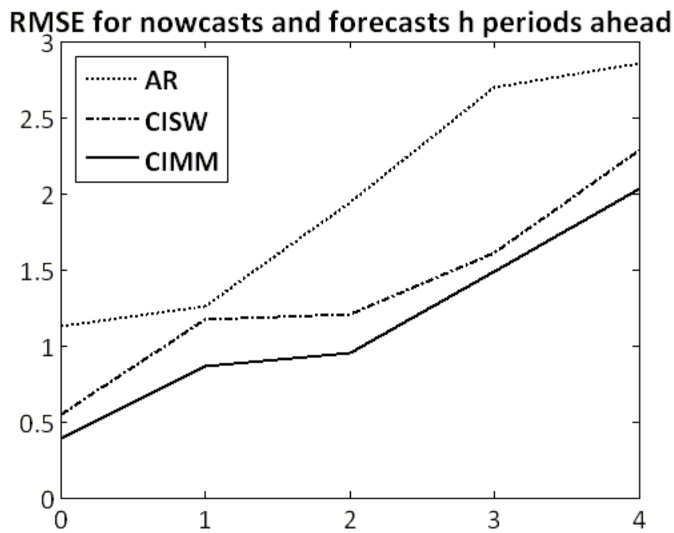


Figure 8.

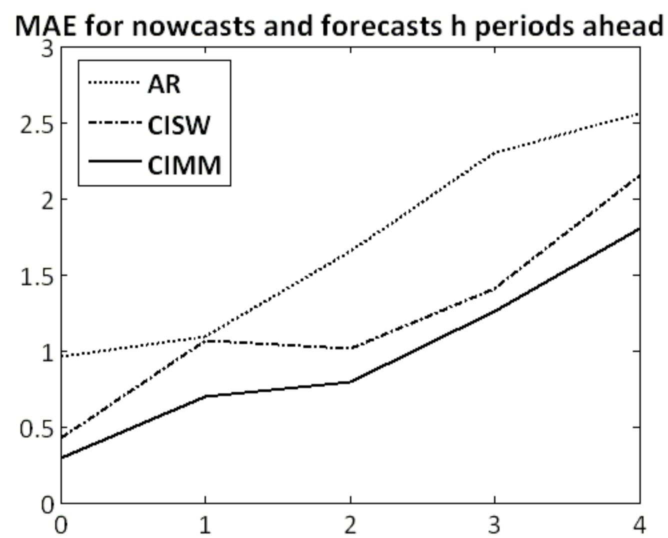
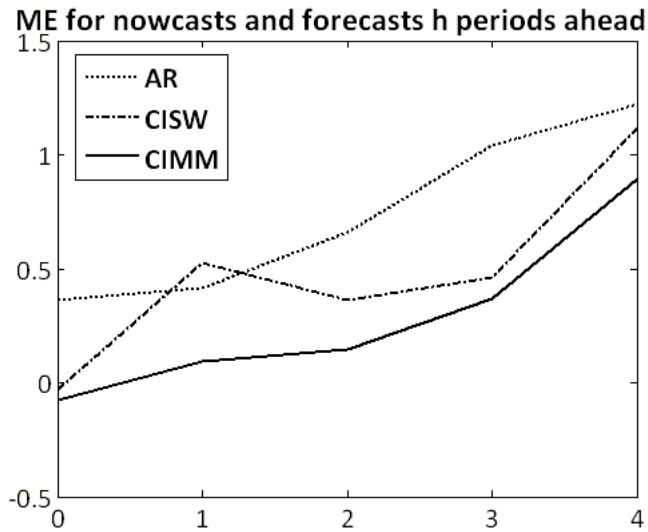


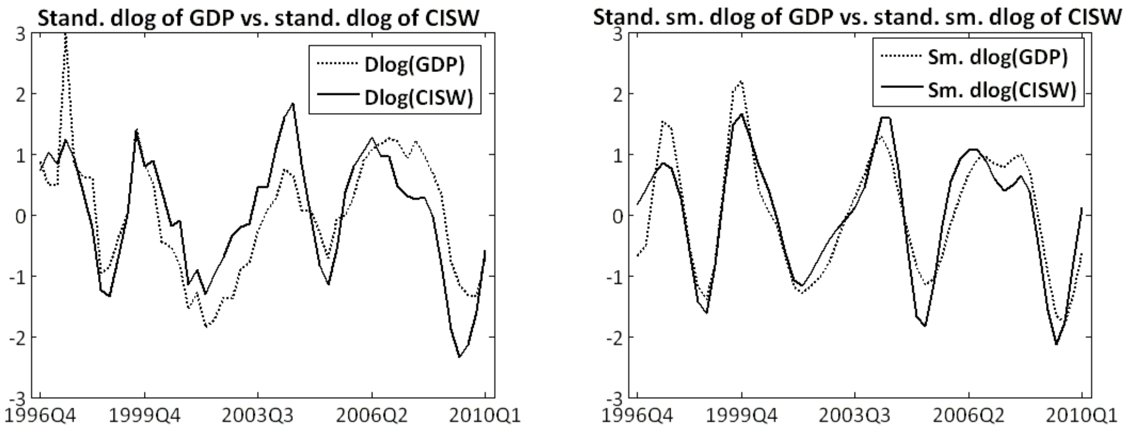
Figure 9.



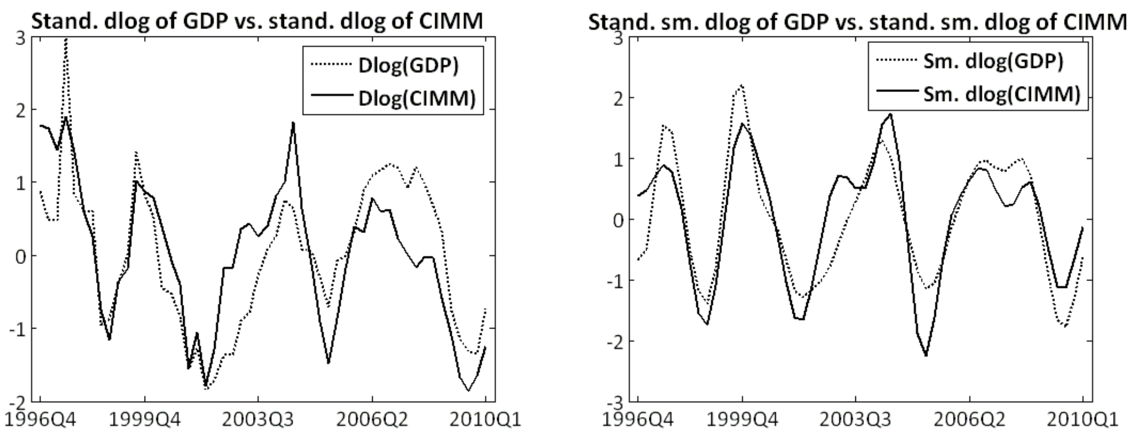
For both indicators the steepest error growth can be noticed for change from nowcast to one quarter ahead forecast and for shift from horizon of 3 to 4 quarters. The biggest discrepancy in the forecast based on coincident indicators was registered for 3 quarters ahead perspective.

Last part of this section is devoted to presentation of the results of pseudo real-time business cycle dating exercise. The aim of this part of the survey was to check usefulness of coincident indicators for the process of early critical points detection. The aforementioned feature is very useful for decision makers and agents observing published coincident indicators because it allows them to determine changes in future trends of economic situation development in particular country. Dating procedure has been ran on indicators transformed to quarter on previous year's quarter growths smoothed with Christiano-Fitzgerald filter. Additionally turning points dates were also determined for original series and gained results were almost the same as in case of transformed data. Set of figures grouped in two panels below shows inputs to the dating procedure for full time span (1996:Q4: 2010:Q1)

Panel 1.



Panel 2.



Basic comparison of above plots confirms that frequency of the business cycles determined by both coincident indicators are consistent with these recorded for the reference indicator. They are also coherent with previous research of the author devoted to synchronization of Polish business cycles (Lupinski [16] , 2009), which suggest that average business cycle span in Poland, measured for GDP series, lasts 3 years.

Pseudo real-time dating procedure, like nowcasting and forecasting exercise, was run on step-by-step growing sample. In the table below first occurrences of peaks and troughs are reported.

Table 8. Business cycles dates determined with Bry-Boschan procedure for GDP, CISW and CIMM

Peak/trough real date	Date of detection			Lag of detection (months)		
	GDP	CISW	CIMM	GDP	CISW	CIMM
Peaks						
June 2006 (2006:Q2)	November 2007	April 2007	February 2007	17	10	8
December 2007 (2007:Q4)	August 2008	-	May 2008	8	-	5
Troughs						
June 2005 (2005:Q2)	August 2006	April 2006	March 2006	14	10	9
June 2007 (2007:Q2)	November 2008	January 2009	November 2008	14	16	14

Source: own computations

It generally takes more than one year to detect turning point on reference time series. However use of M&M indicator allows to shorten this period of time (with exception to the second trough) even by more than 50%. Moreover its use guarantees detecting all turning points observed during pseudo real-time exercise horizon (5 years). This crucial feature is not shared by the Stock and Watson indicator, which misses one peak. It is also not so effective as its counterpart in shortening turning points detection lag.

7. Conclusions

In this paper second generation dynamic factor model with mixed frequencies and missing data handling was applied to create coincident indicator of Polish GDP and formulate with its help reference series' nowcast/short-term forecasts and date business cycles' phases. Despite general drawbacks of the dynamic factor models inherited by their second generation version (e.g. identification problems influencing ability of stochastic trend modeling) and their native problems connected with use of Gaussian distribution to replace missing observations, they turned out to be robust econometric tools for helping decision makers and other agents to summarize current and future state of Polish economy. Results achieved during simulated real-time exercise based on small dataset have shown that coincident indicators built with help of M&M model give basis for formulation of forecast that significantly outperform naive benchmark models and dynamic models of the first generation. When compared with S&W coincident indicators they offer smaller forecast errors especially in the perspective of one and two quarters. They also allow to detect business cycle turning points earlier than their counterparts. Hence, intuitive guess that including more recent monthly information in the econometric model can improve its explanatory and forecasting features has found its measurable proof.

During his planned work the author is going to augment presented econometric framework with module introducing endogenous stochastic trend /this plan is inspired by Fernandez-Macho's idea [9](1997)/. Second considered extension is connected with introduction into base model switching regime schema. The last augmentation is crucial in the context of changes in the structure of the Polish and global economy observed recently. Good example of such approach was presented in the paper of Camacho, Perez-Quirios and Poncela [7](2010) in which Markov switching framework was exploited. Nevertheless both two mentioned extensions need to be substantially customized to Polish economic data characteristics and carefully tested.

References

[1] Arouba S.B., Diebold F.X., Scotti C., 2009, "Real-Time Measurement of Business Conditions", Journal of Business and Economic Statistics, vol 27 (4), pp. 417-27

[2] Barhoumi K., Benk S., Cristadoro R., Den Reijer A., Jakaitiene A., Jelonek P., Rua A., Rünstler G., Ruth K. and Van Nieuwenhuyze Ch., 2004, "Short-Term Forecasting of GDP Using Large Monthly Datasets a Pseudo Real-Time Forecast Evaluation Exercise", ECB Occasional Working Paper, no. 84

[3] Boivin J., Ng S., 2006, "Are More Data Always Better for Factor Analysis?", Journal of Econometrics, 132, pp. 169-194.

[4] Bry G., Boschan C., 1971, "Cyclical Analysis of Time Series: Selected Procedures and Computer Programs", NBER

[5] Burns, A.F., Mitchell W.C., 1946, "Measuring Business Cycles", NBER

[6] Camacho M., Perez-Quiros G., 2009. "Ñ-STING: España Short Term INdicator of Growth," Banco de Espana Working Papers 0912, Banco de España.

[7] Camacho M., Pérez-Quiros G., Poncela P., 2010, "Green Shoots? Where, When and How?", Working Papers 2010-04, FEDEA.

[8] Forni M., et al, 2001. "Coincident and Leading Indicators for the Euro Area," Economic Journal, Royal Economic Society, vol. 111(471), pp. 62-85

[9] Fernandez-Macho F.J., 1997, "A Dynamic Factor Model for Economic Time Series", Kybernetika, vol. 33 (6), pp. 583-606

[10] Mariano R.S., Murasawa Y., 2003, "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," Journal of Applied Econometrics, vol. 18 (4), pp 427-443, John Wiley & Sons

[11] Inklar R., Jacobs J., Romp W. E., "Business Cycle Indexes: Does a Heap of Data Help?", CCSO Working Paper 200312

[12] `http://www.rug.nl/staff/r.c.inklaar/research` (accessed: 16.04.2010)

[13] Kim Ch. J" Nelson Ch., R., 1999 , "State Space Models with Regime Switching", MIT

[14] Kowal P., 2005, "Matlab Implementation of Commonly used Filters" Computer Programs 0507001, EconWPA

[15] Lupinski M., 2007, "Konstrukcja wskaźnika wyprzedzającego aktywnoście ekonomicznej w Polsce", ("Construction of Polish Economic Activity Leading Indicator"), PhD Dissertation, Warsaw University.

[16] Lupinski M., 2009, "Four Years After Expansion. Are Czech Republic, Hungary and Poland Closer to Core or Periphery of EMU?", Economics, vol. 22, pp. 75-103

[17] Stock J. H., Watson M. W., 1989. "New Indexes of Coincident and Leading Economic Indicators" NBER Chapters, in: NBER Macroeconomics Annual 1989, vol. 4, pp. 351-409, NBER.

[18] Stock J. H., Watson M. W., 1998, "Business Cycle Fluctuations in U.S. Macroeconomic Time Series," NBER Working Papers 6528, NBER

[19] Shumway R.H., Stoffer D.S., 1982, "An Approach to Time Series Smoothing and Forecasting using the EM algorithm", Journal of Time Series Analysis, vol. 3, pp. 253-264