van Os, Bram; van Dijk, Dick

Working Paper

Accelerating Peak Dating in a Dynamic Factor Markov-Switching Model

Tinbergen Institute Discussion Paper, No. TI 2020-057/VI

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: van Os, Bram; van Dijk, Dick (2020) : Accelerating Peak Dating in a Dynamic Factor Markov-Switching Model, Tinbergen Institute Discussion Paper, No. TI 2020-057/VI, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/229677

TI 2020-057/VI
Tinbergen Institute Discussion Paper

Accelerating Peak Dating in a Dynamic Factor Markov-Switching Model

Revision: November 29, 2020

Bram van Os[1]
Dick van Dijk[1]

[1] Econometric Institute, Erasmus University Rotterdam

Accelerating Peak Dating in a Dynamic Factor Markov-Switching Model*

Bram van Os† and Dick van Dijk‡

Econometric Institute, Erasmus University Rotterdam

November 29, 2020

Abstract

The dynamic factor Markov-switching (DFMS) model introduced by Diebold and Rudebusch (1996) has proven to be a powerful framework to measure the business cycle. We extend the DFMS model by allowing for time-varying transition probabilities, with the aim of accelerating the real-time dating of turning points between expansion and recession regimes. Time-variation of the transition probabilities is brought about endogenously using the accelerated score-driven approach and exogenously using the term spread. In a real-time application using the four components of The Conference Board's Coincident Economic Index for the period 1959-2020, we find that signaling power for recessions is significantly improved and are able to date the 2001 and 2008 recession peaks four and ten months before the NBER.

Keywords: Business cycles, generalized autoregressive score models, time-varying transition probabilities, turning points.

*We gratefully acknowledge helpful comments and suggestions from Monica Billio, Rutger-Jan Lange, Laurent Ferrara, Francis X. Diebold, Michel van der Wel and the participants of the CFE-CM Statistics 2019 and SNDE 2020 conferences. Any remaining errors are our own.

†vanos@ese.eur.nl

‡djvandijk@ese.eur.nl

1 Introduction

The business cycle is an important driver of many macroeconomic variables. Dating the turning points between the phases of this cycle, especially the peaks marking the transition from expansion to contraction, is of great interest to policy-makers, firms and investors alike. The dynamic factor Markov-switching (DFMS) model proposed by Diebold and Rudebusch (1996) has proven to be a powerful framework to measure the cycle, see Chauvet and Piger (2008) among others. This model extracts a latent business cycle factor by exploiting the cross-sectional information in multiple observed coincident variables. Furthermore, in line with the macroeconomic intuition of expansion and contraction phases in the cycle, the statistical properties of the factor (first and foremost its mean) are allowed to be regime-dependent with a hidden Markov process dictating the regime-switches. Chauvet and Piger (2008) find that the DFMS model compares favorably to the non-parametric dating method of Harding and Pagan (2003). Moreover, they find that the DFMS model is able to call the troughs of the cycle faster in real-time than the National Bureau of Economic Research (NBER), but do not find any improvements in timeliness for the peaks.

In this paper we extend the DFMS model with the aim of accelerating peak dating. With this purpose in mind, we allow the probability to switch from an expansion to a contraction phase to be time-varying. To bring about such time-variation we propose an autoregressive structure driven by endogenous information in the form of the log-likelihood score and additionally by exogenous variables. This framework is therefore a multivariate extension of the methods described by Bazzi et al. (2017) for Markov-switching models for univariate time series. Furthermore, we make use of the accelerated score-driven method recently introduced by Blasques et al. (2019) and tailor it to our needs. In this approach the magnitude of the parameter update for a given value of the score-function is also made time-varying. The resulting accelerated Generalized Autoregressive Score with eXogenous variables (aGASX) model thus combines the ideas of exogenous and endogenous drivers of the transition probabilities from Diebold et al. (1994) and Durland and McCurdy (1994), respectively.

We explore the empirical usefulness of the aGASX framework in an application involving the four components of The Conference Board's (TCB) Coincident Economic Index (CEI)

for the period 1959-2020. We consider both an ex-post analysis using the full sample with currently available revised data and a real-time exercise using appropriate vintages available from December 1976 until March 2020. For the exogenous inputs we consider an indicator for a negative term spread, which is generally considered to be one of the most prominent leading indicators, see e.g. Estrella and Mishkin (1998). In the ex-post analysis, we find that both score-driven endogenous dynamics and exogenous time-variation are significant and improve the signaling ability for recessions. While most of the improvement in dating performance stems from the exogenous information, we find that (accelerated) GAS dynamics are able to meaningfully amplify correct peak signals and reduce false ones. In the real-time analysis, we find similar benefits from the aGASX specification over the DFMS model with constant transition probabilities. This includes a significantly improved signaling power of the real-time contraction state probabilities for the NBER recession periods, with higher such probabilities at the start of four out five of the most recent recessions. Additionally, by converting real-time smoothed state probabilities to turning points, the aGASX specification is able to match or precede the peak announcements made by the NBER without any false signals. Most notably, our proposed model is able to date the peaks of the 2001 and 2008 recessions four and ten months before their NBER announcements, a gain of three and five months over the base DFMS model, respectively.

Our paper is related and contributes to several strands of literature. First, it is related to the vast literature on business cycle measurement. Here factor models have played and continue to play a prominent role. Stock and Watson (1989) propose a dynamic factor model and estimate both a leading and a coincident economic index using a selection of economic indicators. The DFMS model by Diebold and Rudebusch (1996) combines the idea of co-movement in multiple coincident macroeconomic series with the idea of regime-dependence as found by Hamilton (1989). Chauvet (1998) extends their work by jointly estimating both the latent factors and the latent regimes using the filter proposed by Kim (1994). Chauvet and Piger (2008) find that the DFMS model provides superior timeliness in terms of the real-time dating of turning points when compared to the NBER and the non-parametric method of Harding and Pagan (2003), mainly for the troughs. For an overview and comparison with more alternative dating methods, see Hamilton (2011). The DFMS model has also been successfully applied to macroeconomic data of many other countries, see

e.g. Norway by Aastveit et al. (2016), Germany by Carstensen et al. (2020) and Japan by Watanabe et al. (2003) among others, and today still remains a topic of interest. Recently, for example, Camacho et al. (2018) investigate the effects of ragged edges for the DFMS model and Doz et al. (2020) allow for time-varying long-run growth rates. The Markov transition probabilities play a primary role in the construction of the regime probabilities. In the context of empirical macroeconomics, Diebold et al. (1994) and Filardo (1994) argue that the transition probabilities of a Markov-switching model need not be constant and propose using economic variables to guide their evolution. Furthermore, evidence of duration dependence is found by Durland and McCurdy (1994) for GNP growth rates and by Kim and Nelson (1998), using a DFMS model, for coincident indicators.

Second, our paper is related to the literature that exploits leading indicators (LIs) to improve business cycle measurement, see Marcellino (2006) for an overview of the use and construction of LIs. In the context of the DFMS model, Huang and Startz (2020) allow the transition probabilities to depend on the volatility regime of the stock market. In a similar spirit Chauvet and Senyuz (2016) add a set of yield curve variables to the DFMS framework and consider a bi-factor setup. Here both find that turmoil in financial markets often precedes recessions of the real economy. While it is incredibly difficult, if not impossible, to pick a single best exogenous predictor for the entire history of the US business cycle, the term spread has historically been a good predictor, see Estrella and Mishkin (1998), Rudebusch and Williams (2009), Ng and Wright (2013) and Liu and Mönch (2016) among many others.

Third, our paper is related to the literature regarding Generalized Autoregressive Score-driven (GAS) models, introduced by Creal et al. (2013) and Harvey (2013). This method updates a time-varying parameter in the direction of the score and has attractive information-theoretic optimality properties. Koopman et al. (2016) show in an extensive Monte Carlo study that GAS models offer comparable performance to parameter-driven models, even if the latter matches the data generating process. Due to the observation-driven approach, standard maximum likelihood methods may be employed for estimation. Moreover, the approach has been found useful in a variety of empirical applications. In the context of Markov-switching models in particular, Bazzi et al. (2017) introduce score-driven time-varying transition probabilities and find improvements for describing the dynamic variance patterns in industrial production growth rates.

The outline of the paper is as follows. Section 2 presents how the DFMS framework may be enhanced by adding time-varying dynamics to the transition probabilities. Section 3 examines the results of the empirical application both ex post, with currently available revised data, and in real-time. Finally, section 4 concludes.

2 Methodology

2.1 Model specification

For clarity of exposition, we present the DFMS model for N coincident economic variables with two Markov states, a single factor and first-order autoregressive (AR(1)) dynamics. Extensions to more regimes, more common factors and higher lag-orders are relatively straightforward but tedious. To facilitate the discussion of the estimation procedure, we describe the model in state space representation. Let $y_{i,t}$ denote the observation of variable $i = 1, 2, ..., N$ at time $t = 1, 2, ..., T$. We assume that the $y_{i,t}$ are driven by a common latent factor ψ_t with factor loadings λ_i and idiosyncratic components $v_{i,t}$, such that the observation equation of the state space representation is given as follows:

$$\mathbf{y}_t = \mathbf{Z}\boldsymbol{\zeta}_t, \tag{1}$$

where $\mathbf{y}_t$ are the $y_{i,t}$ collected in a column vector, $\mathbf{Z}$ is a matrix of coefficients and $\boldsymbol{\zeta}_t$ is the state vector, i.e. we have that

$$\mathbf{y}_t = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \lambda_1 & 1 & 0 & \dots & 0 \\ \lambda_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_N & 0 & 0 & \dots & 1 \end{bmatrix} \text{ and } \boldsymbol{\zeta}_t = \begin{bmatrix} \psi_t \\ v_{1,t} \\ \vdots \\ v_{N,t} \end{bmatrix}. \tag{2}$$

We assume that the latent factor ψ_t follows a stationary AR(1) process with autoregressive parameter ϕ and intercept α_{S_t}, which is allowed to depend on the hidden Markov state $S_t \in \{0, 1\}$. The variance of the factor innovations is denoted σ_η^2. We also assume stationary AR(1) dynamics for the idiosyncratic components $v_{i,t}$ with autoregressive parameters θ_i and

error variances σ_i^2. The transition equation of the state vector $\boldsymbol{\zeta}_t$ is then given by

$$\boldsymbol{\zeta}_t = \mathbf{d}_{S_t} + \mathbf{T}\boldsymbol{\zeta}_{t-1} + \mathbf{Q}^{\frac{1}{2}}\boldsymbol{\omega}_t, \tag{3}$$

where the system matrices $\mathbf{d}_{S_t}$, $\mathbf{T}$ and $\mathbf{Q}$ are defined as

$$\mathbf{d}_{S_t} = \begin{bmatrix} \alpha_{S_t} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{T} = \begin{bmatrix} \phi & 0 & \dots & 0 \\ 0 & \theta_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_N \end{bmatrix} \text{ and } \mathbf{Q} = \begin{bmatrix} \sigma_\eta^2 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix}, \tag{4}$$

and $\boldsymbol{\omega}_t$ denotes an $(N \times 1)$ i.i.d. innovation vector which we assume to follow a multivariate standard normal distribution.

As the DFMS model contains both latent regimes and a latent factor, estimation makes use of both the Hamilton filter and the Kalman filter. Furthermore, parameter estimation requires either the use of an approximation of the likelihood or Bayesian methods, see Kim and Nelson (1999). This is because the calculation of the exact likelihood quickly becomes computationally infeasible, as the value of the factor at time t depends on all previous Markov states, a problem known as path-dependence. We follow the approach of Chauvet (1998), which makes use of the filter proposed by Kim (1994), and approximate the likelihood. Specifically, this method proposes a collapsing step to avoid the need to track an ever-increasing number of past states, such that only a modest history of states needs to be considered. To maintain sufficient accuracy one requires this history length to be at least one longer than the highest lag-order contained in the model, see Kim (1994) for further details. To obtain the parameter estimates we maximize the associated approximate log-likelihood obtained using the prediction-error decomposition, which for our setting is given by

$$\text{Log}L \approx \sum_{t=1}^{T} \text{Log}[\sum_{i,j \in \{0,1\}} p^{ij} Pr(S_{t-1} = i | I_{t-1}) \phi_t^{ij}(\mathbf{y}_t)]. \tag{5}$$

Here p^{ij} denotes the Markov transition probability $Pr(S_t = j | S_{t-1} = i)$ with $i, j \in \{0, 1\}$ and $\phi_t^{ij}(\mathbf{y}_t)$ denotes the multivariate normal density evaluated in $\mathbf{y}_t$ with mean $\boldsymbol{\mu}_t^{ij}$ and covariance $\boldsymbol{\Sigma}_t^{ij}$, in turn defined as

$$\boldsymbol{\mu}_t^{ij} = \mathbf{Z}\boldsymbol{\zeta}_{t|t-1}^{ij}, \tag{6}$$

$$\mathbf{\Sigma}_t^{ij} = \mathbf{Z}\mathbf{P}_{t|t-1}^{ij}\mathbf{Z}'. \tag{7}$$

Where $\boldsymbol{\zeta}_{t|t-1}^{ij}$ denotes the expectation of the state vector $\boldsymbol{\zeta}_t$ conditional on all information available at time t-1, denoted by the information set I_{t-1}, and the states being i and j at time t-1 and t respectively, and $\mathbf{P}_{t|t-1}^{ij}$ is similarly defined to be its covariance matrix. The complete prediction-update recursion and further details regarding estimation are provided in Appendix A.

2.2 Score-driven time-varying transition probabilities

The transition probabilities of the latent Markov process S_t play a key role in the timely identification of business cycle turning points. With the aim of speeding up this dating process, our novelty therefore pertains to allowing these transition probabilities to vary over time. This is achieved by extending the methods put forward in Bazzi et al. (2017), who consider time-varying transition probabilities in a univariate Markov-switching model by applying the generalized autoregressive score-driven (GAS) approach from Creal et al. (2013). Our extension is in fact threefold. First, we consider not a single but multiple time-series by virtue of the DFMS model. Second, in similar spirit as Diebold et al. (1994) and Filardo (1994), we allow for exogenous variables to guide the transition probabilities. Third, we adopt the novel accelerated GAS approach (aGAS) described in Blasques et al. (2019) and adapt it to our needs. This extension allows for more rapid changes from score information in the time-varying parameter in question if the current score aligns with its immediate predecessor and smaller changes if this is not the case.

Let p_t^{ij} denote the dynamic Markov transition probability $Pr(S_t = j|S_{t-1} = i)$ with $i, j \in \{0, 1\}$ that replaces the time-invariant p^{ij} in (5). To ensure that the transition probabilities remain in the unit interval we consider the following link function

$$p_t^{kk} = \frac{\exp(f_{k,t})}{1 + \exp(f_{k,t})},\ k \in \{0, 1\}. \tag{8}$$

Now assume that the time-variation of the vector $\mathbf{f}_t = [f_{0,t}, f_{1,t}]^T$ is captured by the following vector AR(1) process:

$$\mathbf{f}_{t+1} = \mathbf{w} + \mathbf{A}\mathbf{s}_t + \mathbf{B}\mathbf{f}_t + \mathbf{C}\mathbf{X}_t. \tag{9}$$

We follow the general paradigm of Creal et al. (2013) and consider the score function to construct $\mathbf{s}_t$. Note that the class of observation-driven time-varying parameter models is much broader, for example, recently Blasques et al. (2020) outline a class of models that nests the GAS approach and also allows for different target functions in lieu of the local log-likelihood to construct s_t. Additionally, let $\mathbf{X}_t$ denote some vector of exogenous variables known at time t. Here $\mathbf{w}$ denotes a vector of constants and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are matrices of coefficients.

The gradient or score of the (approximate) predictive log density at time t with respect to the $f_{k,t}$, which determine the transition probabilities according to (8), is denoted by $\boldsymbol{\nabla}_t^f$ and is given by

$$\boldsymbol{\nabla}_t^f = \frac{1}{L_t(\mathbf{y}_t)} \begin{bmatrix} \phi_t^{00}(\mathbf{y}_t) - \phi_t^{01}(\mathbf{y}_t) \\ \phi_t^{11}(\mathbf{y}_t) - \phi_t^{10}(\mathbf{y}_t) \end{bmatrix} \odot \begin{bmatrix} Pr(S_{t-1}=0|I_{t-1}) p_t^{00}(1-p_t^{00}) \\ Pr(S_{t-1}=1|I_{t-1}) p_t^{11}(1-p_t^{11}) \end{bmatrix}. \tag{10}$$

Here $\odot$ denotes the Hadamard product and $L_t(\mathbf{y}_t)$ the (approximate) likelihood at time t. In the context of GAS models, it is common practice to scale the score before using it to drive a time-varying parameter. Scaling by the square root (pseudo-)inverse Fischer matrix is often preferred, because this yields the attractive property of constant unit variances of the scaled scores. However, as the Fischer matrix is obtained from the expectation of the outer product of the score, this presents a large computational burden here. We shall therefore take a more practical approach to scaling instead. To facilitate the discussion of this process and with our empirical application in mind we proceed now with just a single time-varying transition probability.

When considering a single time-varying transition probability p_t^{kk}, we propose to adjust its score in two ways before using it in the update Equation (9). First, we drop the $p_t^{kk}(1-p_t^{kk})$ term found in the score, see Equation (10), which is a remnant of the use of the logistic function to map $f_{k,t}$ to p_t^{kk} by virtue of the chain-rule. From a practical standpoint, with the transition probability empirically often being very close to either 0 or 1, the product $p_t^{kk}(1-p_t^{kk})$ is close to 0, very much dampening any movement in p_t^{kk}. Note also that it is straightforward to show that in the case of scaling with the square root Fischer information this part would also disappear, see Creal et al. (2013).

Second, we propose to consider a function of the resulting simplified score instead of using

it directly. The reason for this is motivated by empirical findings. Namely, it is found that the score in the DFMS model can produce a large number of outliers. This happens when the likelihood found in the denominator in (10) becomes small. As a result the ability to drive the transition probability in a meaningful way is hampered. In a standard regression framework with a strictly positive regressor a straightforward remedy for this issue would be to consider the logarithm instead. However, as the score can be both positive and negative we propose the following intuitive transformation:

$$g(x) = \text{sign}(x)\log(1 + |x|). \tag{11}$$

This monotonic function $g(x)$ is antisymmetric around the origin, coinciding with the zero expectation of the score, that is close to the identity map for 'very small' $|x|$ and close to the logarithm for 'large' $|x|$. Effectively, the direction of the update is maintained while the magnitude of the update by a large absolute score is reduced relative to that by a small one. This approach thus prevents over-updating of our transition probability when a very large score, possibly due to an outlier, occurs. Integrating the transformed score with respect to p_t^{kk} reveals we are now applying a gradient update based on a robustified version of the log-likelihood, such that our simple approach is therefore similar in spirit to the robust-GAS models by Blasques et al. (2020). To conclude, we propose the following process to drive time-variation of the transition probability p_t^{kk} both endogenously using the transformed score and exogenously using additional known explanatory variables as follows:

$$f_{k,t+1} = w + a s_t^{f_k} + b f_{k,t} + \sum_{h=1}^{H} c_h x_{h,t}, \tag{12}$$

whereby $s_t^{f_k}$ is given by

$$s_t^{f_k} = g\left(\frac{\phi_t^{kk}(\mathbf{y}_t) - \phi_t^{k(1-k)}(\mathbf{y}_t)}{L_t(\mathbf{y}_t)} Pr(S_{t-1} = k|I_{t-1})\right), \tag{13}$$

and w, a, b and the c_h are (scalar) parameters to be estimated. Note that the expectation of $s_t^{f_k}$ may deviate from zero and even vary over time. In practice, however, for our models this difference is small and of no significant consequence.

2.3 Accelerated score-driven dynamics

In the accelerated score-driven framework by Blasques et al. (2019) the parameter $\mathbf{A}$ in (9), which determines how much to update in the direction of the score, is allowed to be time-varying. This is done by reapplying the score-driven approach to this parameter and under appropriate scaling can be seen to account for the autocorrelations of past score innovations. However, in view of the non-standard scaling of the previous section and in the absence of a computationally feasible estimator of the Fischer information, we provide a simple and intuitive analog for our specification.

Building upon the general idea of score-alignment to dictate the degree of the update, we propose the following approach. The parameter a in (12) is made a time-varying parameter a_t, such that it fluctuates smoothly between lower and upper bounds $a^l > 0$ and $a^u > a^l$, driven by a feasible proxy of the correlation coefficient between $s_t^{f_k}$ and $s_{t-1}^{f_k}$. Specifically, we specify the parameter a_t as follows:

$$a_t = a^l + a^u u_{t+1}, \tag{14}$$

where we assume that u_{t+1} admits the following dynamics

$$u_{t+1} = \delta u_t + (1 - \delta)\rho_t, \tag{15}$$

for some parameter $\delta \in (0, 1)$. Here, ρ_t denotes a feasible proxy of the correlation coefficient between $s_t^{f_k}$ and $s_{t-1}^{f_k}$ linearly mapped to the unit interval and is constructed as follows:

$$\rho_t = \frac{s_t^{f_k} s_{t-1}^{f_k}}{(s_t^{f_k})^2 + (s_{t-1}^{f_k})^2} + \frac{1}{2}. \tag{16}$$

Note that this expression of ρ_t is similar to that of the (transformed) sample autocorrelation and may more generally be simply interpreted as an intuitive measure of alignment. For example if $s_t^{f_k}$ and $s_{t-1}^{f_k}$ have the same value, then ρ_t is maximal at 1. Conversely, if $s_t^{f_k}$ and $s_{t-1}^{f_k}$ have the same absolute value but are of opposite sign this would put ρ_t to 0. The parameters a^l, a^u and δ are estimated jointly with the other parameters of the DFMS model in the maximum likelihood (ML) procedure.

The idea between this setup is as follows. First, by opting for integrated dynamics for u_t as in (15) and having its innovation term ρ_t in the unit interval, u_t is also contained therein. This approach limits the number of additional parameters to be estimated and also does not require us to consider an additional (non-linear) link function to ensure (at least) the positivity of a_t. Second, because here $s_t^{f_k}$ is a transform of the (scaled) score of $f_{k,t}$, it is not straightforward to apply the framework of Blasques et al. (2019) directly. Therefore by considering an intuitive measure of alignment, that may be interpreted as a first-order sample correlation, we provide an intuitively clear alternative that adheres to the core rationale of their accelerated approach.

3 Empirical application

3.1 Data

For our empirical application, we consider the four components of TCB's CEI for the US economy: employees on nonfarm payrolls (EMP), industrial production (IP), manufacturing and trade sales (MAN) and personal income less transfer payments (INC). We analyze these variables at a monthly frequency from January 1959 until February 2020. The vintages for these series, which are used for the real-time exercise, are obtained from TCB and are supplemented with data from Jeremy Piger[1] for the earlier vintages. A more elaborate discussion of the data and its handling, alongside a graphical illustration of the final vintage data can be found in Appendix B. The DFMS model and its extensions are estimated based on the monthly logarithmic growth rates of these four indicators.

Furthermore, the interest rate spread (IRS), here constructed by subtracting the US Federal Funds (FF) rate from the 10-year US Treasury rate, is used as an exogenous variable. The variable is obtained from the Federal Reserve Bank of St. Louis database (FRED) and is used by TCB as one of the components in their Leading Economic Index (LEI). A graphical illustration of the IRS, an indicator for its negativity and how they relate to the NBER recession periods, is provided in Appendix Figure B.2. Here we observe that business cycle recessions almost always are preceded by a period with negative IRS. We remark that

[1] https://pages.uoregon.edu/jpiger/research/published-papers/raw-real-time-data.zip

unconventional monetary policy implemented after the 2008 recession, quantitative easing in particular, may have artificially suppressed term premiums. When considering also the large differences in the sizes of the interest rates during the 1980s compared with the rest of the sample, we argue that considering a dummy of the IRS may therefore provide an additional level of robustness. Alternately, one could consider so-called shadow rates, see e.g. Wu and Xia (2016). This is left for future research.

For our empirical analysis we will only make use of a single exogenous variable, namely an indicator for a negative IRS. Results of several robustness checks, whereby the IRS is incorporated in a different manner can be found in Appendix Figure C.1 and C.2. These figures additionally contain findings for using the LEI as the exogenous variable, which also appears to be an effective choice. Note however that these results include all revisions of the LEI known at the final date. The reason for considering the IRS over the LEI is first because the former is found to be slightly more timely for dating peaks and has more synergy with the score-driven approach, which is naturally more coincident. Moreover, the IRS is a far simpler option as the composition and weighing of the LEI has changed significantly over the years. Considering more or a weighted combination of leading indicators to drive the transition probability is left for future research.

3.2 Model for empirical application

For our empirical application we consider the two-regime one-factor DFMS model with AR(1) specifications for both the common factor as well as the idiosyncratic components. The choice for two regimes is motivated by our interest in dating business cycle turning points. Although more involved multi-state models, such as the three phase model by Sichel (1994), can provide a better in-sample fit, it appears sensible here to consider just two regimes in view of the relative paucity of recessions found in the sample. The magnitudes of the eigenvalues of the correlation matrix of the growth rates of the four coincident indicators motivate the choice of a single factor. The choice of only a single lag for the factor and idiosyncratic components, similar to Chauvet (1998) for monthly data, is mainly to not over-complicate the already reasonably involved inference and estimation. Further extensions, including a structural break in output volatility to accommodate the Great Moderation, see McConnell and Perez-Quiros (2000), and time-varying mean growth rates, see Eo and Kim (2016), Eo

and Morley (2019) and Doz et al. (2020), are left for future research. To identify the factor, we fix the first factor loading i.e. $\lambda_1 = 1$. Additionally, to identify the regimes we set $\alpha_1 < 0$, such that $S_t = 1$ corresponds to a contraction and consequently $S_t = 0$ to an expansion.

Furthermore, as we are primarily interested in improving peak dating and find little evidence of strong relevant dynamics in p^{11}, we allow only $p^{01} = 1 - p^{00}$ to vary over time. Although equivalent, we present our results in terms of p^{01} instead of p^{00} for expositional purposes. This transition probability p^{01} reflects the probability to switch from expansion to a contraction phase and is therefore henceforth informally referred to as the peak probability. In addition, we propose to not estimate p^{11} alongside the other parameters during the ML procedure, but instead calibrate it directly from the completed NBER recessions. This means that for the ex-post analysis all recessions in the sample are used for calibration, whereas for the real-time exercise the value for p^{11} is updated the month after a trough announcement by the NBER. This can be interpreted as a form of targeting and by fixing p^{11} we guarantee that the recession regimes are sufficiently persistent. In practice, we find for the time-varying models that p^{11} is reduced, if it is estimated in the ML procedure, paired with large movement in p^{01} before and during recession periods. This may lead to patterns that allow for expansion months during recession months with only moderately negative or even positive growth rates of the four coincident indicators. Calibrating p^{11} directly using past NBER data ensures no such undesirable patterns occur. The evolution of this dynamically estimated p^{11} from NBER recession periods is given in Appendix Figure B.3 and shows little variation over time. Unsurprisingly, the NBER based estimator falls well within any sensible confidence interval of the ML-estimate of p^{11} in the base model when estimated alongside the other parameters. Note that due to the non-linear link function, it is not straightforward to apply a similar targeting approach for the unconditional expectation of p_t^{01} from completed NBER recessions.

For the ex-post analysis using the full sample we consider a total of six (nested) DFMS model specifications. The first is the base model as given in (1) and (3) with a time-invariant peak probability p^{01}, while the second and third introduce GAS and aGAS dynamics for this transition probability (using (8) and (12) with $c_h = 0$ for the GAS specification, and further combined with (14)-(16) for aGAS) and are referred to accordingly. The final three specifications are similar to the first three but include the dummy variable for a negative term

spread as an additional driver of the time-varying transition probability p^{01} and are denoted in order by Exo, GASX and aGASX. For the specification with only exogenous information, we allow for a time-varying peak probability by using (12) with $a = 0$, that is, we ignore the score information but maintain the autoregressive structure. With regards to the aGAS specifications it is found for our empirical application that the estimates of δ in (15) are statistically indistinguishable from 0 and for parsimony therefore impose this value for δ. While the estimates of a^l in the aGAS framework are also often found to be insignificant we retain this parameter for the full-sample analysis, such that the model specifications to be considered are nested, allowing for straightforward likelihood-ratio (LR) testing.

3.3 Full-sample estimation results

In this section we discuss the estimation results of the DFMS model and our extensions for the period January 1959 until February 2020 using the data as released in March 2020. Hence, this includes all the revisions known at the final date. In Table 1, for brevity, only key parameter estimates for the six considered model specifications are shown. The remaining parameter estimates can be found in Appendix Table C.1.

In Table 1, we observe from the log likelihood and the Akaike information criterion (AIC) that the various specifications that allow for a time-varying peak probability p_t^{01} improve upon the base model to different degrees. This includes notable improvements from using the IRS as exogenous input and more modest improvements due to the endogenous information captured by the (accelerated) GAS dynamics. In terms of parameter estimates, we find that the parameter of the negative IRS indicator is positive and significant for all specifications. The positive sign is in accordance with economic theory, which suggests a higher probability of a change to a recession state in the face of an inverted yield curve. In addition, the estimates of the autoregressive parameter b in (12) suggest that p_t^{01} is highly persistent.

While the improvements in log likelihood due to the GAS dynamics seem modest, LR tests indicate they are significant, irrespective of the inclusion of the exogenous variable. Specifically, the LR tests for the GAS versus the base specification and the GASX versus the Exo specification reject the null hypothesis at the 5 percent level and are given by $LR = 11.87$ (p-value $= 0.003$) and $LR = 4.6179$ (p-value $= 0.032$) respectively. In addition,

Table 1: Key parameter estimates for the DFMS model with variants that consider a time-varying p^{01}.

	Base	**GAS**	**aGAS**	**Exo**	**GASX**	**aGASX**
w		-0.559	-0.761	-0.463	-0.553	-0.514
		(0.218)	(0.458)	(0.167)	(0.183)	(0.175)
b		0.839	0.743	0.936	0.910	0.902
		(0.050)	(0.112)	(0.023)	(0.025)	(0.029)
a		1.207			0.979	
		(0.399)			(0.461)	
a^l			0.000			0.220
			(0.054)			(0.859)
a^u			2.602			1.261
			(0.996)			(1.295)
IRS<0				0.533	0.685	0.723
				(0.171)	(0.221)	(0.234)
p^{01}	0.017					
	(0.006)					
LogL	-1913.3	-1907.3	-1905.4	-1900.2	-1897.9	-1897.3
k	16	18	19	18	19	20
AIC	3858.5	3850.7	3848.8	3836.5	3833.9	3834.7

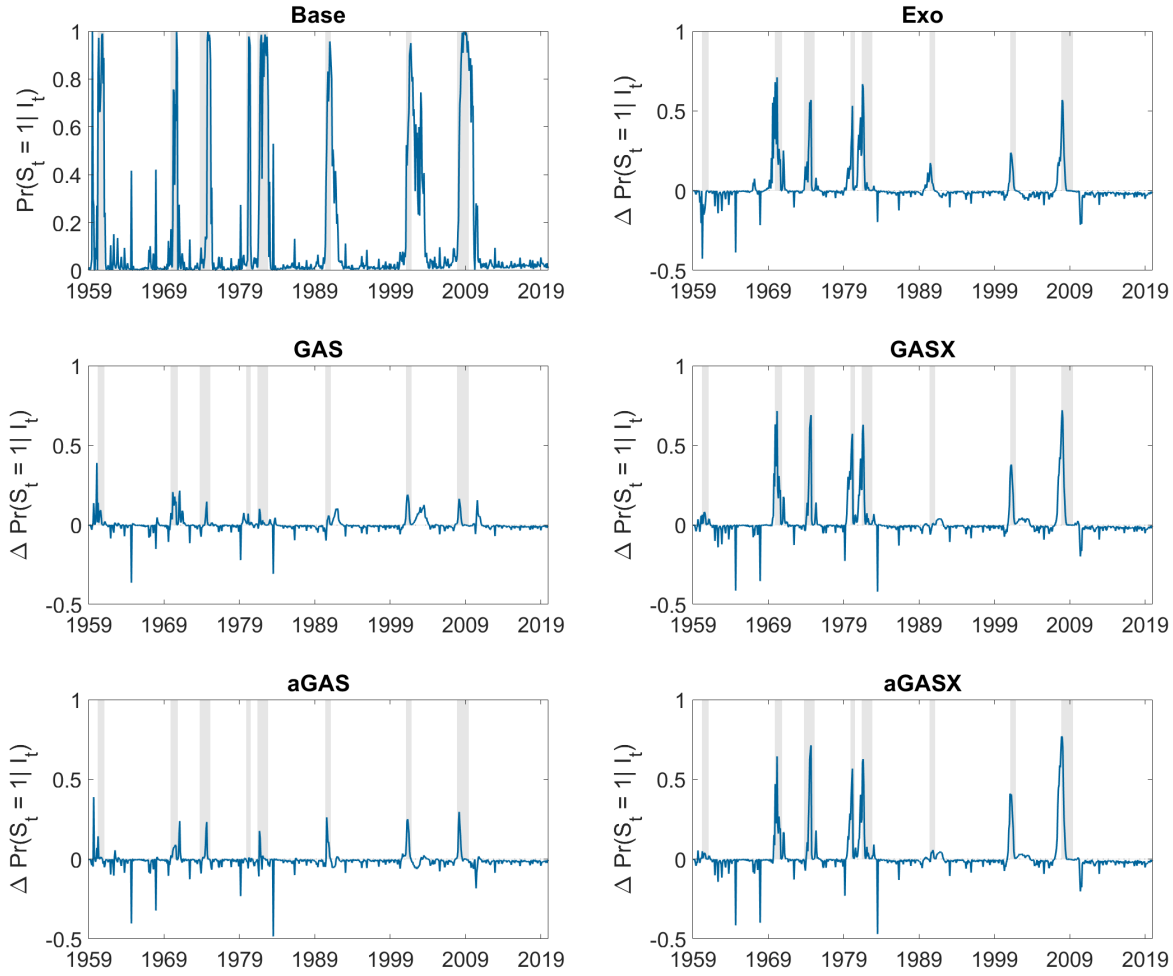
Note: This table presents key parameter estimates for the base model and the extensions that allow for a time-varying transition probability p_t^{01} using monthly growth rates for EMP, IP, MAN and INC over the period January 1959-February 2020. Standard errors are displayed in parentheses and k denotes the number of estimated parameters.

we find that the accelerated GAS dynamics are significant, at the 5 percent level, in the absence of the exogenous variable when testing against regular GAS dynamics ($LR = 3.909$, p-value $= 0.048$), but not in the presence of the exogenous variable when testing against the GASX specification ($LR = 1.201$, p-value $= 0.27$). Considering the AIC or the individual significance of the relevant parameters corroborates these findings. In particular, we find the GAS parameter a to be significant in both specifications, whereas the aGAS parameter a^u is significant only in the absence of the exogenous input. The lower bound a^l in the aGAS specifications is statistically indistinguishable from 0 in either case.

We evaluate the usefulness of the models for describing the business cycle regimes by means of the filtered state probabilities $Pr(S_t = 1|I_t)$. A model is considered better if these state probabilities more closely match the NBER recessions, whereby in particular an eye is kept on the timely identification of business cycle peaks. Figure 1 compares the filtered state probabilities of the six specifications with the NBER recession dates. For

comparison, we present the filtered state probabilities only for the base model, while for the other specifications we show the differences of their filtered state probabilities with the base model. Specifically, we subtract the filtered state probabilities of the base model from the filtered state probabilities of the specifications that allow for a time-varying peak probability p_t^{01}.

Figure 1: Filtered state probabilities for the DFMS model with variants that consider a time-varying peak probability.



Note: Filtered contraction state probabilities ($Pr(S_t = 1|I_t)$) for the base model are depicted in the top left panel. The remaining figures depict the filtered contraction state probabilities of the extensions minus those of the base model, denoted by $\Delta Pr(S_t = 1|I_t)$. Finally, the shaded areas reflect the recession periods as determined by the NBER.

We observe from Figure 1 that the base model by itself is quite successful in the iden-

tification of the business cycle regimes. The filtered probabilities remain close to 0 during expansion periods and rapidly increase to levels close to 1 during recessions. Note, however, that the base model can be somewhat slow in picking up the business cycle peaks, most notably for the recessions starting in 1969 and 1973. If the state probabilities are low during the first month(s) of a recession, this may lead to delays in the detection of the peaks of the cycle. In addition, we find that the model can be noisy at times, for example around 1964 and 1968 relatively large spikes are found. We also observe that the base model can be slow to recognize a trough, particularly so for the 2001 recession. For this recession, these findings could stem from what is known as a jobless recovery, whereby employment remained low while output experienced steady growth. A possible explanation could be structural change whereby a reallocation of workers across industries may have hampered growth, see Groshen and Potter (2003). Particularly the nonfarm payrolls employment series used here contains a very slow recovery in 2001 relative to a civilian employment metric, see Chauvet and Hamilton (2006). However, because the civilian employment metric is much noisier, using it would weaken the dating of most other turning points, see Doz et al. (2020) for a comparison in a similar setup.

Most notably, Figure 1 demonstrates that the addition of the IRS variable provides large improvements over the base model. We observe higher recession state probabilities during recession periods and lower such probabilities during expansions compared to the base model. In particular, the largest probability differences are found around the peak dates, precisely when they are deemed most important. Comparing the three specifications that make use of the IRS, we find that adding (a)GAS dynamics improves slightly upon the Exo specification which makes use of only the exogenous variable by increasing the differences around most peaks. For the GAS and aGAS specifications in the absence of the exogenous variable, Figure 1 indicates that the GAS dynamics modestly improve upon the base model, whereby in turn the aGAS specification improves upon the GAS specification. Here we observe that the aGAS specification features larger contraction state probabilities around the start of recessions and is better at reducing false signals when compared to the GAS specification.

Table 2 contains an overview of the signaling performance for the NBER recessions by the considered model specifications, which confirms the graphical evidence in Figure 1. Here we consider the Area-Under-the-Receiver-Operating-Curve (AUC), a common measure for

evaluating the quality of binary classification ability, see Berge and Jordà (2011) for an application involving recession and expansion classification. Specifically, this metric reflects the area under the curve obtained when plotting the true positive rate against the false positive rate for different signaling thresholds and is thus contained in the unit interval. A perfect classifier and an uninformative classifier have a value of 1 and 0.5 respectively, whereas a value below 0.5 reflects a classifier with an informative signal but with opposite sign. In addition, we consider the average contraction state probability during recessions, non-recession periods and the first month of the recessions, denoted by π^r, π^e and π^p respectively. We observe that all extensions improve upon the signaling ability of the base model, with the largest improvements stemming from the addition of the IRS information, which, for example, leads to an increase of the overall AUC from 0.941 to 0.979. Furthermore, we find for peak dating specifically, that GAS dynamics are a useful addition either on its own, increasing π^p from 0.267 to 0.290, or alongside the exogenous variable increasing π^p from 0.528 to 0.594. When considering the ratio of π^r and π^e, which can be seen seen as measure for contrast, we find that the aGAS specification achieves the highest value. This indicates that in a purely score-driven framework the accelerated framework may bring improvements over simple score-driven dynamics.

Table 2: Signaling performance of the state probabilities of the different DFMS specifications.

	Base	**GAS**	**aGAS**	**Exo**	**GASX**	**aGASX**
AUC	0.941	0.950	0.954	0.979	0.978	0.977
AUC Pre	0.937	0.942	0.950	0.975	0.975	0.973
AUC Post	0.972	0.975	0.984	0.989	0.989	0.992
AUC Pred	0.894	0.916	0.920	0.961	0.965	0.961
π^r	0.647	0.685	0.689	0.779	0.802	0.799
π^e	0.066	0.063	0.052	0.063	0.065	0.064
π^r/π^e	9.731	10.832	13.319	12.337	12.419	12.402
π^p	0.267	0.290	0.288	0.528	0.594	0.596

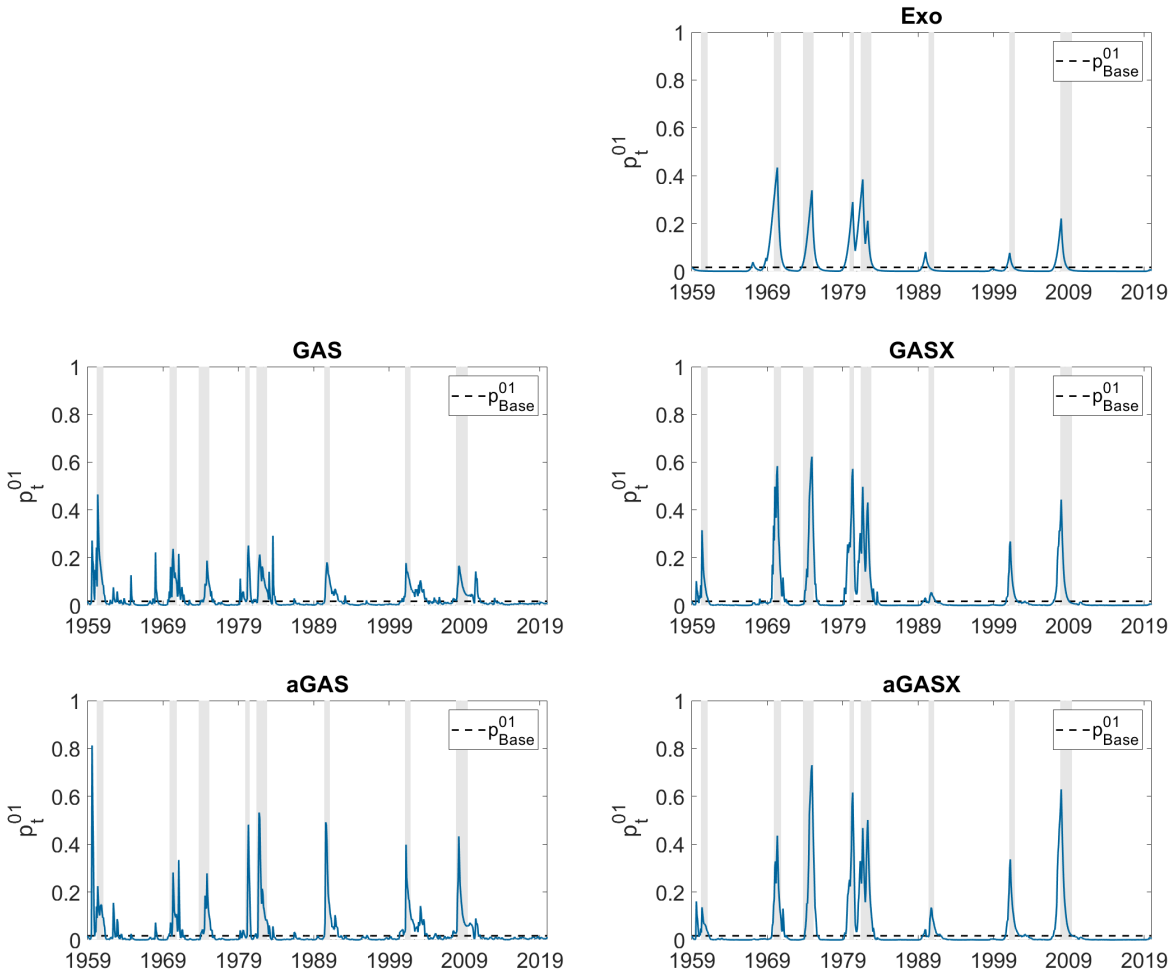
Note: Signaling power of the filtered state probabilities for the NBER recessions is evaluated using the Area-Under-Receiver-Operating-Characteristic curve (AUC), whereby Pre and Post indicate the subperiod before and after 1984 respectively and Pred the AUC for the predicted state probabilities. Furthermore, we have that $\pi^{r(e)}$ represent the average filtered contraction state probability during NBER recession (expansion) periods and π^p the average probability in the first recession month.

In addition, Table 2 contains the AUC of the predicted state probabilities $Pr(S_t = 1|I_{t-1})$. Unsurprisingly, we find the signaling power of the predicted state probabilities to be lower

than their filtered counterparts. More importantly, we find that the addition of a time-varying peak probability yields even greater improvements for the signaling power here. Indicating that with a minimum of a one month delay in data releases and if one is interested in making a nowcast, then the addition of a time-varying peak probability may be even more rewarding. To consider the effects of the Great Moderation, see e.g Bernanke (2004) for an overview, we also consider the AUC of the filtered state probabilities before and after 1984. Here we find that the recessions post 1984 appear easier to date, for example the overall AUC is equal to 0.937 before 1984 and afterwards increases to 0.972. Relative improvements of the addition of a time-varying peak probability p_t^{01} are, however, fairly comparable.

By construction, the relative differences in filtered state probabilities and their signaling performance directly stem from the differences in the specification for the peak probability p_t^{01}. Therefore we now consider the time-evolution of this parameter in more detail. In Figure 2 the estimates of p_t^{01} are compared to the constant estimate of the base model. Naturally, a model specification is considered superior if it has a heightened p_t^{01} just before or at the start of a recession. In Figure 2, we find that variation in the peak probability in the GAS and aGAS specifications appears sensible, but is mostly coincident. Therefore peaks are not signaled in advance, but rather the switches are made more extreme. This is not too surprising considering the fact that the a(GAS) specifications still only make use of coincident information. We do observe that the aGAS specification seems more successful in this regard. Comparing the aGAS specification to the GAS specification, we observe that the peaks are amplified while other signals are dampened. This confirms the idea that, in a purely score-driven framework, the idea of score-alignment for dictating the magnitude of the parameter update appears promising in this application. For the three other specifications, which include the IRS, we observe large movements already before the start of recessions. Here adding GAS dynamics strengthens the movement of p_t^{01} and also adds a small increase before the 1960 recession for which no yield curve signal is contained in the data. The addition of accelerated GAS dynamics to the exogenous information, further increases the movement of p_t^{01} for the 1974, 1990, 2001 and 2008 recessions but reduces it somewhat for the 1960 and 1970 recessions, compared to the GAS specification. Therefore, as indicated by the estimates from Table 1 and the performance metrics from Table 2, accelerated dynamics do not appear to provide much increased benefits over simple GAS dynamics in the presence

Figure 2: Transition probability p_t^{01} for the different DFMS specifications.



Note: This figure displays the evolution of the transition probability p_t^{01} over time for the DFMS model specifications. The dotted line represents the constant p^{01} estimated by the base model. Finally, the shaded areas reflect the recession periods as determined by the NBER.

of the exogenous input.

We conclude that exogenous information, here an indicator for a negative IRS, can help improve the dating of recessions and their peaks in particular, when used to guide the transition probability to switch from an expansion to a contraction directly. Furthermore, more advanced methods using score information may be employed. In the absence of exogenous information, the GAS specification presents modest improvements over the base model, with accelerated GAS dynamics providing increased performance by amplifying correct signals and reducing erroneous ones. When combining exogenous and score information, we find

that the GAS dynamics are still a useful addition, particularly for peak dating, but that additional accelerated dynamics do not provide much further increased benefit.

3.4 Real-time exercise

In the previous section we found that allowing for a time-varying transition probability to switch from an expansion to a recession may aid in the timely identification of business cycle peaks. However, estimating the model on the entire sample yields results which would not be available in real-time. Furthermore, a plethora of revisions in the coincident variables make it necessary that we analyze the performance of the DFMS specifications using real-time data vintages for a fair assessment of their potential. In this section we perform such a real-time exercise for the base DFMS model and the aGASX extension. For the aGASX specification we set a^l, that is the lower bound in the smooth transition interval for the GAS parameter a_t in (14), to 0 as it is mostly insignificant. Combined with again $\delta = 0$, we have now that $a_t = a^u \rho_t$, such that the magnitude of the update by the score is made proportional to a proxy of its correlation with its predecessor. The choice for this aGASX specification over the GASX specification is twofold. First, this specification provides the lowest information criteria, as it produces a slightly better likelihood than the GASX model with now an equal number of parameters. Second, the accelerated model variant appears more stable for earlier vintages when not much data is available by reducing false signals due to its consideration of score-alignment. However, results are qualitatively similar without the accelerated component here.

Chauvet and Piger (2008) consider the data at the end of the month and restrict the sample at each point in time to the series for which the least amount of information is available. We instead proceed with the maximum amount of data at our disposal in the third week of each month, in line with Camacho et al. (2018). Specifically, this entails that at time t, we have EMP and IP available up to and including time $t-1$. For MAN and INC we are presented with additional delays in data publication, such that we have observations only up to and including time $t-3$ and $t-2$ respectively. Parameter estimation at time t is thus based on all observations up to and including time $t-1$, whereby the final two observations for MAN and the final observation for INC are considered missing.

The results of the outlined real-time exercise, whereby the models are repeatedly esti-

mated with all information available at each point in time, are supplied below. Table 3 presents a comparison of the signaling performance of the real-time filtered and predicted state probabilities. We observe in Table 3 that the aGASX specification trumps the base

Table 3: Real-time signaling performance for the base DFMS model and the aGASX extension.

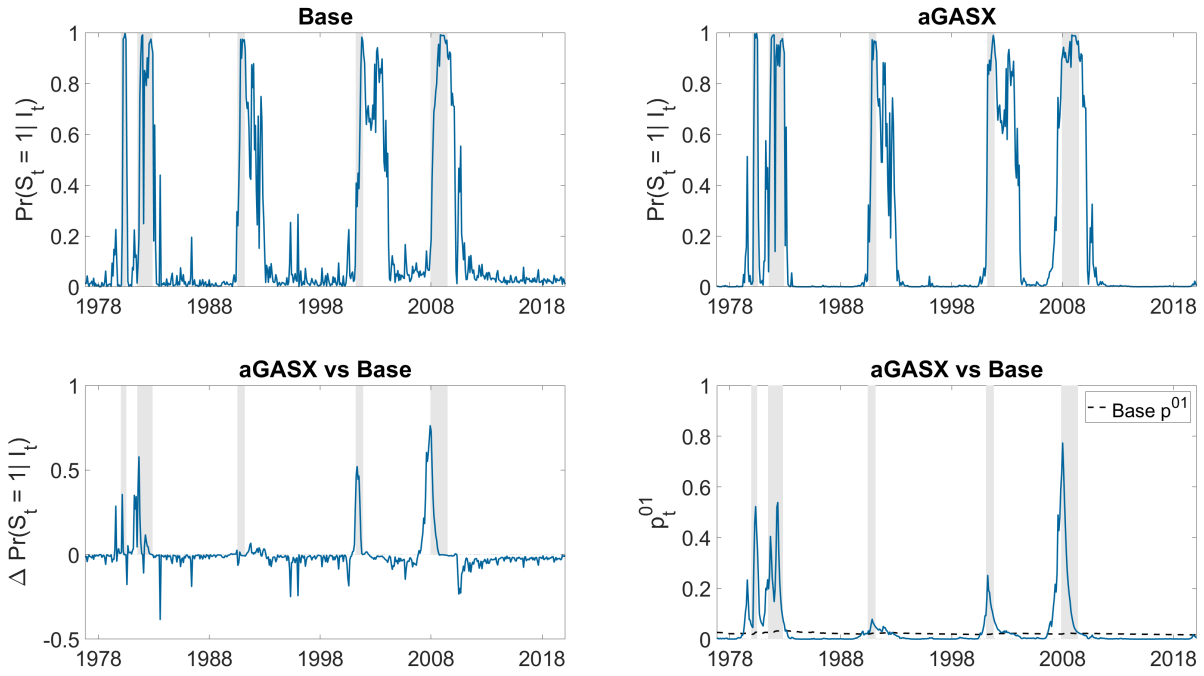
Filtered	**Base**	**aGASX**	*Predicted*	**Base**	**aGASX**
AUC	0.940	0.973	AUC	0.894	0.960
$\pi^r_{t-1\|t-1}$	0.751	0.864	$\pi^r_{t\|t-1}$	0.620	0.761
$\pi^e_{t-1\|t-1}$	0.123	0.113	$\pi^e_{t\|t-1}$	0.140	0.119
$\pi^r_{t-1\|t-1}/\pi^e_{t-1\|t-1}$	6.083	7.625	$\pi^r_{t\|t-1}/\pi^e_{t\|t-1}$	4.440	6.374
π^p	0.185	0.491	π^p	0.116	0.369

Note: Signaling power of the probabilities for the NBER recessions is evaluated using the Area-Under-Receiver-Operating-Characteristic Curve (AUC). Furthermore, we have that $\pi^{r(e)}_{i|j}$ represent the average contraction state probability at time i during NBER recession (expansion) periods with observations up to and including time j, which corresponds to the real-time estimation at time $j+1$. Finally, π^p denotes the average state probability during the first month of the recessions within the evaluation sample.

model in all considered metrics, with higher contraction state probabilities during NBER recession periods (π^r) and lower such probabilities during NBER expansion phases (π^e). Most importantly, in the interest of dating peaks, we find that the average contraction state probability for the first 'official' (NBER) recession month (π^p) is more than doubled(tripled) for the filtered(predicted) state probabilities. In addition, we find overall improved signaling performance of both the filtered and predicted contraction state probabilities for the NBER recessions in terms of the AUC. Moreover, these improvements of the aGASX method over the base methodology are found to be significant for both the filtered and predicted contraction states probabilities (p-value = 0.035 and p-value = 0.005 respectively) using the (one-sided) testing methodology of Hanley and McNeil (1983).

Figure 3 depicts the diagonal of the full history of filtered state probabilities for the base DFMS model and the aGASX specification, which can be found in Appendix Figure D.3. Specifically, Figure 3 compares for all points in the evaluation sample the most recent filtered state probability at time t, which is that of month $t-1$ due to the data delay, to the NBER recession indicator at time $t-1$. Additionally, the figure displays the evolution of the peak probability p^{01}_t, which is due to the observation-driven update equation known at time t and thus compared to the NBER indicator at time t. We observe in Figure 3

Figure 3: Real-time results for the base DFMS model and the aGASX extension.



Note: The top left and top right figures display the filtered state probabilities for the base and aGASX model respectively and the bottom left presents their difference (Base minus aGASX). Furthermore, the bottom right figure displays the evolution of the transition probability p_t^{01}. Finally, the shaded areas reflect the recession periods as determined by the NBER.

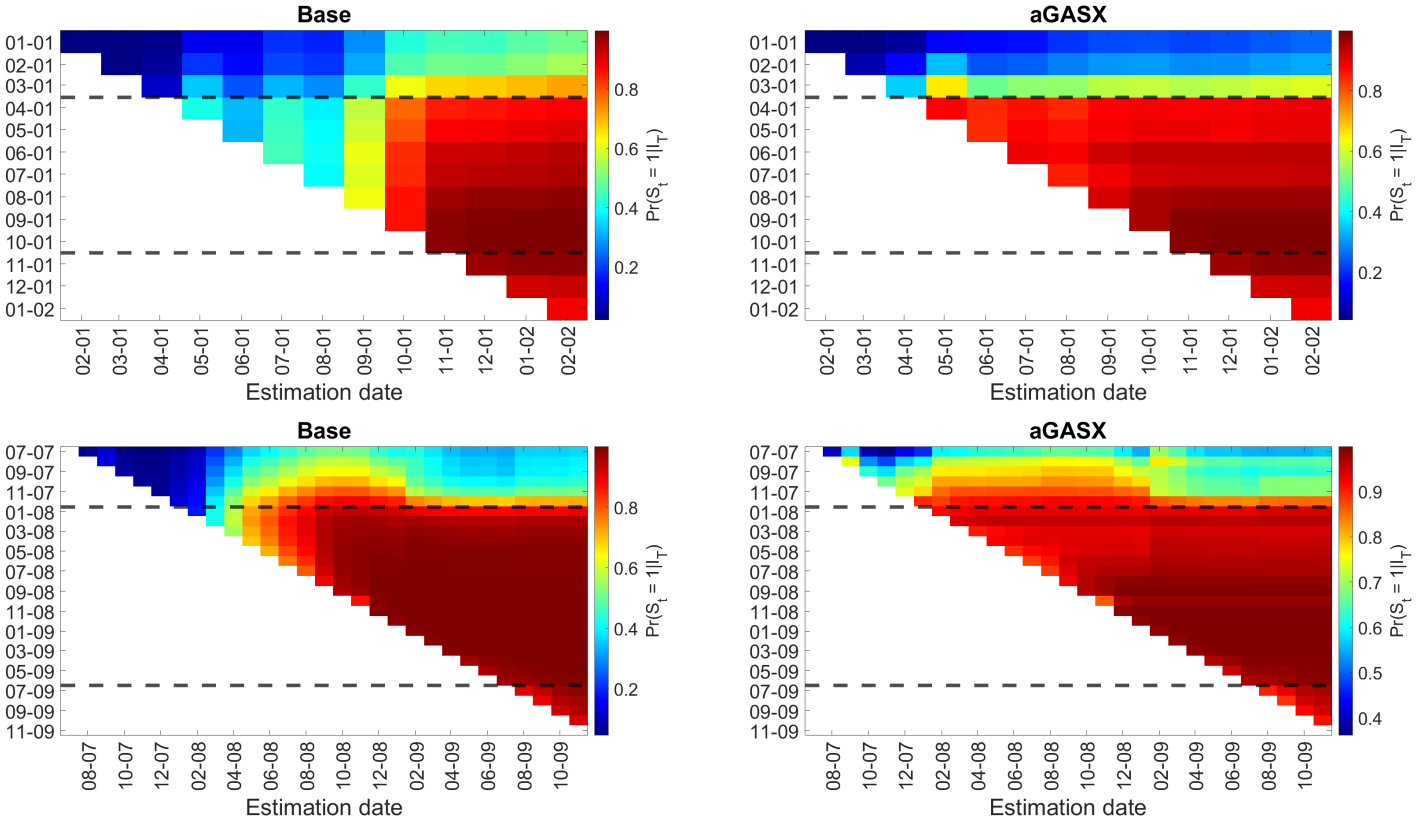
that the aGASX specification much better matches the NBER recession periods relative to the base model, in line with the statistics from Table 3. Moreover, the extension seems to be better able to identify business cycle peaks in four out of the five recessions that are contained in the evaluation sample. Most notably, it indicates that the 2008 recession could have been anticipated well in advance using the proposed framework. For the 1990 recession, the aGASX model initially appears to slightly dismiss the signal relative to the base model, although this difference is very minor. Interestingly, Stock and Watson found similar results using their leading indicator at the time when incorporating the IRS, see Hamilton (2011) for a discussion. Moreover, in Appendix Figure D.1 a brief comparison between the performance of the model that uses just the exogenous variable (i.e. the Exo specification) and the aGASX specification is given. There we find that the small dip for the 1990 peak is found to be much worse without aGAS dynamics. This indicates that while the IRS appears to provide the most benefit, the aGAS dynamics play a role as well, boosting

the correct signals of the spread and dampening its mistakes. In sum, this underlines that a combination of exogenous and endogenous information for driving the transition probability may be particularly powerful. To investigate the effects of data revisions, Appendix Figure D.2 contains the results obtained when making use of final vintage data. Although fit is improved somewhat for both specifications, findings remain qualitatively unchanged.

As we observed the largest benefits for the aGASX specification in the two most recent recessions, we will further investigate the evolution of the state probabilities around these periods in more detail. Figure 4 depicts the evolution of the real-time smoothed contraction state probabilities for the recessions that followed the peaks in March 2001 and December 2007. The full history of smoothed state probabilities can be found in Appendix Figure D.3. In Figure 4 we observe that the aGASX specification produces much higher contraction state probabilities than the base model during the first months of the 2001 and 2008 recessions. As a result the aGASX specification is able to accurately provide a strong contraction signal as soon as these recessions start, whereas the base model requires multiple months of additional information to do so. These differences are substantial and naturally of direct relevance for dating the peaks of the cycle. Also note that as more time passes the models appear to largely agree on past states. Appendix Figure D.4 contains a plot similar to Figure 4 for the filtered state probabilities, here improvements are even more pronounced. All in all these findings indicate that the aGASX specification may be able to date the 2001 and 2008 peaks several months ahead of the base model depending on the conversion rule, which we consider next.

As our aGASX model specification pertains to the transition probability to enter a recession, we shall focus our attention on peak dating. For this purpose we use a straightforward two-step conversion rule, similar to the one used by Chauvet and Piger (2008). Specifically, we propose the following identification scheme for each estimation date (i.e. for each vintage 'column'). First, to call a recession we require $Pr(S_t|I_T) < \tau$ and $Pr(S_{t+k}|I_T) \geq \tau$, for k = 1,2,3 for some threshold $0.5 < \tau < 1$. Second, the peak associated with this recession period is identified by the point in time where the probability crosses a half, immediately preceding this period. Meaning that we find the smallest non-negative integer q such that $Pr(S_{t-q-1}|I_T) < 0.5$ and $Pr(S_{t-q}|I_T) \geq 0.5$. The peak date is subsequently identified to be $t - q - 1$ and as such refers to the final expansion month. Before another peak can be

Figure 4: History of real-time smoothed state probabilities for the base DFMS model and the aGASX extension around the 2001 and 2008 recession.



Note: This figures depicts the real-time smoothed state probabilities around the 2001 (top) and 2008 (bottom) recessions for the base (left) and aGASX specification (right). The x-axis contains the estimation date and the y-axis the sample date for which a probability is constructed. Finally, the black dotted lines reflect the turning points as determined by the NBER.

called we require that the state probability remains below τ for three consecutive periods. The first such a period after a peak may then be established to be its corresponding trough, although more generally it may be given its own threshold. Here for simplicity, we label the first period t, after a peak has been dated, for which $Pr(S_t|I_T) \geq \tau$ and $Pr(S_{t+k}|I_T) < \tau$, for $k = 1,2,3$ as a trough and as such characterizes the final recession month. By construction, we now have that a peak can only be called after a trough and vice versa. By considering the recent state probabilities of the initial estimation date, December 1976, we determine that our evaluation window begins in an expansion and as such begin looking for a peak. In Table 4 the initial peak dates obtained from the method outlined above for $\tau = 0.65$ and

$\tau = 0.8$ are presented and reflect identification in real-time.

Table 4: Comparison peak dates from the DFMS specifications with the NBER recessions.

$\tau = 0.65$

Peak date			*Ann. date*		
Base	**aGASX**	**NBER**	**Base**	**aGASX**	**NBER**
0 (1)	1 (0)	Jan 1980	0	0	Jun 1980
0 (0)	0 (0)	Jul 1981	-1	-2	Jan 1982
-1 (-2)	-1 (-1)	Jul 1990	-4	-4	Apr 1991
-1 (-2)	-1 (-2)	Mar 2001	-1	-4	Nov 2001
-3 (0)	-5 (-4)	Dec 2007	-7	-12	Dec 2008

$\tau = 0.8$

Peak date			*Ann. date*		
Base	**aGASX**	**NBER**	**Base**	**aGASX**	**NBER**
0 (1)	1 (0)	Jan 1980	0	0	Jun 1980
0 (0)	1 (0)	Jul 1981	-1	0	Jan 1982
-1 (-2)	-1 (-1)	Jul 1990	-4	-4	Apr 1991
-1 (-2)	-1 (-2)	Mar 2001	-1	-4	Nov 2001
-5 (0)	-6 (-4)	Dec 2007	-5	-10	Dec 2008

Note: This table contains the monthly differences in obtained initial peak dates of the base model and aGASX extension with the NBER database. The differences in parentheses reflect the dating at the final estimation date March 2020. Peaks are constructed from the smoothed contraction state probabilities using a threshold of $\tau = 0.65$ (top) and $\tau = 0.8$ (bottom). The NBER turning points and their respective announcement dates are obtained from https://www.nber.org/cycles.html.

In Table 4 we observe that the DFMS models are able to match or precede the peak announcement (ann.) dates of the NBER, regularly by a substantial margin. Specifically, the aGASX specification is able to date the 2001 and 2008 peaks three and five months earlier respectively for both thresholds, in line with the probability differences found in Figure 4. For the most recent recession of 2008 the aGASX specification is even able to signal the peak a year before the announcement by the NBER using the threshold $\tau = 0.65$. For the first three recessions we note comparable performance of the base model and the extension, but note that both are able to match or precede the NBER announcements. However, because the dating procedure is done at each point in time with the available data (i.e. for each 'column'), it might be that at later estimation dates different turning points are established than before. The values in parentheses in Table 4 reflect the peak dates as established at the final estimation date March 2020. Here we observe that the DFMS specifications make some adjustments as more data becomes available, unlike the NBER which has not made any revisions since the inception of their dating method. For example, the 2008 recession peak

is dated closer to the date established by the NBER at the final estimation date for both specifications. However, using our dating rule the aGASX model still dates it earlier than the NBER. This is consistent with Figure 4, where we observe for the aGASX specification that even for estimation in 2009 that the smoothed contraction state probabilities are elevated above a half several months before the NBER peak date.

With regards to the troughs corresponding to the peaks of Table 4, we confirm the results of Chauvet and Piger (2008), in that the DFMS model in general is able to call troughs much earlier than the NBER. However, no noteworthy differences between the base DFMS model and the aGASX specification are found, with only a minor advantage for the 2001 trough for the base model for $\tau = 0.8$. This is somewhat unsurprising as the transition probability p^{11} plays a much larger role here than p_t^{01} and is set the same for both specifications. For this reason and for brevity the troughs corresponding to the peaks of Table 4 can be found in Appendix Table D.1. We do note that the 2001 recession trough is initially put at a much later date than the NBER date for both specifications for $\tau = 0.65$ and undergoes large revisions as more data becomes available. This issue for the 2001 recession is also discussed in Chauvet and Piger (2008) and is presumably, as indicated before, related to a jobless recovery. Opting for a different measure of employment, such as civilian employment instead of the nonfarm payrolls used here and by TCB, is found by Doz et al. (2020) to remedy this specific dating delay.

We conclude that qualitatively our findings in real-time closely match those of our ex-post analysis. This entails that the aGASX specification that allows for a time-varying p^{01} using both exogenous and endogenous information provides clear benefits over the base DFMS model. This includes a significant increase in binary signaling ability of the filtered and predicted state probabilities for the NBER recessions, as measured by the AUC, and higher contraction state probabilities at the start of recessions. Additionally, we find that when converting the smoothed state probabilities to peaks that the DFMS specifications may be able to call turning points as soon or sooner than the NBER. Some adjustments of the turning point dates, when more data and revisions become available, are however noted. Moreover, the aGASX specification is able to date the two most recent peaks associated with the 2001 and 2008 recessions three and five months ahead of the base model respectively.

4 Concluding remarks

In this paper we extend the DFMS model by Diebold and Rudebusch (1996) by allowing for time-varying transition probabilities. Specifically, we propose a score-driven framework to guide the time-evolution of the transition probabilities endogenously by building upon the techniques found in Bazzi et al. (2017). Here we additionally allow for relevant exogenous economic drivers, as suggested by Diebold et al. (1994) and Filardo (1994). Furthermore, we consider a simplified analog of the recently developed aGAS framework by Blasques et al. (2019), which considers the recent alignment of scores in determining the magnitude of the parameter update. In an empirical application, using the components of TCB's CEI from 1959 until 2020 and the term spread as an exogenous input, we find that the proposed method can significantly improve the real-time signaling power for NBER recessions.

References

Aastveit, K.A., A.S. Jore, and F. Ravazzolo (2016). Identification and real-time forecasting of Norwegian business cycles. *International Journal of Forecasting* **32**, 283–292.

Bazzi, M. et al. (2017). Time-Varying Transition Probabilities for Markov Regime Switching Models. *Journal of Time Series Analysis* **38**, 458–478.

Berge, T.J. and Ò. Jordà (2011). Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics* **3**, 246–77.

Bernanke, B. (2004). The great moderation. *Speech available at https://www.federalreserve.gov/boarddocs/speeches/2004/20040220/.*

Blasques, F., C. Francq, and S. Laurent (2020). A New Class of Robust Observation-Driven Models. *Tinbergen Institute Discussion Papers* **20-073/III**.

Blasques, F., P. Gorgi, and S. Koopman (2019). Accelerating score-driven time series models. *Journal of Econometrics* **212**, 359–376.

Camacho, M., G. Perez-Quiros, and P. Poncela (2018). Markov-switching dynamic factor models in real time. *International Journal of Forecasting* **34**, 598–611.

Carstensen, K. et al. (2020). Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model: An application to the German business cycle. *International Journal of Forecasting* **36**, 829–850.

Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review* **39**, 969–996.

Chauvet, M. and J.D. Hamilton (2006). Dating business cycle turning points. *Contributions to Economic Analysis*. Ed. by C. Milas, P. Rothman, and D. van Dijk. Vol. **276**. Elsevier, 1–54.

Chauvet, M. and J. Piger (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics* **26**, 42–49.

Chauvet, M. and Z. Senyuz (2016). A dynamic factor model of the yield curve components as a predictor of the economy. *International Journal of Forecasting* **32**, 324–343.

Creal, D., S.J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**, 777–795.

Croushore, D. and T. Stark (2003). A real-time data set for macroeconomists: Does the data vintage matter? *Review of Economics and Statistics* **85**, 605–617.

Diebold, F.X., J.-H. Lee, and G.C. Weinbach (1994). Regime switching with time-varying transition probabilities. *Business Cycles: Durations, Dynamics, and Forecasting* **1**, 144–165.

Diebold, F.X. and G.D. Rudebusch (1996). Measuring Business Cycles: A Modern Perspective. *Review of Economics and Statistics* **78**, 67–77.

Doz, C., L. Ferrara, and P.-A. Pionnier (2020). Business cycle dynamics after the Great Recession: An extended Markov-Switching Dynamic Factor Model. *OECD Statistics Working Papers* **2020/01**.

Durland, J.M. and T.H. McCurdy (1994). Duration-dependent transitions in a Markov model of US GNP growth. *Journal of Business & Economic Statistics* **12**, 279–288.

Eo, Y. and C.-J. Kim (2016). Markov-switching models with evolving regime-specific parameters: Are postwar booms or recessions all alike? *Review of Economics and Statistics* **98**, 940–949.

Eo, Y. and J. Morley (2019). Why has the US economy stagnated since the Great Recession? *Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3077864*.

Estrella, A. and F.S. Mishkin (1998). Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics* **80**, 45–61.

Filardo, A.J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics* **12**, 299–308.

Groshen, E.L. and S. Potter (2003). Has structural change contributed to a jobless recovery? *Current issues in Economics and Finance* **9**.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357–384.

— (2011). Calling recessions in real time. *International Journal of Forecasting* **27**, 1006–1026.

Hanley, J.A. and B.J. McNeil (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839–843.

Harding, D. and A. Pagan (2003). A comparison of two business cycle dating methods. *Journal of Economic Dynamics and Control* **27**, 1681–1690.

Harvey, A.C. (2013). Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series. Vol. **52**. Cambridge University Press.

Huang, Y.-F. and R. Startz (2020). Improved recession dating using stock market volatility. *International Journal of Forecasting* **36**, 507–514.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics* **60**, 1–22.

Kim, C.-J. and C.R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics* **80**, 188–201.

— (1999). State-space models with regime switching: classical and Gibbs-sampling approaches with applications. *MIT Press Books* **1**.

Koopman, S.J., A. Lucas, and M. Scharth (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics* **98**, 97–110.

Liu, W. and E. Mönch (2016). What predicts US recessions? *International Journal of Forecasting* **32**, 1138–1150.

Marcellino, M. (2006). Leading indicators. *Handbook of Economic Forecasting*. Ed. by G. Elliott, C.W. Granger, and A. Timmermann. Vol. **1**. Elsevier, 879–960.

McConnell, M.M. and G. Perez-Quiros (2000). Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review* **90**, 1464–1476.

Ng, S. and J.H. Wright (2013). Facts and challenges from the great recession for forecasting and macroeconomic modeling. *Journal of Economic Literature* **51**, 1120–54.

Rudebusch, G.D. and J.C. Williams (2009). Forecasting recessions: the puzzle of the enduring power of the yield curve. *Journal of Business & Economic Statistics* **27**, 492–503.

Sichel, D.E. (1994). Inventories and the three phases of the business cycle. *Journal of Business & Economic Statistics* **12**, 269–277.

Watanabe, T. et al. (2003). Measuring business cycle turning points in Japan with a dynamic Markov switching factor model. *Monetary and Economic Studies* **21**, 35–68.

Wu, J.C. and F.D. Xia (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking* **48**, 253–291.

Appendix A Prediction-update recursion

The Hamilton prediction step is given as

$$Pr(S_t = j, S_{t-1} = i|I_{t-1}) = p_t^{ij} Pr(S_{t-1} = i|I_{t-1}) \tag{A.1}$$

and the Kalman prediction steps can be written as

$$\boldsymbol{\zeta}_{t|t-1}^{ij} = \mathbf{d}_{S_t=j} + \mathbf{T}\boldsymbol{\zeta}_{t-1|t-1}^{i}, \tag{A.2}$$

$$\mathbf{P}_{t|t-1}^{ij} = \mathbf{T}\mathbf{P}_{t-1|t-1}^{i}\mathbf{T}' + \mathbf{Q}. \tag{A.3}$$

Using the observation at time t, we can update our Markov state probability and the mean and covariance matrix of the state vector. The Hamilton update step is given by

$$Pr(S_t = j, S_{t-1} = i|I_t) = \frac{p_t^{ij} Pr(S_{t-1} = i|I_{t-1})\phi_t^{ij}(\mathbf{y}_t)}{\sum_{l,k\in\{0,1\}} p_t^{lk} Pr(S_{t-1} = l|I_{t-1})\phi_t^{lk}(\mathbf{y}_t)} \tag{A.4}$$

and the Kalman update steps are given by

$$\boldsymbol{\zeta}_{t|t}^{ij} = \boldsymbol{\zeta}_{t|t-1}^{ij} + \mathbf{P}_{t|t-1}^{ij}\mathbf{Z}'(\mathbf{Z}\mathbf{P}_{t|t-1}^{ij}\mathbf{Z}')^{-1}(\mathbf{y}_t - \mathbf{Z}\boldsymbol{\zeta}_{t|t-1}^{ij}), \tag{A.5}$$

$$\mathbf{P}_{t|t}^{ij} = (\mathbf{I} - \mathbf{P}_{t|t-1}^{ij}\mathbf{Z}'(\mathbf{Z}\mathbf{P}_{t|t-1}^{ij}\mathbf{Z}')^{-1}\mathbf{Z})\mathbf{P}_{t|t-1}^{ij}. \tag{A.6}$$

To prevent an increasing length of path dependence, we collapse the states, a method first proposed in this context by Kim (1994). For the state probability this is relatively straightforward and given as

$$Pr(S_t = j|I_t) = \sum_{i\in\{0,1\}} Pr(S_t = j, S_{t-1} = i|I_t). \tag{A.7}$$

The collapse step for the factor and its covariance are slightly more involved and given as

$$\boldsymbol{\zeta}_{t|t}^{j} = \frac{\sum_{i\in\{0,1\}} Pr(S_t = j, S_{t-1} = i|I_t)\boldsymbol{\zeta}_{t|t}^{ij}}{Pr(S_t = j|I_t)}, \tag{A.8}$$

$$\mathbf{P}^{j}_{t|t} = \frac{\sum_{i\in\{0,1\}} Pr(S_t = j, S_{t-1} = i|I_t)(\mathbf{P}^{ij}_{t|t} + (\boldsymbol{\zeta}^{j}_{t|t} - \boldsymbol{\zeta}^{ij}_{t|t})(\boldsymbol{\zeta}^{j}_{t|t} - \boldsymbol{\zeta}^{ij}_{t|t})')}{Pr(S_t = j|I_t)}. \tag{A.9}$$

We now have all the components to begin the next iteration, i.e. start the prediction steps for $t+1$ and so on. To obtain the parameter estimates we maximize the associated approximated log-likelihood, which is here given as

$$\text{Log}L \approx \sum_{t=1}^{T} \text{Log}[\sum_{i,j\in\{0,1\}} p^{ij}_t Pr(S_{t-1} = i|I_{t-1})\phi^{ij}_t(\mathbf{y}_t)]. \tag{A.10}$$

In terms of initialization, the contraction state probability at time $t = 0$ is set to 0, with virtually identical results obtained if this is treated as a parameter which is optimized alongside the other parameters. For the specifications that include a time-varying transition probability, the initial values are set to those obtained from the base model with all transition probabilities being constant. Optimization is done using standard quasi-Newton methods and standard errors are obtained from the inverse hessian. Finally, for the initialization of the factor, a diffuse prior with zero mean is considered at time $t = 0$.

Appendix B Data

B.1 Employees on nonfarm payrolls and industrial production

The vintages for US total nonfarm payroll employment and the US industrial production index between December 1976 up to and including December 1988 are retrieved from the real-time dataset for macroeconomists from the Philadelphia Fed, see Croushore and Stark (2003). Data is reported after the monthly publication of the previous month's employment and industrial production, such that in month t we have vintages up to and including month $t-1$ for these variables. For January 1989 through March 2020 the vintages are obtained from TCB. Here data up to January 2001 are reported at start of the month before publication of the previous month's employment, while the remainder of the data reports values later in the month after the publication of last month's values. Because no new information becomes available between the end of each month and the beginning of its subsequent month for these variables, we shift the data for this period one month, such that it is available one month earlier. For December 2001 two observations are reported, one in the first and one in the

third week, such that the time-shift does result in missing data for this month.

B.2 Manufacturing and trade sales

Similar to Chauvet and Piger (2008), data for the vintages from December 1976 up to and including December 1988 are obtained from *Business Conditions Digest*. For the remaining period of January 1989 through March 2020, data is retrieved from TCB. For both datasets the reporting moment in each month t is after the publication of the observation for month $t-3$.

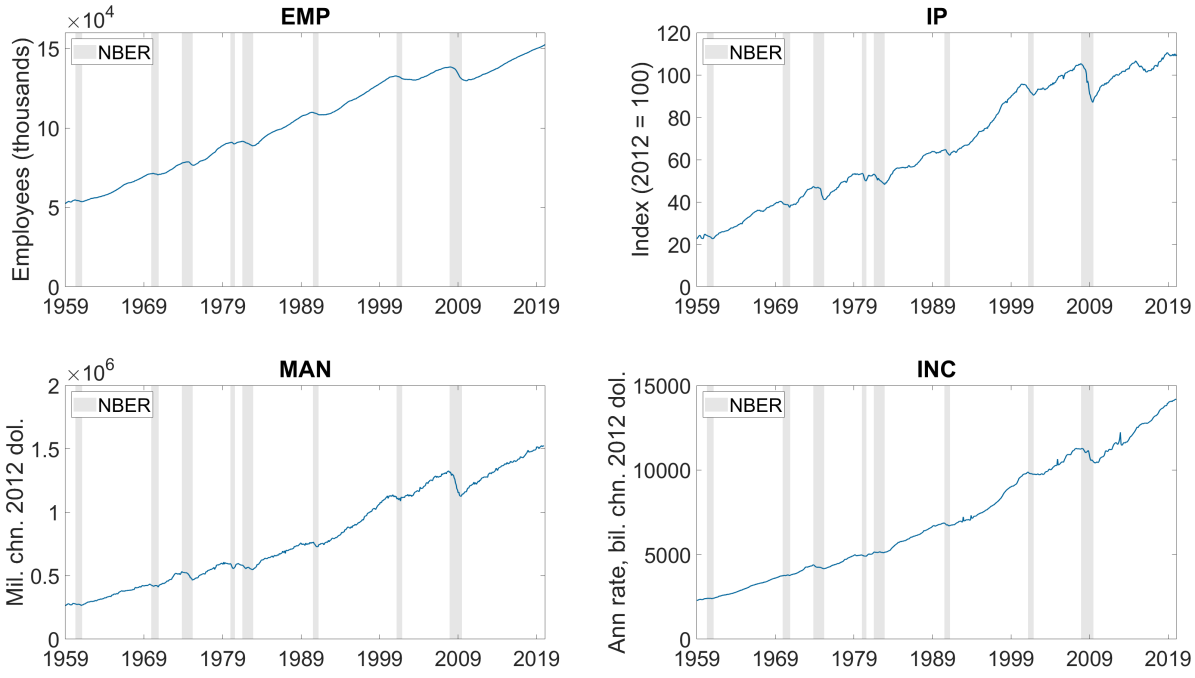
B.3 Personal income less transfer payments

Similar to Chauvet and Piger (2008), data for the vintages from November 1976 up to and including December 1988 are obtained from *Business Conditions Digest*. For the remaining period of January 1989 through March 2020, data is retrieved from TCB. For the first part of the data the reporting moment in month t is just after the publication of the observation of month $t-1$, while for the second part the reporting moment precedes this publication. Therefore, to synchronise the datasets, and to ensure we only include data up to the third week of each month, we lag the observations of the first dataset with one month. Now we have that for each month t we have vintages up to and including month $t-2$, with data available for the real-time analysis from December 1976 on. Furthermore, missing data are encountered for January 1997, where data starts in December 1992. The resulting missing log growth rates are filled with a constructed series, which is obtained using the same steps Chauvet and Piger (2008) employ to construct their personal income data after 1995. Specifically, nominal transfer payments are subtracted from nominal personal income and finally divided by the ratio of nominal to real disposable income. These components are collected from the ALFRED database, maintained by the Federal Reserve Bank of St. Louis. For nominal transfer payments, data from Economic Indicators, Business Statistics and the Survey of Current Business are considered.

Due to the presence of some large outliers in this variable, see Appendix Figure B.1, particularly in December 2012, an adjustment is made during estimation of the DFMS model specifications. Specifically, we first winsorize the (logarithmic) monthly growth rates of this

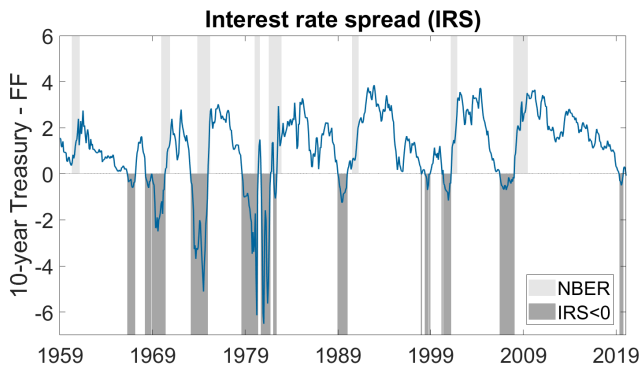
variable on both sides at a 1 percent level and estimate the model parameters. Second, we return to the original data and search for a maximum in a sensible neighbourhood of this winsorized optimum. This approach therefore excludes outlier solutions and prevents structural breaks in the parameter estimates in real-time when an outlier enters the estimation window.

Figure B.1: Final vintage US employees on nonfarm payrolls (EMP), the index of industrial production (IP), manufacturing and trade sales (MAN) and personal income less transfer payments (INC), January 1959 - February 2020.



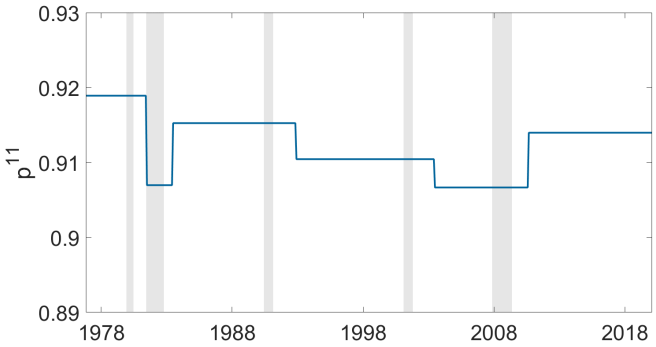
Note: Graphical illustration of the raw data, whereby the shaded areas reflect the recession periods as determined by the NBER.

Figure B.2: US interest rate spread, an indicator for its negativity and the NBER recessions, January 1959 - February 2020.



Note: The plot depicts the IRS constructed by subtracting the Federal Funds (FF) rate from the 10-year Treasury rate. The dark shaded areas in the negative domain reflect the periods of a negative IRS. Finally, the light shaded areas in the positive domain reflect the recession periods as determined by the NBER.

Figure B.3: Dynamic estimates of p^{11} from completed NBER recessions.



Note: This figure depicts the evolution of the estimate of p^{11} when estimated in an expanding window fashion from completed NBER recessions. The estimate is updated the month after a trough announcement and boils down to dividing the total number of recession months minus the number of recessions by the total number of recession months. Finally, the shaded areas reflect the recession periods as determined by the NBER.

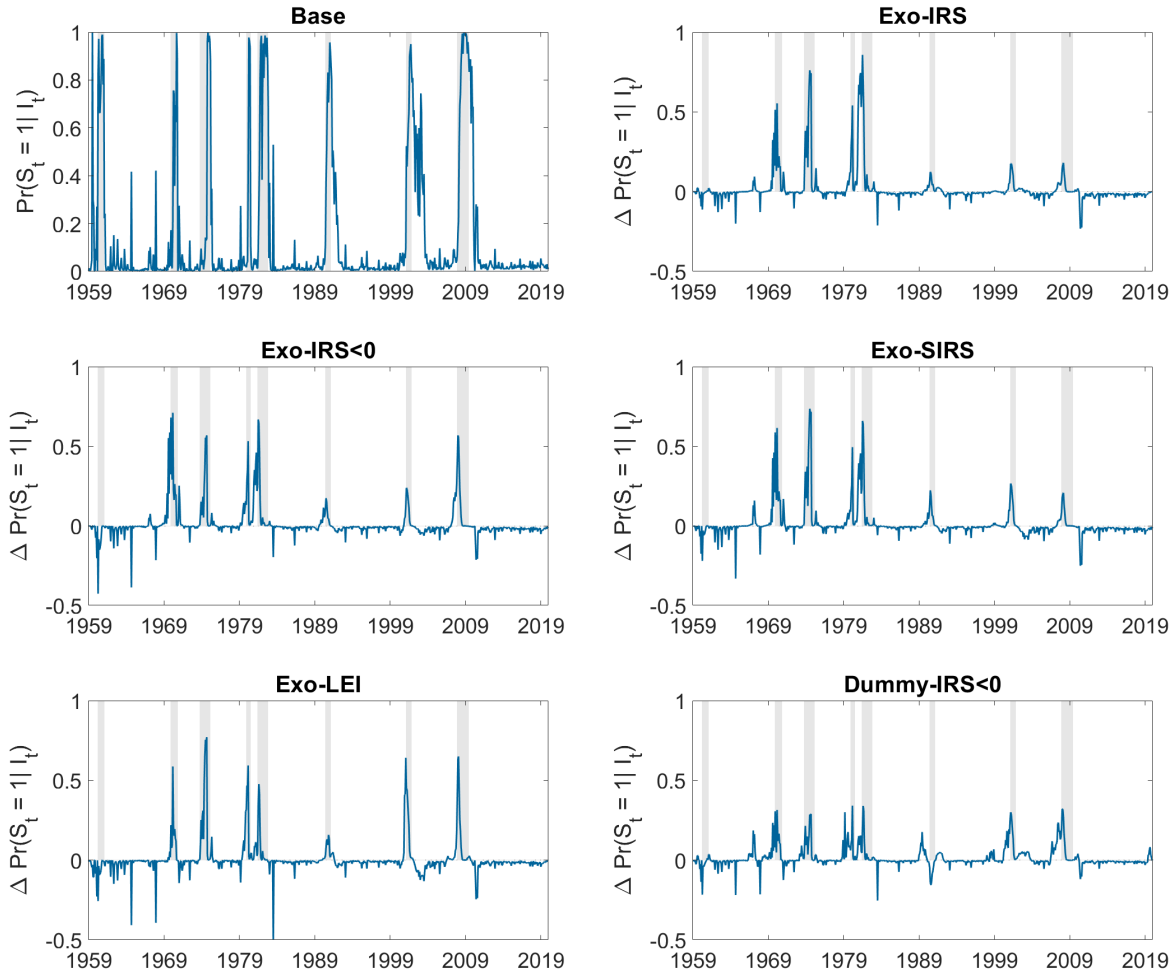
Appendix C Full-sample

Table C.1: Remaining parameter estimates for the DFMS model with variants that consider a time-varying p^{01}.

	Base	**GAS**	**aGAS**	**Exo**	**GASX**	**aGASX**
α_0	0.094	0.091	0.091	0.093	0.093	0.092
	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)
α_1	-0.097	-0.095	-0.111	-0.094	-0.089	-0.092
	(0.023)	(0.020)	(0.021)	(0.023)	(0.022)	(0.025)
ϕ	0.546	0.556	0.549	0.552	0.552	0.556
	(0.050)	(0.049)	(0.047)	(0.051)	(0.048)	(0.053)
λ_{IP}	2.298	2.310	2.305	2.302	2.305	2.308
	(0.114)	(0.115)	(0.116)	(0.115)	(0.115)	(0.116)
λ_{MAN}	1.907	1.914	1.915	1.909	1.910	1.910
	(0.109)	(0.109)	(0.110)	(0.109)	(0.109)	(0.109)
λ_{INC}	1.326	1.331	1.335	1.328	1.328	1.329
	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)
$\sigma^2\varepsilon_{EMP}$	0.006	0.006	0.006	0.006	0.006	0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\sigma^2\varepsilon_{IP}$	0.370	0.369	0.370	0.369	0.369	0.369
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
$\sigma^2\varepsilon_{MAN}$	0.705	0.704	0.704	0.705	0.704	0.705
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
$\sigma^2\varepsilon_{INC}$	0.268	0.268	0.268	0.268	0.268	0.268
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
θ_{EMP}	-0.486	-0.478	-0.475	-0.480	-0.479	-0.480
	(0.069)	(0.068)	(0.066)	(0.070)	(0.069)	(0.069)
θ_{IP}	0.156	0.154	0.151	0.155	0.157	0.157
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
θ_{MAN}	-0.234	-0.235	-0.236	-0.235	-0.235	-0.235
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
θ_{INC}	-0.110	-0.112	-0.112	-0.111	-0.111	-0.111
	(0.038)	(0.038)	(0.037)	(0.038)	(0.038)	(0.038)
σ^2_η	0.015	0.014	0.014	0.014	0.015	0.014
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
LogL	-1913.3	-1907.3	-1905.4	-1900.2	-1897.9	-1897.3
k	16	18	19	18	19	20
AIC	3858.5	3850.7	3848.8	3836.5	3833.9	3834.7

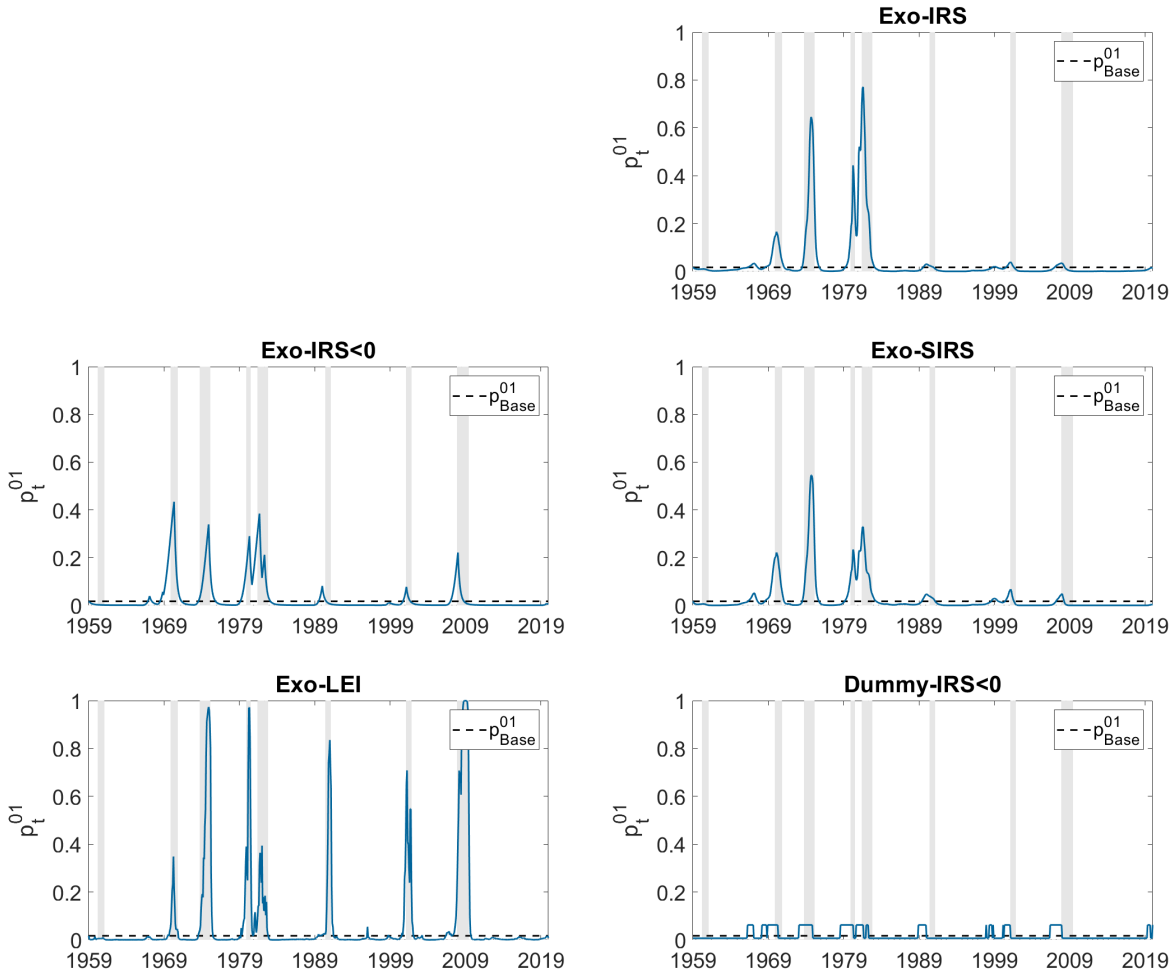
Note: This table presents the leftover parameter estimates for the base model and the extensions that allow for a time-varying transition probability p_t^{01} using monthly growth rates for EMP, IP, MAN and INC over the period January 1959-February 2020. Standard errors are displayed in parentheses and k denotes the number of estimated parameters.

Figure C.1: Robustness exogenous variable choice, filtered state probabilities.



Note: Filtered contraction state probabilities ($Pr(S_t = 1|I_t)$) for the base model are depicted in the top left panel. The remaining figures depict the filtered contraction state probabilities of the extensions minus those of the base model, denoted by $\Delta Pr(S_t = 1|I_t)$. Exo-$(\cdot)$ denotes the autoregressive specification that only makes use of $(\cdot)$ to drive p_t^{01}, whereby SIRS denotes the standardized IRS obtained by dividing the spread by the sum of the long and short-term rates. For the LEI, the monthly logaritmic growth rates are used. Dummy-IRS<0 denotes a specification without autoregressive dynamics that simply considers two different levels of p^{01} depending on the sign of the IRS. Finally, the shaded areas reflect the recession periods as determined by the NBER.

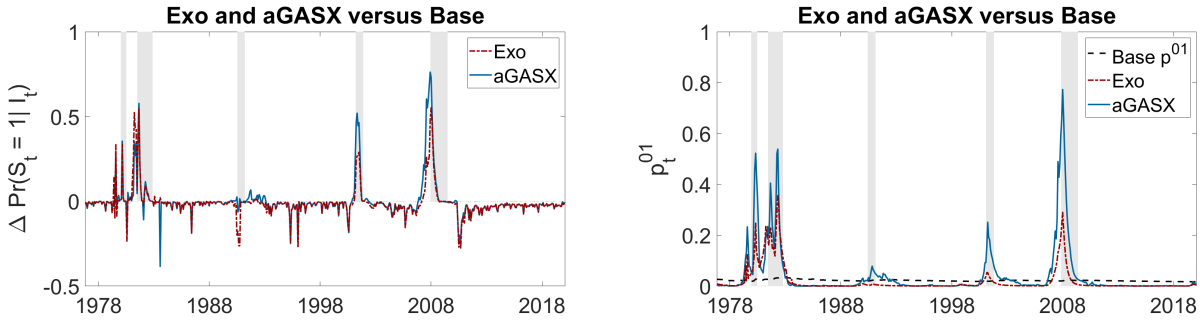
Figure C.2: Robustness exogenous variable choice, transition probability p_t^{01}.



Note: This figure displays the evolution of the transition probability p_t^{01} over time for the alternative DFMS model specifications. The dotted line represents the constant p^{01} estimated by the base model. Exo-$(\cdot)$ denotes the autoregressive specification that only makes use of $(\cdot)$ to drive p_t^{01}, whereby SIRS denotes the standardized IRS obtained by dividing the spread by the sum of the long and short-term rates. For the LEI, the monthly logaritmic growth rates are used. Dummy-IRS<0 denotes a specification without autoregressive dynamics that simply considers two different levels of p^{01} depending on the sign of the IRS. Finally, the shaded areas reflect the recession periods as determined by the NBER.

Appendix D Real-time

Figure D.1: Comparison real-time filtered performance for the Exo and the aGASX specification.



Note: The left figure displays the differences of the real-time filtered contraction state probabilities of the Exo and aGASX specification with those of the base model (Base minus extension). Furthermore, the right figure displays the evolution of the transition probabilities p_t^{01}. Finally, the shaded areas reflect the recession periods as determined by the NBER.

Table D.1: Comparison trough dates from the DFMS specifications with the NBER recessions.

$\tau = 0.65$

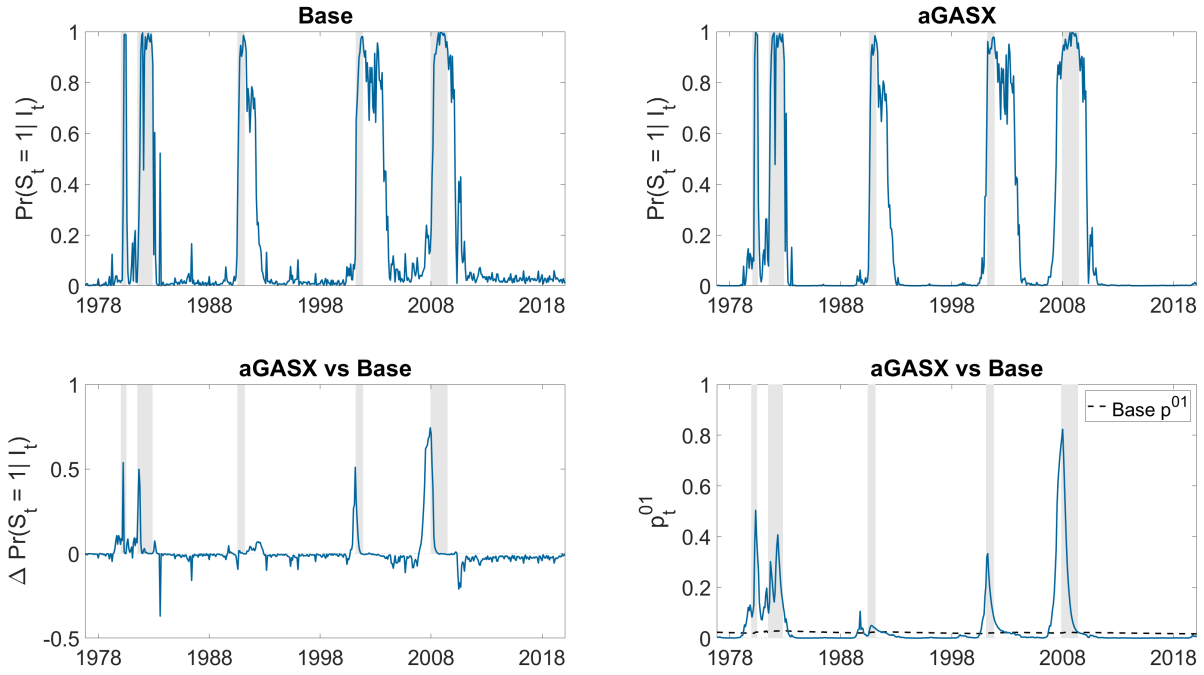
Trough date			*Ann. date*		
Base	**aGASX**	**NBER**	**Base**	**aGASX**	**NBER**
-1 (-1)	-1 (0)	Jul 1980	-9	-9	Jul 1981
-1 (-1)	-1 (-1)	Nov 1982	-5	-5	Jul 1983
1 (1)	1 (1)	Mar 1991	-14	-14	Dec 1992
21 (5)	20 (8)	Nov 2001	5	4	Jul 2003
4 (4)	5 (4)	Jun 2009	-5	-5	Sep 2010

$\tau = 0.8$

Trough date			*Ann. date*		
Base	**aGASX**	**NBER**	**Base**	**aGASX**	**NBER**
-1 (-1)	-1 (-1)	Jul 1980	-9	-9	Jul 1981
-1 (-1)	-1 (-1)	Nov 1982	-5	-5	Jul 1983
1 (0)	1 (0)	Mar 1991	-16	-16	Dec 1992
1 (2)	4 (3)	Nov 2001	-15	-12	Jul 2003
4 (3)	4 (3)	Jun 2009	-7	-7	Sep 2010

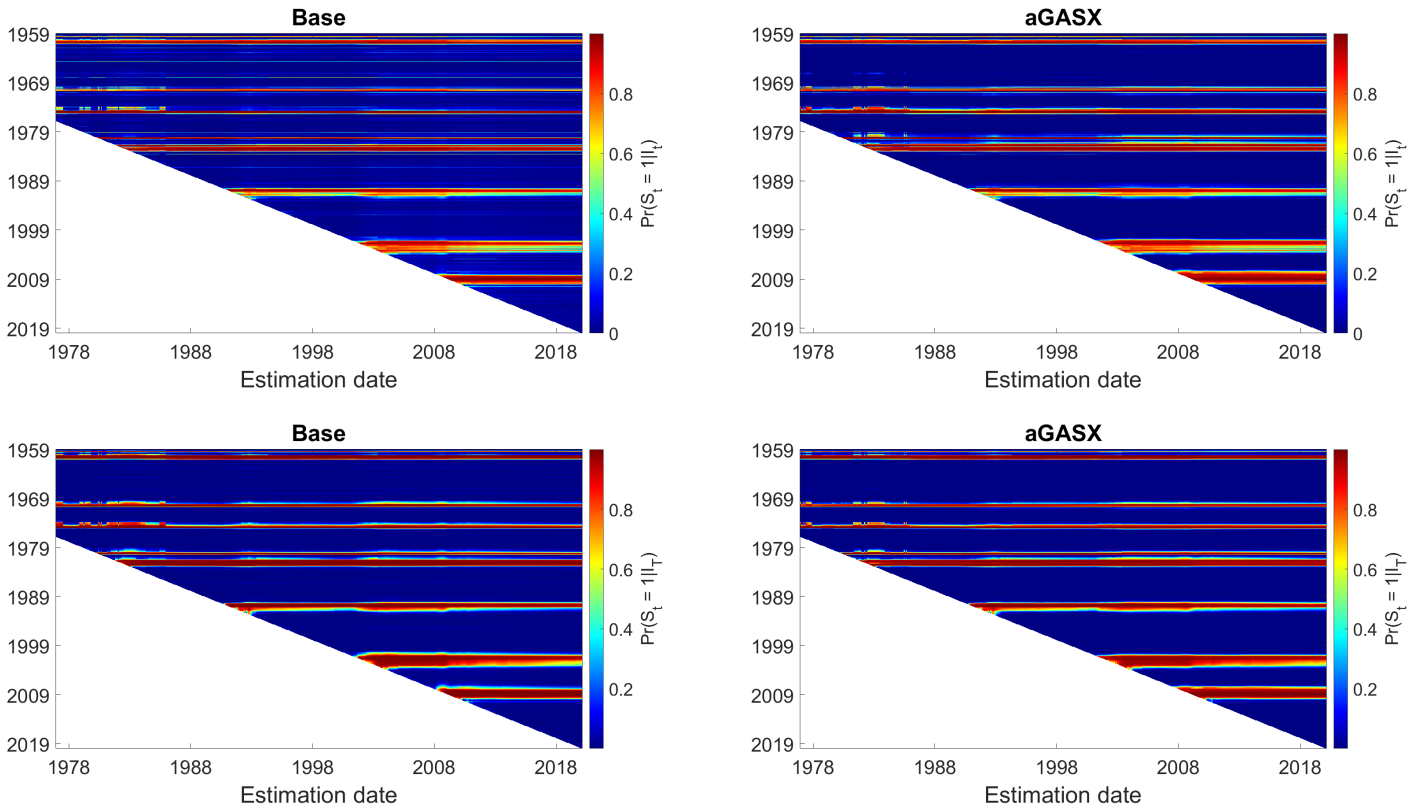
Note: This table contains the monthly differences in obtained initial trough dates of the base model and aGASX extension with the NBER database. The differences in parentheses reflect the dating at the final estimation date March 2020. Troughs are constructed from the smoothed contraction state probabilities using a threshold of $\tau = 0.65$ (top) and $\tau = 0.8$ (bottom). The NBER turning points and their respective announcement dates are obtained from https://www.nber.org/cycles.html.

Figure D.2: Comparison real-time filtered performance for the base and the aGASX specification using final vintages.



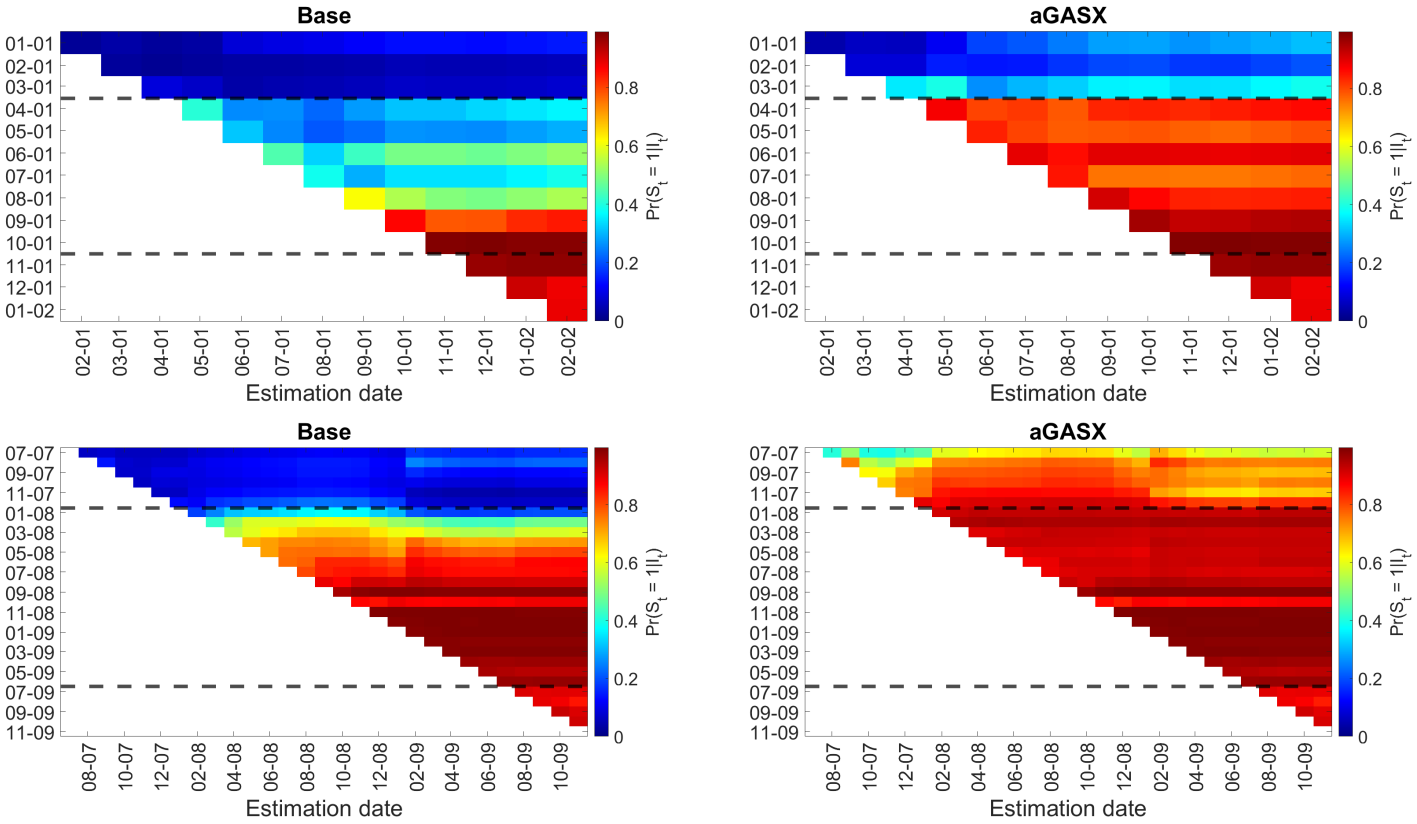
Note: The top left and top right figures display the filtered state probabilities for the base and aGASX model respectively and the bottom left presents their difference (Base minus aGASX) using final vintage data. Furthermore, the bottom right figure displays the corresponding evolution of the transition probability p_t^{01}. To prevent large movement of p_t^{01} for some of the early vintages when not much data is available, we constrain the autoregressive parameter b to lie between 0.8 and 1. For all points in time a maximum is found here that does not lie on the boundary. Finally, the shaded areas reflect the recession periods as determined by the NBER.

Figure D.3: Complete history of real-time filtered and smoothed state probabilities for the base DFMS model and the aGASX extension.



Note: This figure depicts the filtered (top) and smoothed (bottom) state probabilities for the Base (left) and aGASX (right) specification. The x-axis contains the estimation date and the y-axis the sample date for which a probability is constructed.

Figure D.4: History of real-time filtered state probabilities for the base DFMS model and the aGASX extension around the 2001 and 2008 recession.



Note: This figures depicts the real-time filtered state probabilities around the 2001 (top) and 2008 (bottom) recessions for the base (left) and aGASX specification (right). The x-axis contains the estimation date and the y-axis the sample date for which a probability is constructed. Finally, the black dotted lines reflect the turning points as determined by the NBER.