TI 2020-057/VI
Tinbergen Institute Discussion Paper

Accelerating Peak Dating in a Dynamic Factor Markov-Switching Model

Revision: November 29, 2022

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

1 Econometric Institute, Erasmus University Rotterdam

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

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

Bram van Os[†] and Dick van Dijk[‡]

Econometric Institute, Erasmus University Rotterdam

November 19, 2022

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 business cycle peaks. Time-variation of the transition probabilities is brought about endogenously using the 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 we are able to date the 2001 and 2008 recession peaks four and two months after the peak date, which is four and ten months before the NBER.

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

*We gratefully acknowledge helpful comments and suggestions from Monica Billio, Francis X. Diebold, Laurent Ferrara, Rutger-Jan Lange, 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, such as output, employment and the like. Dating the turning points between the phases of this cycle is of utmost importance to policy-makers, firms and investors. Especially the peaks (marking the transitions from expansions to contractions) of the cycle are of crucial interest for downside risk management (Adrian et al., 2019; Caldara et al., 2020). The dynamic factor Markov-switching (DFMS) model proposed by Diebold and Rudebusch (1996) has proven to be a powerful framework to measure the cycle. This model extracts a latent business cycle factor from multiple coincident variables and allows the dynamics of the factor to be regime-dependent using a hidden Markov process. Chauvet and Piger (2008) find that the DFMS model is able to call the troughs (marking the transitions from contractions to expansions) of the cycle faster in real-time than the National Bureau of Economic Research (NBER). However, they find no such improvements in timeliness for the peaks.

In this paper, we address the latter issue and extend the DFMS model with the aim of accelerating peak dating. With this purpose in mind, we allow for a time-varying transition probability (TVTP) to switch from an expansion to a contraction phase. This probability is a key ingredient for peak dating and assuming it to be time-invariant may be overly restrictive. For example, one expects the probability of entering a new recession to vary depending on economic fundamentals, see e.g. Diebold et al. (1994). To bring about time-variation, we propose an autoregressive structure driven by the log-likelihood score and additionally by leading indicators (LIs). The resulting Generalized Autoregressive Score with eXogenous variables (GASX) model thus combines the ideas of endogenous and exogenous drivers of the transition probabilities from Durland and McCurdy (1994) and Filardo (1994), respectively.

We apply the GASX approach to date US business cycle peaks, based upon the four components of The Conference Board's (TCB) Coincident Economic Index (CEI) for the period 1959-2020. We consider both an ex-post full-sample analysis using revised data and an ex-ante real-time analysis using data vintages available from December 1976 until March 2020. For the exogenous input we consider an indicator for a negative term spread, which is generally considered to be one of the most prominent LIs, see e.g. Estrella and Mishkin (1998). We find in both ex-post and ex-ante analyses that the GASX model significantly improves upon the DFMS model with static transition probabilities in terms of signaling recessions. Additionally, by converting real-time smoothed state probabilities to turning points, the GASX 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 two months after the peak date. This is four and ten months before their NBER announcements and a gain of three and five months over the base DFMS model. In line with the reputation as a powerful LI, we find most improvements to stem from the use of the term spread. Notably, GAS dynamics are found to further amplify correct peak signals. The combination of both drivers in the GASX model is therefore particularly attractive.

Our paper contributes to the vast literature on US business cycle measurement and dating turning points (e.g. Boldin, 1994; Berge and Jordà, 2011; Hamilton, 2011; Stock and Watson, 2014; Doz et al., 2020). In particular, we build upon the DFMS framework of Diebold and Rudebusch (1996), which combines the dynamic factor structure of Stock and Watson (1989, 1993, 2005, 2010) with the Markov-switching (MS) approach of Hamilton (1989). This setup captures both the co-movement in multiple coincident series and the regime-dependence that characterize the business cycle (Burns and Mitchell, 1946). The factor structure allows for a larger information input relative to univariate MS models, which are already able to effectively date turning points (Layton, 1996; Layton and Katsuura, 2001; Chauvet and Piger, 2003). Furthermore, Chauvet and Piger (2008) find that the DFMS model fares well against the non-parametric dating method of Harding and Pagan (2003), which aggregates the approach of Bry and Boschan (1971) for multiple series. Both methods are shown to provide faster real-time dating of the troughs compared to the NBER. Similar results have been found for other countries (Watanabe et al., 2003; Aastveit et al., 2016; Carstensen et al., 2020).

Our paper is most closely related to the literature on macroeconomic MS models with TVTPs. One way to drive these TVTPs is to use exogenous information in the form of LIs. In particular, Filardo (1994) finds that TVTPs driven by LIs can aid in the identification of turning points for output growth. In the context of the DFMS model, Huang and Startz (2020) allow the transition probabilities to depend on the stock market volatility, while Chauvet and Senyuz (2016) add a set of yield curve variables. Both find that turmoil in financial markets often precedes economic recessions. Our proposed GASX model incorporates this insight by allowing the TVTPs to depend on the term spread, which has historically been among the best LIs (e.g. Estrella and Mishkin, 1998; Rudebusch and Williams, 2009; Ng and Wright, 2013; Liu and Mönch, 2016).

Alternatively, one may drive the TVTPs using only endogenous information. This can be done using duration dependence (Durland and McCurdy, 1994; Kim and Nelson, 1998), directly specifying the TVTPs as functions of dependent variables (Diebold et al., 1994; Caldara et al., 2020), or using a score-driven approach (Bazzi et al., 2017). The latter updates the TVTPs in the direction of the log-likelihood score as suggested by Creal et al. (2013) and Harvey (2013). This Generalized Autoregressive Score-driven (GAS) approach has favorable properties (Blasques et al., 2015), allows for standard likelihood estimation and is known to produce accurate filters in a variety of settings, see e.g. Koopman et al. (2016). Our GASX approach presents a multivariate version of the GAS setup of Bazzi et al. (2017) and combines it with exogenous LI information in a single framework.

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. 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, multiple common factors and higher lag orders are relatively straightforward but tedious. 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 constant factor loadings λ_i and idiosyncratic components $v_{i,t}$, such that the observation equation of the state space representation is given by

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

where $\mathbf{y}_t$ collects the $y_{i,t}$ in a $(N \times 1)$ vector, $\mathbf{Z}$ is a $(N \times N + 1)$ matrix of coefficients and $\boldsymbol{\zeta}_t$ is the $(N + 1 \times 1)$ state vector, that is,

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

We assume that the latent factor ψ_t follows a stationary AR(1) process with autoregressive parameter ϕ and intercept α_{S_t}, which depends on the state $S_t \in \{0, 1\}$ of a hidden inhomogeneous Markov process with dynamic transition probabilities $p_t^{ij} := \Pr(S_t = j|S_{t-1} = i)$ with $i, j \in \{0, 1\}$. For identification purposes we impose $\alpha_{S_t=0} > \alpha_{S_t=1}$, such that regime 0 (1) reflects an expansion (contraction) period. The variance of the factor innovations is denoted by σ_η^2. We assume stationary zero mean 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

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

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

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

and $\boldsymbol{\omega}_t$ denotes an $(N+1\times1)$ i.i.d. innovation vector which we assume to follow a multivariate standard normal distribution.

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 accelerating peak dating in mind we consider dynamic transition probabilities starting from an expansion period (p_t^{00} and $p_t^{01} = 1 - p_t^{00}$), but keep the probabilities starting from a contraction fixed ($p_t^{11} = p^{11}$ and $p_t^{10} = p^{10} = 1 - p^{11}$). For expositional purposes we formulate our model in terms of p_t^{01}, which reflects the probability to switch from an expansion to a contraction phase, such that we may informally label it as the (conditional) peak probability.

We build upon the general framework of Creal et al. (2013) and make use of the endogenous information available in the log-likelihood score to drive p_t^{01}. In addition, we also allow for exogenous information as suggested by Diebold et al. (1994) and Filardo (1994). To ensure that the transition probability remains in the unit interval we adopt the logistic link function. Specifically, we model the conditional peak probability p_t^{01} as

$$p_t^{01} = \frac{\exp(f_t)}{1 + \exp(f_t)}, \tag{5}$$

$$f_{t+1} = w + as_t + bf_t + cx_t, \tag{6}$$

$$s_t = g(H_t \nabla_t^f), \quad \nabla_t^f = \frac{\partial}{\partial f_t} \log p(y_t | \mathcal{I}_{t-1}), \tag{7}$$

where f_t reflects the log odds ratio of the transition probability p_t^{01} (i.e. $f_t = \log(p_t^{01}/(1-p_t^{01}))$, x_t is an exogenous variable known at time t, and w, $a > 0$, $b \in (-1, 1)$ and c are static parameters. Furthermore, s_t denotes the endogenous innovation term, which is composed of three components. First and foremost, ∇_t^f denotes the score with respect to f_t and is obtained by taking the derivative of the (approximate) predictive log density at time t, denoted by $\log p(\mathbf{y}_t|\mathcal{I}_{t-1})$. Here $\mathcal{I}_{t-1}$ denotes the information set containing all information available at time $t-1$. Second, H_t is a positive scaling factor known at time t. Third, $g(\cdot)$ denotes a strictly monotonic transformation with $g(0) = 0$. We note that the true p_t^{01} need not follow the evolution as outlined in (5)-(7). It is therefore perhaps more appropriate to view our setup as an intuitive filter that tracks the true time-varying parameter using a steepest ascent-type search. Our setup can be seen as a generalization of Bazzi et al. (2017), who show that score-driven TVTPs yield an effective filter for various data generating processes in a univariate setting.

Popular choices for H_t in the GAS literature are no scaling ($H_t = 1$) or powers of the Fischer matrix in order to account for the curvature of the likelihood. We consider a scaling that allows for more variation of p_t^{01}. Specifically, the score ∇_t^f contains the term $p_t^{01}(1-p_t^{01})$ because of the logistic link due to the chain-rule. This term is often close to 0, very much dampening and delaying movement in p_t^{01}. We therefore suggest to use $H_t = 1/(p_t^{01}(1-p_t^{01}))$

to remove this effect. Note that it is straightforward to show that this part would also disappear in the case of scaling with the square root Fischer information. The Fischer information, however, presents a large computational burden here. For $H_t = 1/(p_t^{01}(1-p_t^{01}))$ and using the likelihood approximation of Kim (1994) our scaled score $H_t\nabla_t^f$ is given by

$$H_t\nabla_t^f = \Pr(S_{t-1} = 0|\mathcal{I}_{t-1})\frac{\phi_t^{01}(\mathbf{y}_t|\mathcal{I}_{t-1}) - \phi_t^{00}(\mathbf{y}_t|\mathcal{I}_{t-1})}{p(\mathbf{y}_t|\mathcal{I}_{t-1})}, \tag{8}$$

where $\phi_t^{ij}(\mathbf{y}_t|\mathcal{I}_{t-1})$ denotes the multivariate normal density evaluated in $\mathbf{y}_t$ conditional on all information available at time $t-1$ and the states being i and j at time $t-1$ and t, respectively (see Appendix A for details). The scaled score $H_t\nabla_t^f$ thus considers the difference in likelihood between currently being in a contraction or an expansion phase when the previous period is assumed to be an expansion, weighted by the likelihood. This relative difference is then multiplied with the state probability $\Pr(S_{t-1} = 0|\mathcal{I}_{t-1})$, such that we update p_t^{01} more (less) during an expansion (contraction), precisely when the data tells us much (little) about the expansion dynamics.

The addition of the transformation $g(\cdot)$ in (7) can be seen as a score-equivalent special case of the more general quasi score-driven (QSD) framework of Blasques et al. (2022). The QSD framework outlines 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. By integrating s_t with respect to f_t the corresponding target function is obtained. The score-equivalence property refers to the fact that by construction $g(\cdot)$ maintains the update direction (i.e. $\text{sgn}(s_t) = \text{sgn}(\nabla_t^f)$). As a result, we preserve the local optimality properties as laid out in Blasques et al. (2015).

We argue that a transformation $g(\cdot)$ that decreases large (absolute) scores, essentially shrinking large updates, may be especially useful for forecasting. In our case in particular, we find empirically that the score in the DFMS model can produce a substantial number of outliers. This happens when the likelihood in the denominator in (8) becomes small. As a result the learning rate parameter a in (6) is underestimated, limiting the ability to drive the transition probability in a meaningful way. 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:

$$g(x) = \text{sgn}(x)\log(1 + |x|), \tag{9}$$

which is a monotonic antisymmetric function 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' $|x|$. This transformation therefore maintains the direction of the update, while shrinking the updates for large absolute scores. This in turn prevents over-updating of our transition probability when a very large score, possibly due to an outlier, occurs. Integrating the transformed score reveals we are now essentially applying a gradient update based on a flattened version of the log-likelihood, similar in spirit to the robust-GAS models proposed by Blasques et al. (2022), who consider for example Huber-type target functions. We argue that our choice of $g(\cdot)$ is perhaps the most obvious for our purposes, being a natural extension of the logarithm capable of handling both negative numbers and 0. For robustness, we include results in Appendix C.2 using the inverse hyperbolic sine ($g(x) = \log(x + \sqrt{x^2 + 1})$), a popular transformation in regression analysis used to curb extreme values. There we find near identical results as using the transformation in (9). In addition, we provide results when omitting the transformation ($g(x) = x$) or the scaling ($H_t = 1$) in Appendix C.3.

2.3 Estimation

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 due to the problem of path-dependence, meaning that the value of the factor at time t depends on all previous Markov states. 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 this history length is required to be at least one longer than the highest lag order 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. Because of the observation-driven nature of the dynamics of p_t^{01} the likelihood procedure is unaltered relative to the model with static transition probabilities. The complete prediction-update recursion and further details regarding estimation are provided in Appendix A.

3 Empirical application

3.1 Data

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 the monthly logarithmic growth rates of these four indicators from January 1959 until February 2020. Vintages for the real-time exercise are obtained from TCB and supplemented with data from Jeremy Piger[1]. In light of its specific and unusual features, the COVID-19 period from March 2020 until December 2020 is investigated separately in Appendix E.

Furthermore, we use the interest rate term spread (TS) as an exogenous driver of the time-varying peak probability. Specifically, we use the TS as included in TCB's Leading Economic Index (LEI), that is, the 10-year US Treasury rate minus the US Federal Funds rate. We collect the TS from the FRED database. Because of large differences in level over time, we transform the TS to a dummy variable. Specifically, we set our dummy equal to 1 if the TS is negative and 0 otherwise. This balances the contrast between elevated interest rates during the 1980s and the suppressed term premiums after 2008 due to quantitative easing. Empirically, we find that recessions are almost always preceded by periods of an inverted yield curve[2], i.e. a negative TS. The TS dummy therefore present a simple and intuitive predictor for future peaks.

Results of several robustness checks, whereby the TS is incorporated in a different manner

[1] Obtained from https://pages.uoregon.edu/jpiger/research/published-papers/raw-real-time-data.zip, see Appendix B for details.

[2] A graphical illustration is provided in Appendix B.4

can be found in Appendix C.4. These figures additionally contain findings in case the LEI is used as the exogenous variable, which also appears to be an effective choice. We prefer the TS over the LEI because the former tends to signal peaks earlier and does not contain revisions. Considering more or a weighted combination of LIs to drive the peak probability is left for future research.

3.2 Model specification details

We use the two-regime single-factor DFMS model with AR(1) specifications for the common factor and the idiosyncratic components as described in Section 2. Although more involved multi-state models can provide a better in-sample fit (e.g. Sichel, 1994), a two-regime setup appears more appropriate for our goal of peak dating. The magnitudes of the eigenvalues of the correlation matrix of the four coincident indicators in turn motivate the choice of a single common factor. AR(1) dynamics are selected mainly to not overcomplicate the already reasonably involved estimation. Our setup is therefore similar to the one of Chauvet (1998) for monthly data. Further extensions, including structural breaks to accommodate the Great Moderation (McConnell and Perez-Quiros, 2000) and time-varying mean growth rates (Eo and Kim, 2016; Doz et al., 2020; Eo and Morley, 2022) are left for future research.

The addition of a time-varying peak probability p_t^{01} substantially enhances the already high flexibility of the DFMS model. In a Bayesian context, Filardo and Gordon (1998) impose strong priors on the TVTPs in their MS model for output to prevent the underestimation of recession durations. This underestimation may lead to 'recession rallies', by which we mean that during a recession period specific months with only moderately negative (or even positive) growth rates may be classified in the expansion regime. We propose to similarly impose structure by calibrating the static transition probabilities starting from a recession regime (p^{11} and $p^{10} = 1 - p^{11}$). For this, we use a simple count-based estimator using completed NBER recessions. That is, we obtain p^{11} by dividing the total number of recession months minus the number of recessions by the total number of recession months. This approach is readily applicable in real-time[3] and effectively avoids the issue of 'recession rallies' discussed above. For completeness, results without targeting are provided in Appendix C.3.

[3]For the real-time analysis, we update p^{11} the month after an NBER trough announcement, see Appendix B.5 for its evolution. The ex-post analysis uses all recessions in the sample.

For both the ex-post and the real-time analysis, we consider four (nested) DFMS model specifications. The first is the base model as given in (1) and (3) with a constant peak probability p^{01}. The remaining three consider a time-varying peak probability p_t^{01} using only endogenous information (GAS), only exogenous drivers (Exo), or both (GASX). Specifically, we use (5)-(9) with $c = 0$ for GAS and with $a = 0$ for Exo, where x_t is the TS dummy. To identify the common factor, we fix the factor loading of employment $\lambda_{EMP} = 1$.

3.3 Full-sample results

We estimate the DFMS model and our extensions for the period January 1959 until February 2020 using the data vintage as released in March 2020. Hence, this includes all revisions in the coincident variables known at the final date. In Table 1, for brevity, only key parameter estimates for the four considered model specifications are shown. The remaining parameter estimates can be found in Appendix C.1. Furthermore, Table 1 provides an overview of the signaling performance of the filtered state probabilities $\Pr(S_t = 1|\mathcal{I}_t)$ for the NBER recessions. This includes the Area-Under-the-Receiver-Operating-Curve (AUROC), a common measure for evaluating binary classification ability (Berge and Jordà, 2011). Perfect and uninformative classifiers have AUROC values of 1 and 0.5, respectively. In addition, we consider the average contraction state probability during NBER recessions, non-recession periods and the first month of the recessions, denoted by π^r, π^e and π^p, respectively.

In Table 1, we observe from the log likelihood and the Akaike information criterion (AIC) that the extended specifications improve upon the base model to different degrees. This includes notable improvements from using the TS as exogenous input and somewhat more modest improvements due to the endogenous GAS dynamics. GASX is the preferred model according to the AIC. In terms of parameter estimates, we find that the parameter of the negative TS indicator c is positive and significant for both the Exo and GASX specifications. The positive sign is in accordance with economic intuition, 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 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

Table 1: Key parameter estimates and NBER recession signaling performance of the DFMS model and extensions.

	Base	**GAS**	**Exo**	**GASX**
w		-0.559	-0.463	-0.553
		(0.218)	(0.167)	(0.183)
a		1.207		0.979
		(0.399)		(0.461)
b		0.839	0.936	0.910
		(0.050)	(0.023)	(0.025)
c			0.533	0.685
			(0.171)	(0.221)
p^{01}	0.017			
	(0.006)			
LogL	-1913.3	-1907.3	-1900.2	-1897.9
k	16	18	18	19
AIC	3858.5	3850.7	3836.5	3833.9
AUROC	0.941	0.950	0.979***	0.978***
π^r	0.647	0.685	0.779	0.802
π^e	0.066	0.063	0.063	0.065
π^r/π^e	9.731	10.832	12.337	12.419
π^p	0.267	0.290	0.528	0.594

Note: This table presents the key parameter estimates for the base model and GAS, Exo and GASX specifications 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 parameters. Signaling power of the filtered states for the NBER recessions is evaluated using the AUROC, $\pi^{r(e)}$ represent the average filtered contraction state probability during NBER recession (expansion) periods and π^p the average probability in the first recession month. Finally, a significant difference of the AUROC in comparison with the base model is indicated with a *, ** and *** for a p-value below 0.10, 0.05 and 0.01.

Exo specification reject the null hypothesis at the 5 percent level ($LR = 11.87$, p-value $=$ 0.003 and $LR = 4.6179$, p-value $= 0.032$ respectively). We do note that in the absence of a formal proof of the asymptotic distribution of the LR test, some care must be taken with the interpretation of these p-values. However, considering the AIC or the individual significance of the relevant parameters corroborates these findings. That is, we find the GAS parameter a to be significant in both GAS and GASX specifications with lower AICs compared to the base and Exo models.

To formally compare the AUROC between the different models, we follow Aastveit et al. (2019) and use a (two-sided) Wald-type test as suggested by DeLong et al. (1988). In

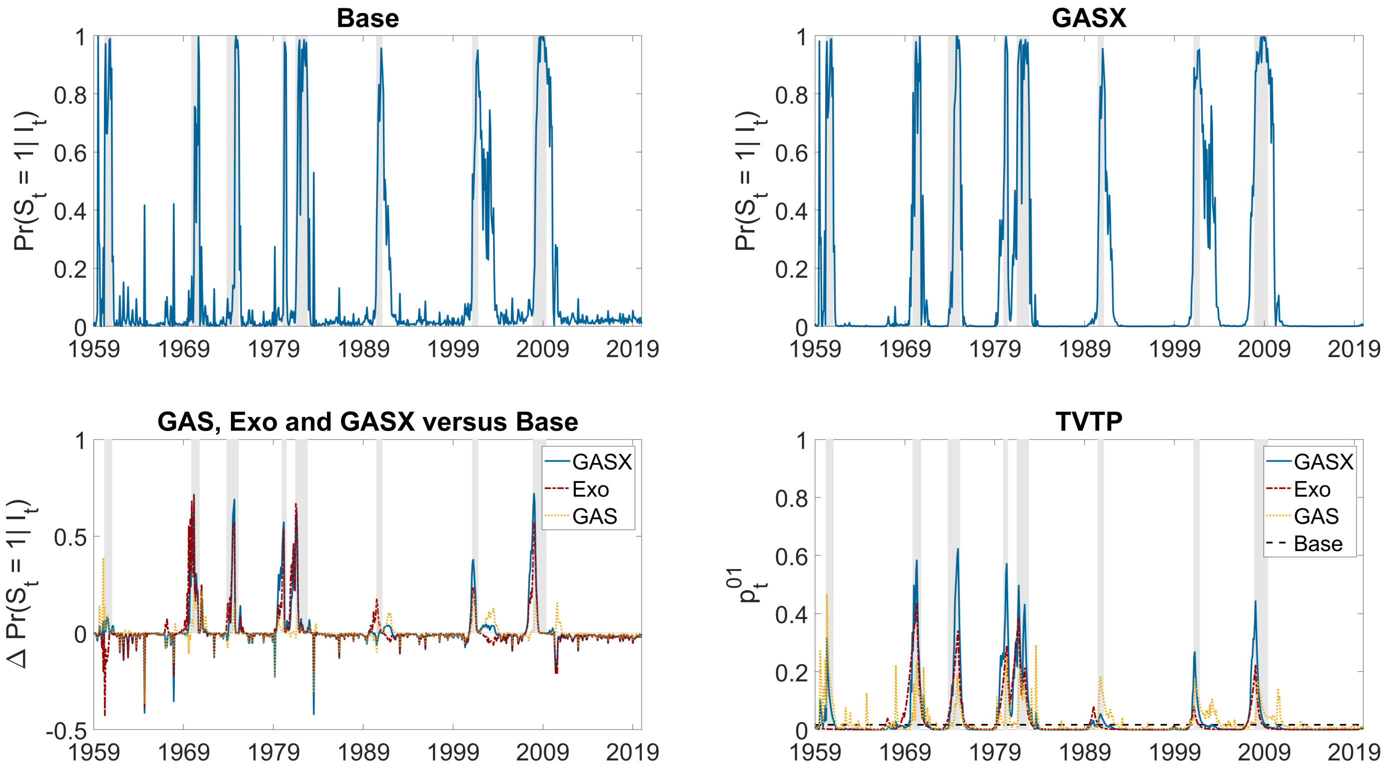
Table 1, we observe that all three extensions improve upon the signaling ability relative to the base model. The largest and also significant improvements stem from the addition of the TS information. For example, this leads to an increase of the overall AUROC from 0.941 to 0.979 for the Exo specification. While the overall AUROC does not improve when adding GAS dynamics, they do benefit peak dating by further increasing π^p. In particular, π^p increases from 0.528 to 0.594 in favor of the GASX model relative to the Exo specification. Similarly, the GASX model achieves the highest value of the ratio between π^r and π^e, which can be seen seen as an informal measure of signal clarity.

Figure 1 displays the filtered state probabilities $\Pr(S_t = 1|\mathcal{I}_t)$ of the base model and the GASX specification. To disentangle the performance gains of the GASX model, we additionally present the differences of the filtered probabilities of all three extended models (i.e. GAS, Exo and GASX) with the base model. Positive differences therefore indicate higher recession probabilities in the more extensive model specifications. Finally, Figure 1 also contains the estimated paths of the peak probability p_t^{01}. Naturally, a model specification is considered superior if it has a heightened p_t^{01} just before or at the start of a recession.

We observe from Figure 1 that both the base and the GASX model are highly successful in the identification of the business cycle regimes. The filtered state probabilities generally remain close to 0 during expansion periods and rapidly increase to levels close to 1 during recessions. Between the two, we observe a clear advantage of the GASX model in terms of recession signaling ability. In particular, the extended models dramatically increase the contraction probabilities at the start of recessions relative to the base model. In addition, the extended models reduce false signals, such as those in 1964, 1968 and 1983. Comparing the three extended specifications, we observe the largest benefits from the inclusion of the TS and a secondary but nevertheless meaningful role for GAS dynamics. In line with the results of Table 1, the combined GASX model seems to offer the best signaling performance.

From the evolution of the dynamic peak probability p_t^{01}, we can more clearly see the differences between the extended models. Specifically, the variation in p_t^{01} for the GAS model 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 GAS specification still only make use of coincident information. For the two other specifications, which include the TS, we observe large movements already before the start of

Figure 1: Filtered state probabilities and peak probabilities for the DFMS model and extensions.



Note: The top left (right) plot depicts the filtered state probabilities $\Pr(S_t = 1|\mathcal{I}_t)$ for the base (GASX) model. The bottom left plot contains the filtered state probabilities of the extensions minus those of the base model, denoted by $\Delta\Pr(S_t = 1|\mathcal{I}_t)$. The bottom right plot contains the estimated path of the peak probability p_t^{01}. Finally, the shaded areas reflect the recession periods as determined by the NBER.

recessions. This timeliness obviously reflects the nature of the TS as a leading indicator. For example, if we currently have that $p_t^{01} = 0.01$, then for the Exo specification three periods of an inverted yield curve will nearly triple the peak probability to $p_{t+3}^{01} = 0.0274$. Adding GAS dynamics on top further strengthens the movement of p_t^{01} and also adds a small increase before the 1960 recession. For this recession no yield curve signal is contained in the data. We conclude that a time-varying peak probability p_t^{01}, driven by a combination of exogenous and endogenous sources, can help improve the (ex-post) dating of recessions.

3.4 Real-time analysis

We perform a real-time analysis in order to make a fair assessment of the applicability of the GASX model in practice. Specifically, we recursively estimate the model parameters for each month from December 1976 until March 2020 using the most recent data vintage available at the time. We consider an expanding window approach where the observations start in January 1959. Following Camacho et al. (2018) and in line with TCB's release schedule for their LEI, we estimate the model parameters in the third week of each month. Specifically, this entails that in month t, we have EMP and IP available up to and including month $t-1$. For MAN and INC their data publication scheme implies that we have observations only up to and including month $t-3$ and $t-2$, respectively. Parameter estimation in month t is thus based on all observations up to and including month $t-1$, whereby the final two observations for MAN and the final observation for INC are considered missing. Here we differ from Chauvet and Piger (2008), who restrict the sample at each point in time to the series for which the least amount of information is available.

Table 2 presents an overview of the recession signaling performance of the real-time filtered and predicted state probabilities. To assess the economic value of our predictions, we also include the Anxious Index Nowcast (AIN) of Scavette (2014). The AIN is based on the Survey of Professional Forecasters (SPF) and corresponds with the probability of a current contraction in real GDP. To obtain monthly values of the AIN, we set all months within a quarter equal to the value of the AIN in that quarter.

We observe in Table 2 that all extensions that allow for a time-varying peak probability trump the base model in all considered metrics. This includes higher contraction state probabilities during NBER recession periods (π^r) and lower such probabilities during NBER expansion phases (π^e). In addition, the AUROC of both the filtered and predicted contraction state probabilities is significantly improved for all but the plain GAS model for the filtered states. Furthermore, the AUROC of the GASX model is on par with that of the AIN (p-value $= 0.770$ for a test of equality). Interestingly, the gain of the AIN over the base model is smaller in a statistical sense, being only significant at the 10 percent level. This is because the variance of the AUROC of the AIN is relatively large. This is in line with the other metrics of Table 2, where the values of $\pi^r_{t-1|t-1}$ indicate that the AIN produces a

Table 2: Real-time signaling performance of the DFMS model and extensions.

Filtered	**Base**	**GAS**	**Exo**	**GASX**	**AIN**
AUROC	0.940	0.944	0.967**	0.970**	0.973*
$\pi^r_{t-1\|t-1}$	0.751	0.783	0.828	0.845	0.609
$\pi^e_{t-1\|t-1}$	0.123	0.119	0.106	0.110	0.121
$\pi^r_{t-1\|t-1}/\pi^e_{t-1\|t-1}$	6.083	6.601	7.846	7.697	5.034
π^p	0.185	0.215	0.387	0.422	0.409

Predicted	**Base**	**GAS**	**Exo**	**GASX**
AUROC	0.894	0.908**	0.948***	0.955***
$\pi^r_{t\|t-1}$	0.620	0.669	0.710	0.740
$\pi^e_{t\|t-1}$	0.140	0.133	0.111	0.116
$\pi^r_{t\|t-1}/\pi^e_{t\|t-1}$	4.440	5.014	6.423	6.389
π^p	0.116	0.136	0.241	0.308

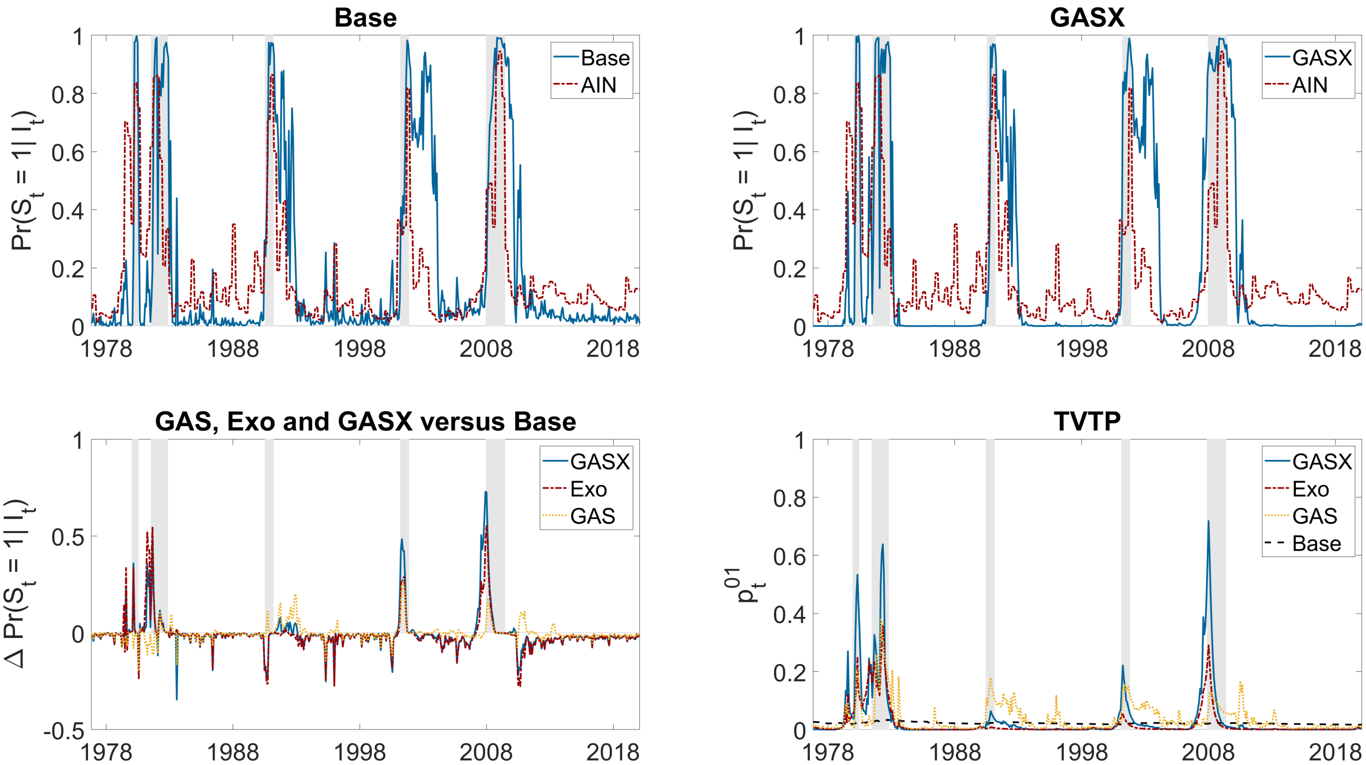
Note: Signaling power of the probabilities for the NBER recessions is evaluated using the AUROC. Furthermore, we have that $\pi^{r(e)}_{i|j}$ represent the average contraction state probability at time i during NBER recession (expansion) periods with observations up to and including time j, which corresponds to the real-time estimation at time $j+1$. In addition, π^p denotes the average state probability during the first month of the recessions within the evaluation sample. The evaluation sample ranges from November 1976 until February 2020. Finally, a significant difference of the AUROC in comparison with the base model is indicated with a *, ** and *** for a p-value below 0.10, 0.05 and 0.01.

weaker signal during recessions relative to the GASX model.

Comparing the three extended DFMS models, we find again that the addition of the exogenous information yields the largest improvements. Interestingly, the GASX model improves significantly upon the Exo model in terms of AUROC for the predicted state probabilities (p-value = 0.032). Endogenous dynamics are thus significant irregardless of the inclusion of the spread. For the filtered state probabilities this difference does not yet reach statistical significance (p-value = 0.150). Even though the absolute differences between the Exo and GASX models are not that large, the estimated standard errors of these differences are very small. This in turn leads to large test statistics. In the interest of dating peaks, we find that the average contraction state probability for the first 'official' (NBER) recession month (π^p) is the highest for the GASX model at 0.422 (0.308) for the filtered (predicted) state probabilities. This is more than double the value of the base model at 0.185 (0.116).

Figure 2 depicts the real-time evolution of the filtered state probabilities for the base model and the GASX specification. This can be understood to reflect the diagonal of the full history of state probabilities found in Appendix D.1. For comparison purposes, we also display the AIN. Additionally, Figure 2 displays the evolution of the peak probability p_t^{01}.

Figure 2: Real-time results for the base DFMS model and the GASX extension.

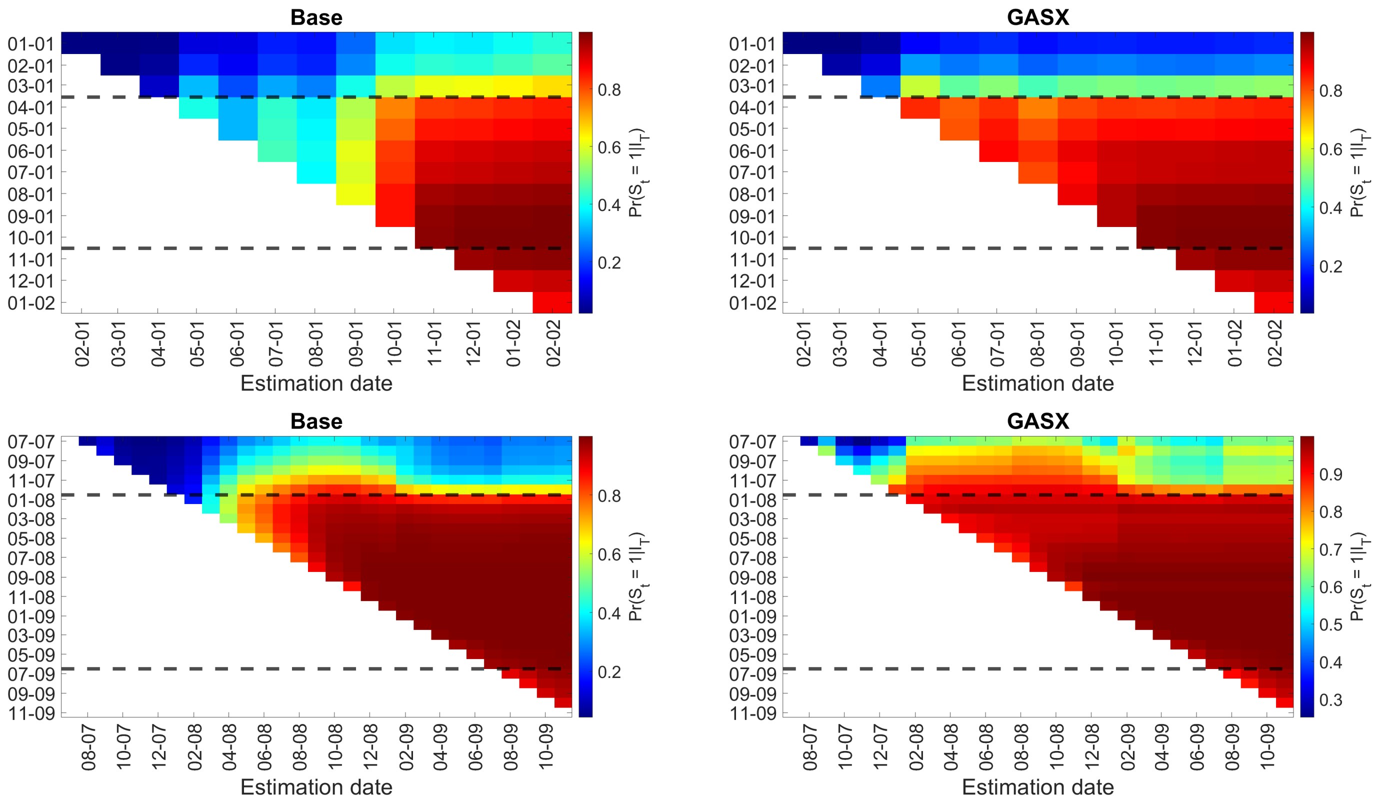


Note: The top left (right) figure displays the filtered state probabilities for the base (GASX) model. The bottom left plot displays the difference in filter state probabilities with the base model, while the bottom right figure contains the evolution of the peak probability p_t^{01}. Finally, the shaded areas reflect the recession periods as determined by the NBER.

We observe in Figure 2 that also in real-time the GASX specification matches the NBER recession periods much better than the base model. Compared to the AIN, we find the GASX model to provide the best signal most of the time, while the AIN appears to better signal troughs. Particularly, the trough of the 2001 recession is identified late by the DFMS models as also found by Chauvet and Piger (2008). This may be due to what is known as a jobless recovery, see Groshen and Potter (2003). While the average signaling power of the AIN measured by the AUROC is similar to the GASX model, we therefore still prefer the latter for peak dating. Compared to the base model, the GASX specification is better able to identify business cycle peaks for four out of the five recessions during the period 1976-2019. Only for the 1990 recession the GASX model initially appears to slightly dismiss the signal relative to the base model. Interestingly, Stock and Watson (1993) found similar results using their leading indicator at the time when incorporating the TS, see Hamilton (2011) for a discussion. Most strikingly, the 2001 and 2008 recessions are identified much earlier by the GASX model, compared to both the base model and the AIN.

Having observed the largest benefits for the GASX specification in the two most recent recessions, we further investigate the evolution of the state probabilities around these periods. Figure 3 depicts the evolution of the real-time smoothed contraction state probabilities for the recessions that followed the peaks in March 2001 and December 2007.

Figure 3: History of real-time smoothed state probabilities for the base DFMS model and the GASX extension around the 2001 and 2008 recession.



Note: This figure depicts the real-time smoothed state probabilities around the 2001 (top) and 2008 (bottom) recessions for the base (left) and GASX 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.

In Figure 3 we observe that the GASX 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 GASX specification is able to accurately provide a strong contraction signal as soon as these recessions start. In contrast, the base model requires several months of additional information to do so. As more time passes the models appear to largely agree on past states. Appendix D.2 contains a plot similar to Figure 3 for the filtered state probabilities. Here improvements are even more pronounced. All in all these findings indicate that the GASX 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.

To identify peak dates 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 $\Pr(S_t = 1|\mathcal{I}_T) < \tau$ and $\Pr(S_{t+k} = 1|\mathcal{I}_T) \geq \tau$, for $k = 1, 2, 3$ for some threshold $0.5 < \tau < 1$, where T is the last month of the vintage. Second, the peak associated with this recession period is identified by the point in time where immediately preceding this period the probability crosses 0.5. That is, we find the smallest non-negative integer q such that $\Pr(S_{t-q} = 1|\mathcal{I}_T) < 0.5$ and $\Pr(S_{t-q+1} = 1|\mathcal{I}_T) \geq 0.5$. The peak date is subsequently identified to be month $t - q$ and as such refers to the final expansion month. Before another peak can be 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 month t, after a peak has been dated and for which $\Pr(S_t = 1|\mathcal{I}_T) \geq \tau$ and $\Pr(S_{t+k} = 1|\mathcal{I}_T) < \tau$, for $k = 1, 2, 3$ as a trough. The trough therefore marks the final recession month. By construction, we now have that peaks and troughs alternate. By considering the recent state probabilities of the initial estimation date, December 1976, we determine that our evaluation window begins in an expansion. We therefore begin looking for a peak. In Table 3 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.

In Table 3, we observe that the DFMS models are able to match or precede the peak announcement dates of the NBER, often by a substantial margin. Specifically, the GASX specification is able to date the 2001 and 2008 peaks 4 and 11 (10) months earlier than the NBER for the threshold $\tau = 0.65$ (0.80). This is a gain of three and four (five) months over the base model, respectively. The large gain of the GASX model in timeliness relative

Table 3: Comparison peak dates from the DFMS specifications with the NBER recessions.

$\tau = 0.65$

Peak date			*Announcement date*		
Base	**GASX**	**NBER**	**Base**	**GASX**	**NBER**
0 (1)	1 (0)	Jan 1980	0	0	Jun 1980
0 (0)	0 (0)	Jul 1981	-1	-2	Jan 1982
-1 (-2)	-1 (-2)	Jul 1990	-4	-4	Apr 1991
-1 (-2)	-1 (-1)	Mar 2001	-1	-4	Nov 2001
-2 (0)	-4 (-1)	Dec 2007	-7	-11	Dec 2008

$\tau = 0.8$

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

Note: This table contains the monthly differences in obtained initial peak dates of the base model and GASX extension with the NBER database. The numbers 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.

to the NBER therefore appear about half due to the GASX structure and half due to the DFMS model itself. For the first three recessions we find comparable performance of the base model and the extension, but note that both are generally able to match or precede the NBER announcements. Of course, the NBER has a different aim with their dating procedure, valuing accuracy above speed. In terms of the peak dates, we find that the DFMS model produces similar dates as the NBER for the first four recessions. The peak date of the 2008 recession, however, is initially identified by the GASX model 4 (6) months before the corresponding NBER peak date for the threshold $\tau = 0.65$ (0.80). This is consistent with Figure 3, where we observe that the smoothed contraction state probabilities are elevated above 0.5 several months before the NBER peak date.

Because the dating procedure is done at each point in time (i.e. for each 'column'), it might be that later estimation dates produce different turning points than initially established. The values in parentheses in Table 3 reflect the peak dates as determined 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. This suggests a trade-off between timeliness and accuracy when comparing our approach with that of the NBER.

With regards to the troughs corresponding to the peaks of Table 3, we confirm the results of Chauvet and Piger (2008). That is, the DFMS model is generally able to call troughs much earlier than the NBER, with the exception of the 2001 recession associated with a jobless recovery. Doz et al. (2020) find that opting for a different measure of employment (specifically, civilian unemployment) may remedy this specific dating delay. Differences between the base DFMS model and the GASX specification are small. This finding is not surprising as the transition probability p^{11} plays a much larger role here than p_t^{01} and is set to the same value for both specifications. For this reason and for brevity, the troughs corresponding to the peaks of Table 3 can be found in Appendix D.3.

We conclude that qualitatively our findings in real-time closely match those of our ex-post analysis. This entails that the GASX model provides clear benefits over the base DFMS model in terms of recession signaling ability. In particular, the GASX model is able to identify peaks faster in real-time than the NBER.

4 Concluding remarks

We analyze the possibilities to accelerate the dating of business cycle peaks by extending the DFMS model of Diebold and Rudebusch (1996) to allow for TVTPs. Specifically, we propose to use a score-driven approach to guide the transition probabilities, building upon the techniques found in Bazzi et al. (2017). Additionally, we allow for relevant exogenous economic drivers, as suggested by Filardo (1994). Using the components of TCB's CEI from 1959 until 2020 and the term spread as an exogenous input, we allow the transition probability to switch from an expansion to a contraction phase to be time-varying. We find that the proposed method can significantly accelerate the real-time peak dating for US recessions. Most benefit is obtained from the use of the term spread, with a more limited but economically meaningful role for score-driven dynamics. We therefore recommend a combined approach, such as the GASX model, to predict whether there is a recession approaching.

References

Aastveit, K.A., A.K. Anundsen, and E.I. Herstad (2019). Residential investment and recession predictability. *International Journal of Forecasting* **35**, 1790–1799.

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.

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review* **109**, 1263–89.

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.

Blasques, F., C. Francq, and S. Laurent (2022). Quasi score-driven models. *Journal of Econometrics*, In Press, Available at: https://doi.org/10.1016/j.jeconom.2021.12.005.

Blasques, F., S.J. Koopman, and A. Lucas (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika* **102**, 325–343.

Boldin, M.D. (1994). Dating turning points in the business cycle. *Journal of Business*, 97–131.

Bry, G. and C. Boschan (1971). Standard business cycle analysis of economic time series. *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. NBER, 64–150.

Burns, A.F. and W.C. Mitchell (1946). The basic measures of cyclical behavior. *Measuring Business Cycles*. NBER, 115–202.

Caldara, D. et al. (2020). "Understanding Growth-at-Risk: A Markov-Switching Approach". *Philaldephia Fed Conference on Real-Time Data Analysis, Methods and Applications*.

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. Piger (2003). Identifying business cycle turning points in real time. *Federal Reserve Bank of St. Louis Review* **85**, 47–47.

— (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.

DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845.

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 (2022). Why has the US economy stagnated since the Great Recession? *Review of Economics and Statistics* **104**, 246–258.

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.

Filardo, A.J. and S.F. Gordon (1998). Business cycle durations. *Journal of Econometrics* **85**, 99–123.

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.

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.

Layton, A.P. (1996). Dating and predicting phase changes in the US business cycle. *International Journal of Forecasting* **12**, 417–428.

Layton, A.P. and M. Katsuura (2001). Comparison of regime switching, probit and logit models in dating and forecasting US business cycles. *International Journal of Forecasting* **17**, 403–417.

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

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.

Scavette, A. (2014). Are We in a Recession? The 'Anxious Index Nowcast'Knows! *Federal Reserve Bank of Philadelphia Special Report.*

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

Stock, J.H. and M.W. Watson (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual* **4**, 351–394.

— (1993). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. *Business cycles, Indicators and Forecasting* **28**, 95–156.

— (2005). Implications of Dynamic Factor Models for VAR Analysis. *NBER Working Paper*, Available at: https://www.nber.org/papers/w11467.

Stock, J.H. and M.W. Watson (2010). Indicators for dating business cycles: Cross-history selection and comparisons. *American Economic Review* **100**, 16–19.

— (2014). Estimating turning points using large data sets. *Journal of Econometrics* **178**, 368–381.

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.

Online Appendix to Accelerating Peak Dating in a Dynamic Factor Markov-Switching Model

A Prediction-update recursion

In this Appendix, we provide details of the prediction and update steps in the Hamilton filter and the Kalman filter, which are used to estimate the DFMS model as discussed in Section 2. The Hamilton prediction step is given as

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

and the Kalman prediction steps can be written as

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

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

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

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

where $\phi_t^{ij}(\mathbf{y}_t|\mathcal{I}_{t-1})$ 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

$$\boldsymbol{\mu}_t^{ij} = \mathbf{Z}\boldsymbol{\zeta}_{t|t-1}^{ij}, \tag{A.5}$$

$$\boldsymbol{\Sigma}_t^{ij} = \mathbf{Z}\mathbf{P}_{t|t-1}^{ij}\mathbf{Z}'. \tag{A.6}$$

Here $\boldsymbol{\zeta}_{t|t-1}^{ij}$ denotes the expectation of the state vector $\boldsymbol{\zeta}_t$ conditional on all information available at time $t-1$ 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 Kalman update steps are given by

$$\boldsymbol{\zeta}_{t|t}^{ij} = \boldsymbol{\zeta}_{t|t-1}^{ij} + \mathbf{P}_{t|t-1}^{ij}\mathbf{Z}'(\mathbf{Z}\mathbf{P}_{t|t-1}^{ij}\mathbf{Z}')^{-1}(\mathbf{y}_t - \mathbf{Z}\boldsymbol{\zeta}_{t|t-1}^{ij}), \tag{A.7}$$

$$\mathbf{P}_{t|t}^{ij} = (\mathbf{I} - \mathbf{P}_{t|t-1}^{ij}\mathbf{Z}'(\mathbf{Z}\mathbf{P}_{t|t-1}^{ij}\mathbf{Z}')^{-1}\mathbf{Z})\mathbf{P}_{t|t-1}^{ij}. \tag{A.8}$$

Due to the Markovian nature of the regime process in the DFMS model, the predictions and updates depend on the complete history of the state evolution. This feature substantially complicates the estimation procedure. To prevent an increasing length of path dependence, we collapse the states after each update step as proposed by Kim (1994). In this context, 'collapsing' means that we eliminate the dependence of our updated quantities on the regime of the previous period. For the state probability this is relatively straightforward and given as

$$\Pr(S_t = j|\mathcal{I}_t) = \sum_{i\in\{0,1\}} \Pr(S_t = j, S_{t-1} = i|\mathcal{I}_t). \tag{A.9}$$

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

$$\boldsymbol{\zeta}^j_{t|t} = \frac{\sum_{i\in\{0,1\}} \Pr(S_t = j, S_{t-1} = i|\mathcal{I}_t)\boldsymbol{\zeta}^{ij}_{t|t}}{\Pr(S_t = j|\mathcal{I}_t)}, \tag{A.10}$$

$$\mathbf{P}^j_{t|t} = \frac{\sum_{i\in\{0,1\}} \Pr(S_t = j, S_{t-1} = i|\mathcal{I}_t)(\mathbf{P}^{ij}_{t|t} + (\boldsymbol{\zeta}^j_{t|t} - \boldsymbol{\zeta}^{ij}_{t|t})(\boldsymbol{\zeta}^j_{t|t} - \boldsymbol{\zeta}^{ij}_{t|t})')}{\Pr(S_t = j|\mathcal{I}_t)}. \tag{A.11}$$

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

$$\mathrm{Log}L \approx \sum_{t=1}^{T} \mathrm{Log}[\sum_{i,j\in\{0,1\}} p^{ij}_t \Pr(S_{t-1} = i|\mathcal{I}_{t-1})\phi^{ij}_t(\mathbf{y}_t|\mathcal{I}_{t-1})]. \tag{A.12}$$

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 that is estimated 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. For the initialization of the factor, a diffuse prior with zero mean is considered at time $t = 0$.

The precise statistical properties of the filter, such as stationarity and invertibility, and the consistency of the MLE estimator depend on the true DGP. We may formalize this by investigating the stochastic recurrence equations of the stacked vector of all time-varying components of the filter and then define a parameter space that yields a well-behaved filter by imposing a boundedness and contraction condition. Practically, here, this boils down to a restriction on the (absolute) magnitudes of a and b. For a more elaborate discussion, see Blasques et al. (2022) and Bazzi et al. (2017). If stability is a concern, one may opt to restrict the range of the time-varying transition probability as pointed out by Bazzi et al. (2017). This can be done directly by altering the specification of the link function in (5).

Finally, we remark that we encountered some (near-)identification issues with the estimation of the purely score-driven GAS model in the real-time exercise for some early vintages. To remedy this, we search in a region for which the dynamics of p_t^{01} are sufficiently persistent ($b > 0.6$) and do not allow for erratic patterns with low persistence and excessive update steps. No boundary solutions are found.

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 and industrial production, while the remainder of the data reports values later in the month after the publication of last month's values. The data from TCB before January 2001 is also already available around the third week of the previous month, because no new information becomes available between the end of each month and the beginning of its subsequent month for these variables. To account for this difference in monthly reporting time, we therefore shift the vintages for this period one month, such that it is available one month earlier. For December 2001 two vintages are available, one in the first and one in the third week, such that this shift does not result in missing data for this month.

B.2 Manufacturing and trade sales

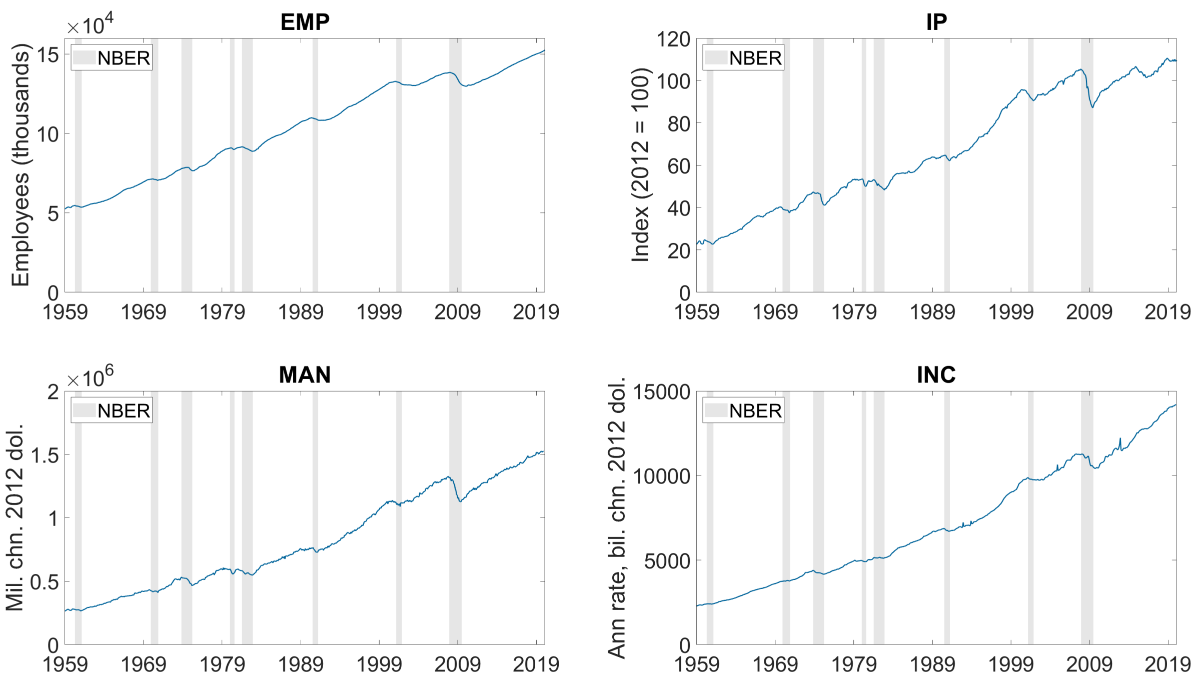
Similar to Chauvet and Piger (2008), data for the vintages from December 1976 up to and including December 1988 are obtained from *Business Conditions Digest*. For the remaining period of January 1989 through March 2020, data is retrieved from TCB. For both datasets the reporting moment in each month 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 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 the estimates found with the winsorized data. This approach therefore excludes outlier solutions and prevents structural breaks in the parameter estimates in real-time when an outlier enters the estimation window. For the models that allow for a time-varying peak probability p_t^{01}, we search for an optimum in a sensible neighbourhood of the base model, with respect to the common parameters.

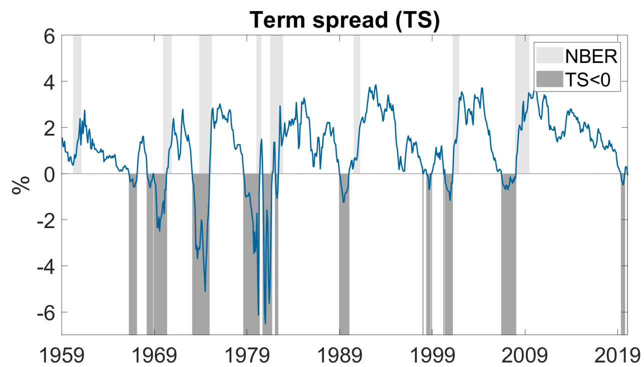
Figure B.1: Final vintage US employees on nonfarm payrolls (EMP), the index of industrial production (IP), manufacturing and trade sales (MAN) and personal income less transfer payments (INC), January 1959 - February 2020.



Note: The shaded areas reflect the recession periods as determined by the NBER.

B.4 Term spread

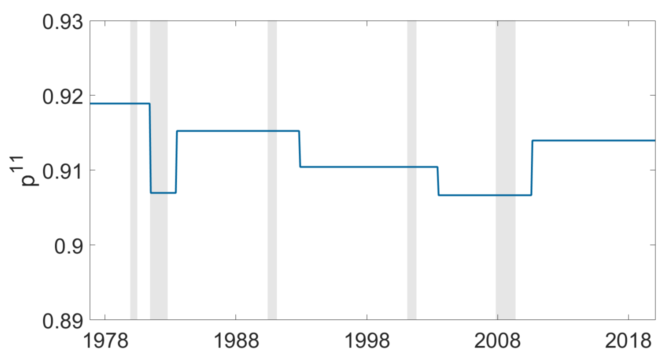
Figure B.2: US term spread, January 1959 - February 2020.



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

B.5 Real-time p^{11} target

Figure B.3: Evolution of the target p^{11} estimated using completed NBER recessions used for the real-time analysis.



Note: This figure depicts the evolution of the estimate of p^{11} when estimated using an expanding window 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.

C Full-sample

C.1 Leftover parameter estimates

Table C.1: Remaining parameter estimates for the DFMS model and extensions.

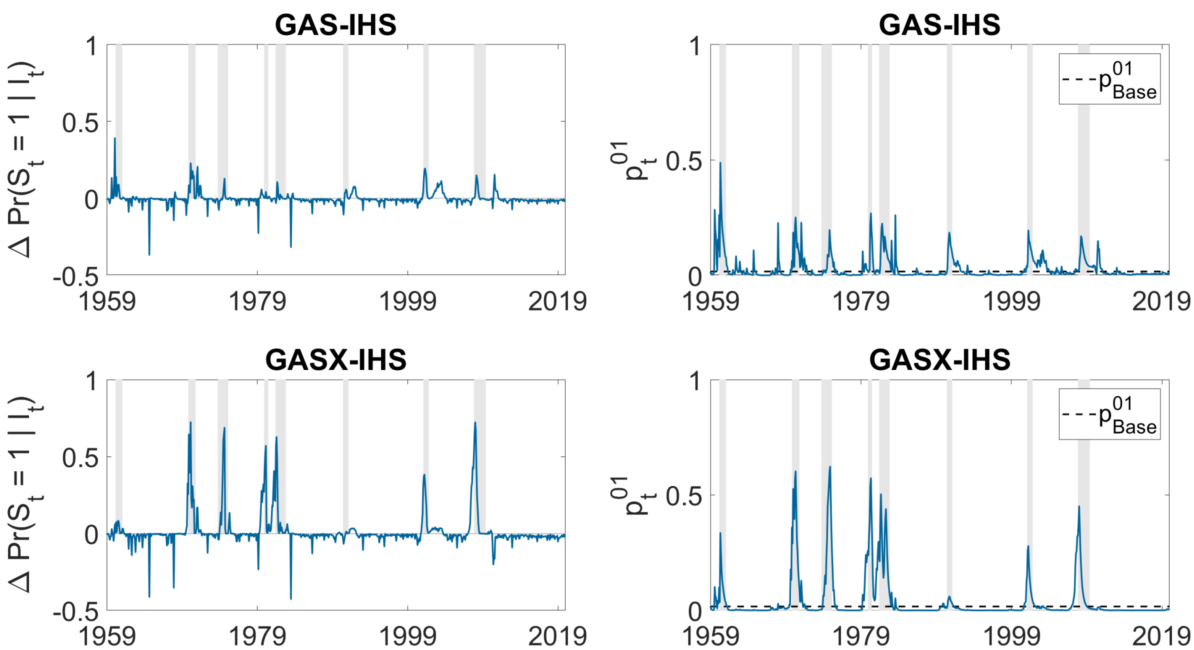
	Base	**GAS**	**Exo**	**GASX**
α_0	0.094	0.091	0.093	0.093
	(0.012)	(0.012)	(0.012)	(0.012)
α_1	-0.097	-0.095	-0.094	-0.089
	(0.023)	(0.020)	(0.023)	(0.022)
ϕ	0.546	0.556	0.552	0.552
	(0.050)	(0.049)	(0.051)	(0.048)
λ_{IP}	2.298	2.310	2.302	2.305
	(0.114)	(0.115)	(0.115)	(0.115)
λ_{MAN}	1.907	1.914	1.909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

Note: This table presents the leftover parameter estimates for the DFMS model and the extensions 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.

C.2 Robustness transformation

In this Appendix, we investigate the effects of a different transformation $g(\cdot)$ of the score used in (5)-(9). Specifically, Figure C.1 displays ex-post results using the inverse hyperbolic sine (IHS) transformation $g(x) = \log(x + \sqrt{x^2 + 1})$. We observe highly similar results to those in the ex-post analysis of Section 3.3.

Figure C.1: Ex-post results for the score-driven specifications using the inverse hyperbolic sine transformation.



Note: The left plots depict the differences in filtered contraction state probabilities with the base model (extension minus base) denoted by $\Delta\Pr(S_t = 1|\mathcal{I}_t)$. The right plots provide the estimated path of the conditional peak probability p_t^{01}. All models are estimated using the inverse hyperbolic sine (IHS) transformation $g(x) = \log(x + \sqrt{x^2 + 1})$. Finally, the shaded areas reflect the recession periods as determined by the NBER.

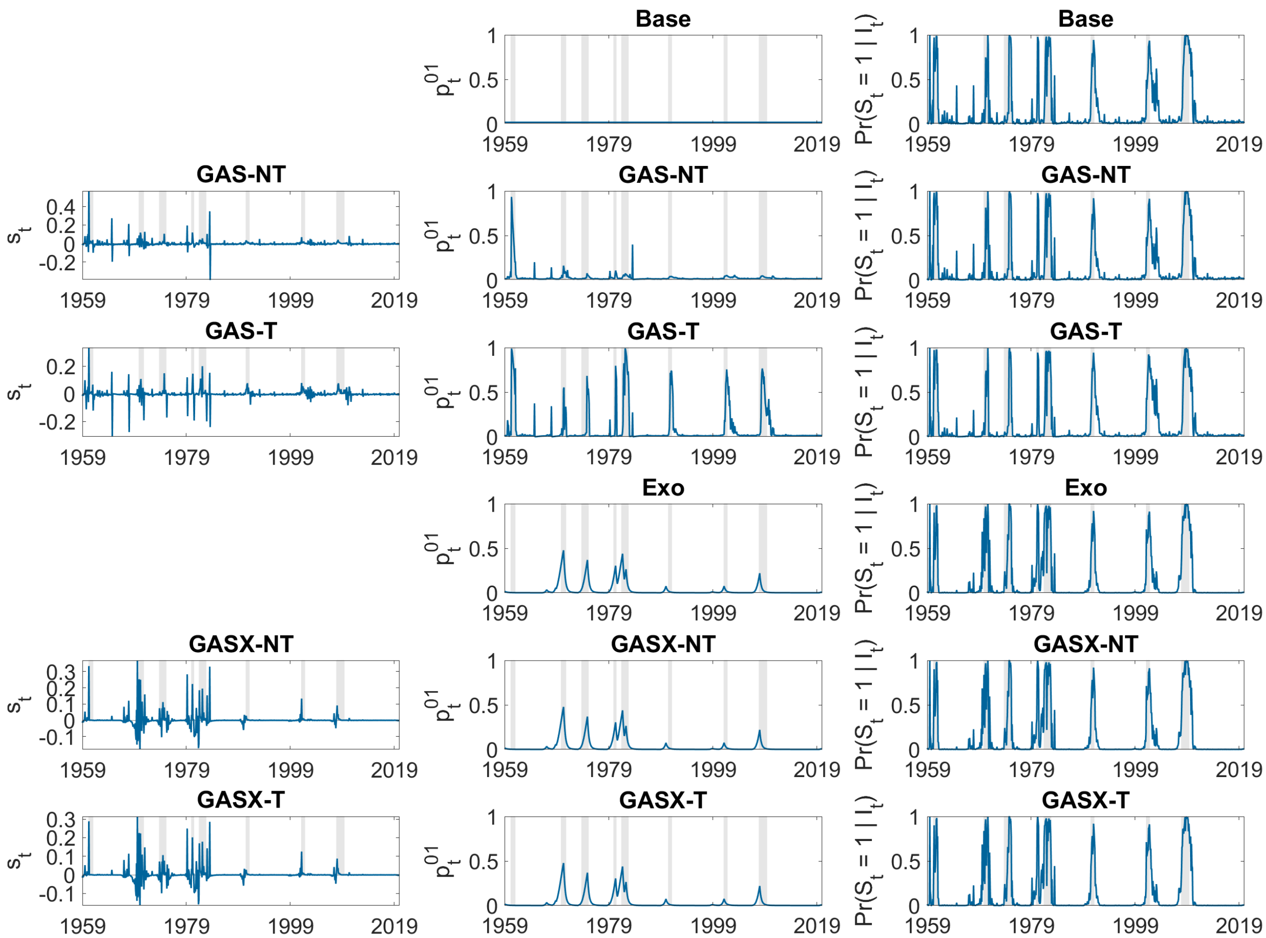
C.3 Robustness of scaling, transformation and targeting

This Appendix investigates the robustness of three model specification choices of the GASX model outlined in (5)-(9). First, this concerns the scaling of the score. In particular, for the analysis as presented in the paper we set $H_t = 1/(p_t^{01}(1 - p_t^{01}))$ in order to offset the effects of the link function on the score. Another common choice that may be considered is to omit scaling altogether and use $H_t = 1$. Second, we apply the transformation $g(x) = \text{sgn}(x)\log(1 + |x|)$ to reduce the impact of large scores. Third, we choose to target the recession-to-recession transition probability p^{11}, instead of estimating it alongside the other parameters. This regularization vastly increases stability in the presence of large time-variation in p_t^{01}.

In order to provide a complete picture of the effects of each these choices and possible interactions, we estimate and provide the results of every possible model specification without these interventions. Specifically, Figure C.2 and Figure C.3 contain results for the score-driven models using $H_t = 1$ and using no targeting or targeting for p^{11}, respectively. Similarly, Figure C.4 and Figure C.5 contain results for the score-driven models using $H_t = 1/(p_t^{01}(1 - p_t^{01}))$ and again using either no targeting or targeting for p^{11}. Within each of the figures we consider both with/without transformation (i.e. $g(x) = \text{sgn}(x)\log(1 + |x|)$ or $g(x) = x$) and with/without exogenous input in the form of the TS dummy. We estimate the models on the full-sample, such that the main ex-post results of Section 3.3 therefore overlap with Figure C.5.

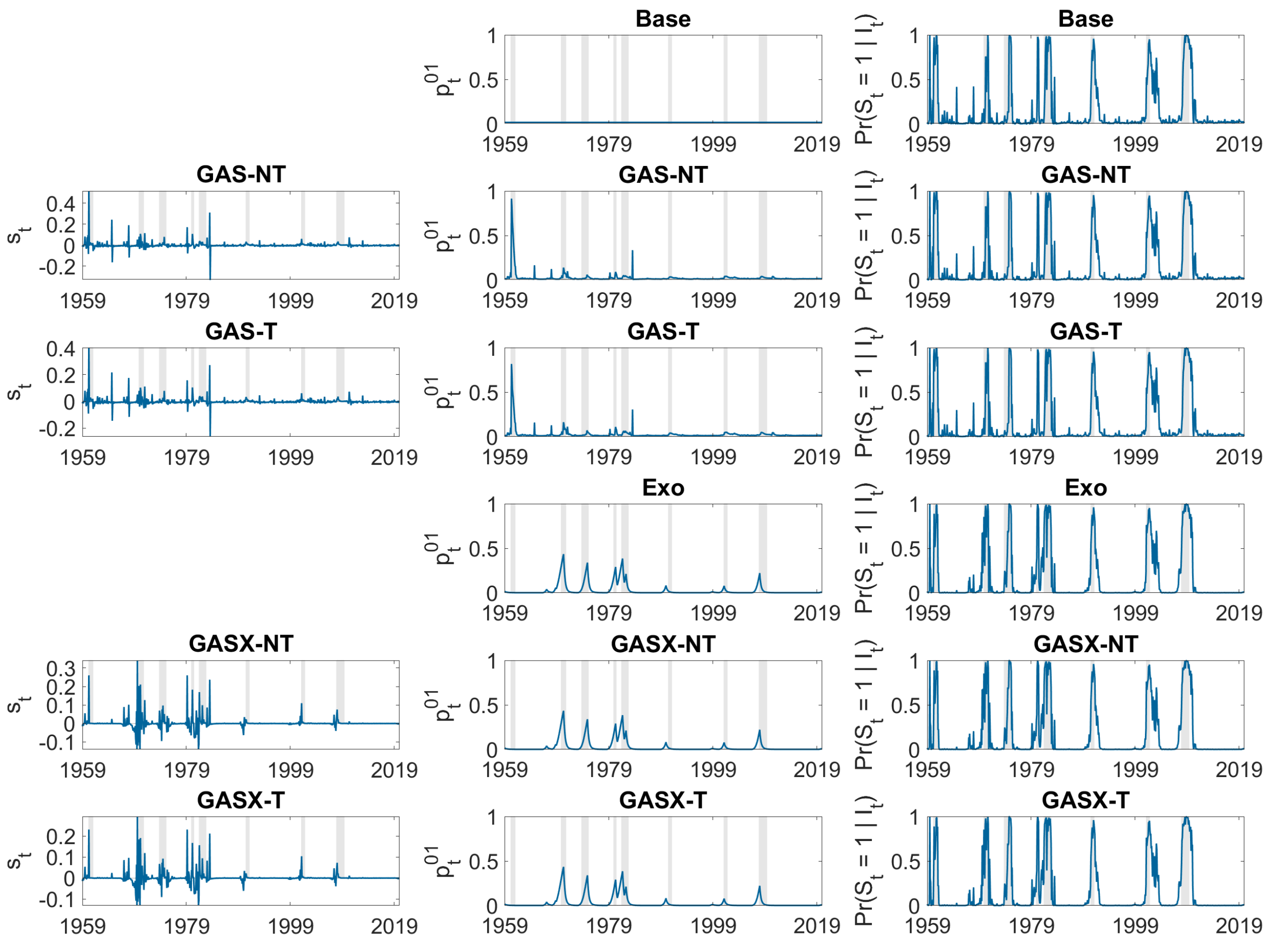
In Figures C.2-C.5, we observe that the addition of the scaling and the transformation generally increase the time-variation in p_t^{01}, while targeting p^{11} prevents 'recession rallies'. Hence, these specification choices achieve the aim for which they are intended. Comparing the models that include the TS against those we do not, we again confirm that the exogenous input provides substantial benefits.

Figure C.2: Ex-post results for different DFMS specifications without scaling and without targeting.



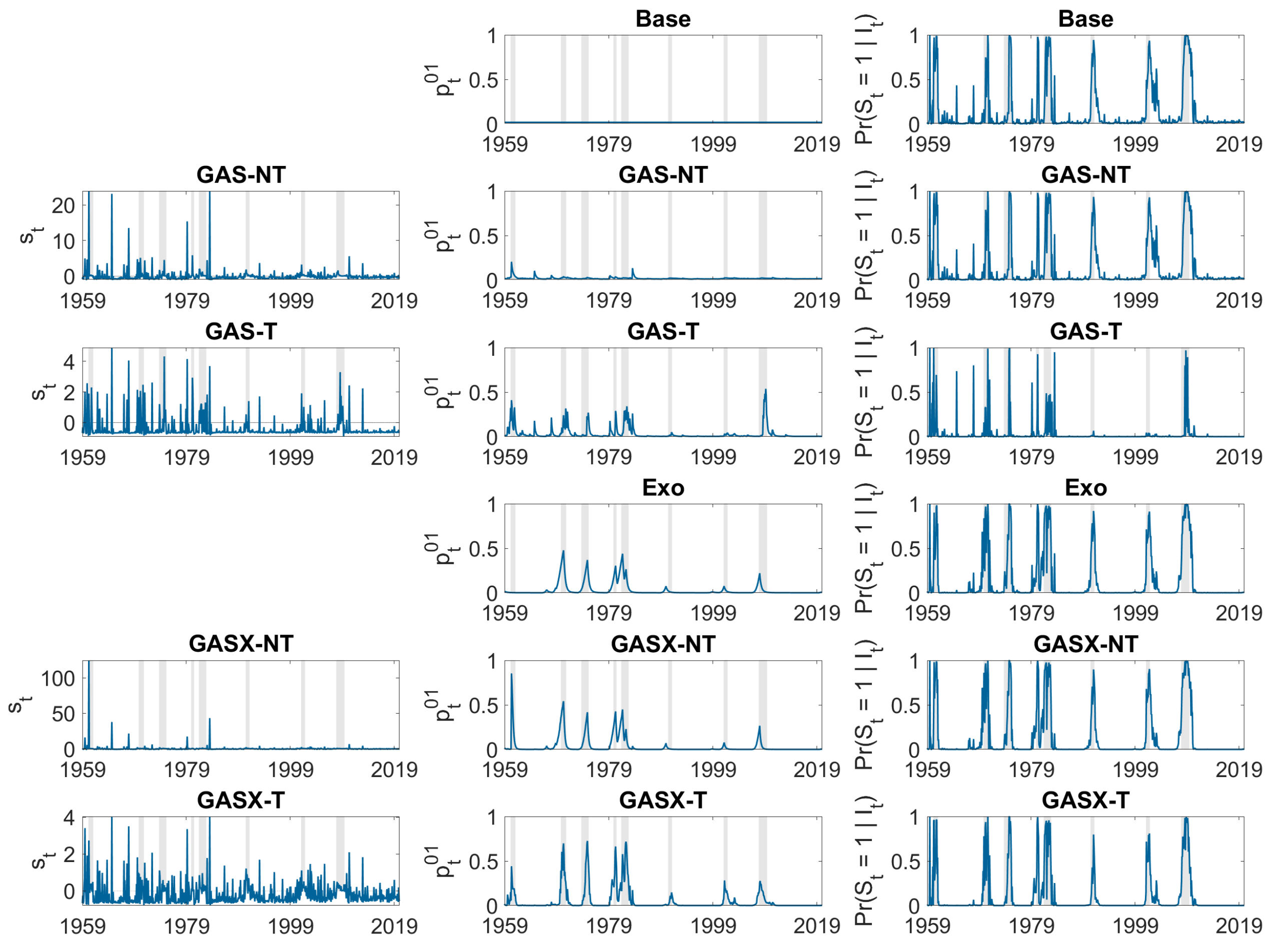
Note: This figure contains the evolution of the transformed scaled score s_t, the peak probability p_t^{01} and the updated state probability $\Pr(S_t = 1|\mathcal{I}_t)$ for the GAS models with (-T) and without transformation of the score (-NT). Finally, the shaded areas reflect the recession periods as determined by the NBER.

Figure C.3: Ex-post results for different DFMS specifications without scaling and with targeting.



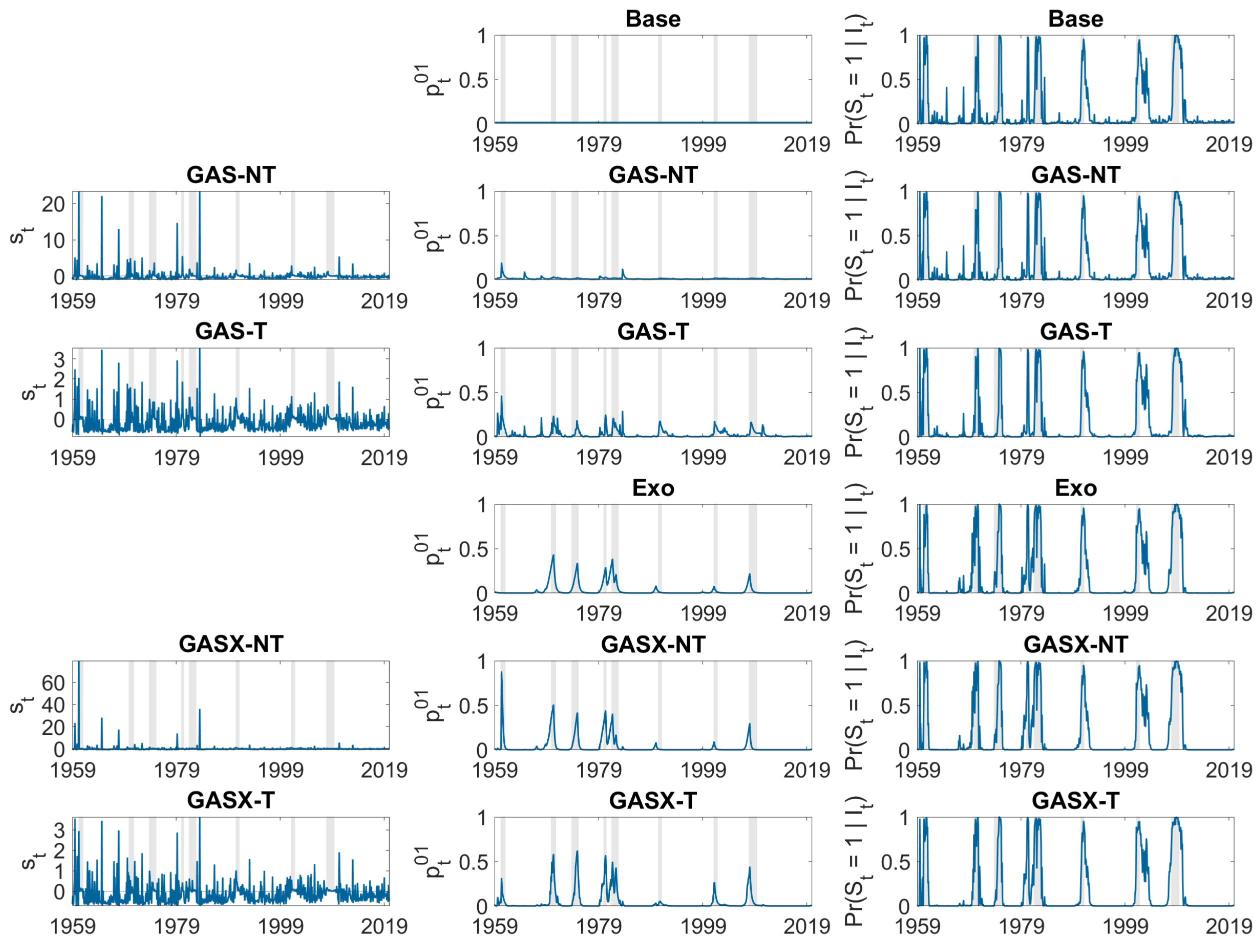
Note: This figure contains the evolution of the transformed scaled score s_t, the peak probability p_t^{01} and the updated state probability $\Pr(S_t = 1|\mathcal{I}_t)$ for the GAS models with (-T) and without transformation of the score (-NT). Finally, the shaded areas reflect the recession periods as determined by the NBER.

Figure C.4: Ex-post results for different DFMS specifications with scaling and without targeting.



Note: This figure contains the evolution of the transformed scaled score s_t, the peak probability p_t^{01} and the updated state probability $\Pr(S_t = 1|\mathcal{I}_t)$ for the GAS models with (-T) and without transformation of the score (-NT). Finally, the shaded areas reflect the recession periods as determined by the NBER.

Figure C.5: Ex-post results for different DFMS specifications with scaling and with targeting.

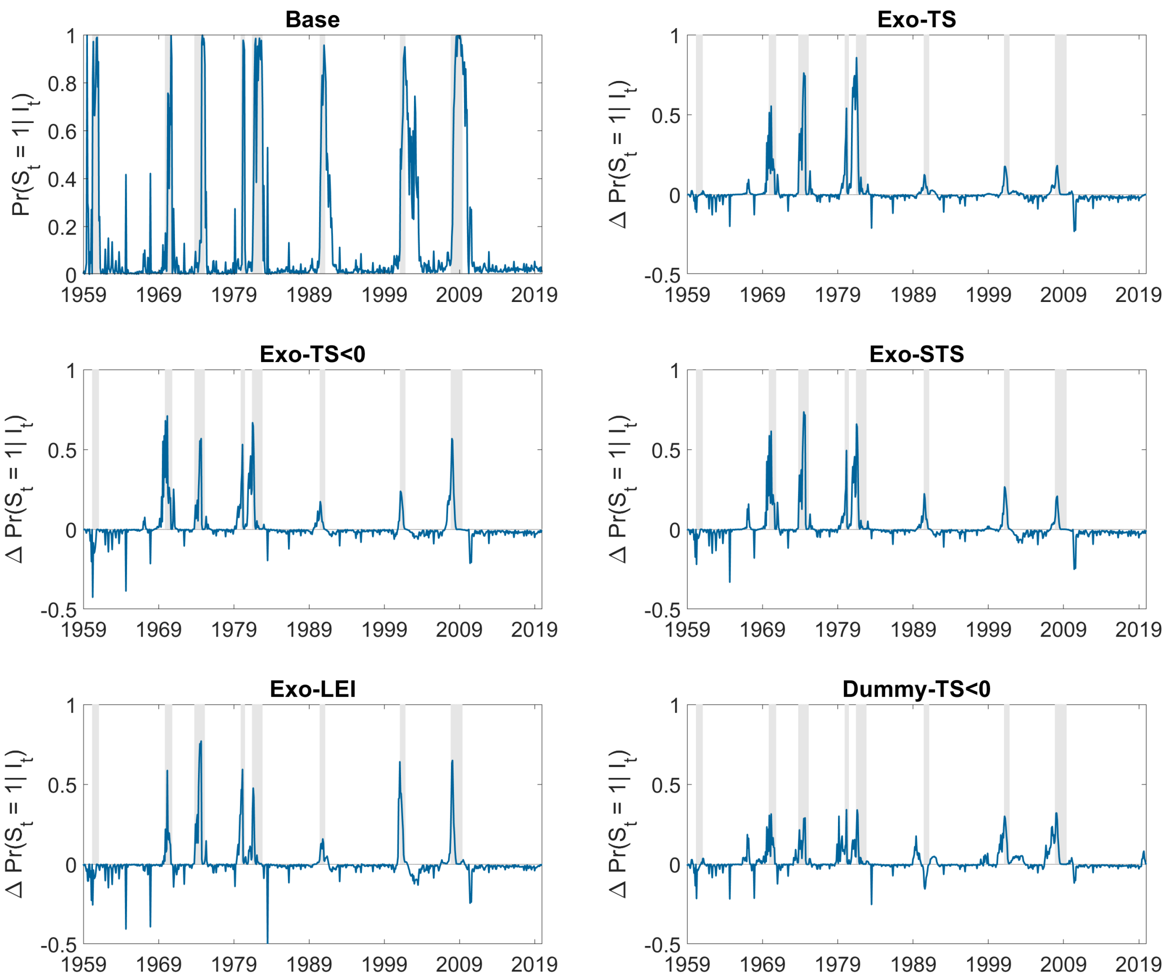


Note: This figure contains the evolution of the transformed scaled score s_t, the peak probability p_t^{01} and the updated state probability $\Pr(S_t = 1|\mathcal{I}_t)$ for the GAS models with (-T) and without transformation of the score (-NT). Finally, the shaded areas reflect the recession periods as determined by the NBER.

C.4 Robustness exogenous variable choice

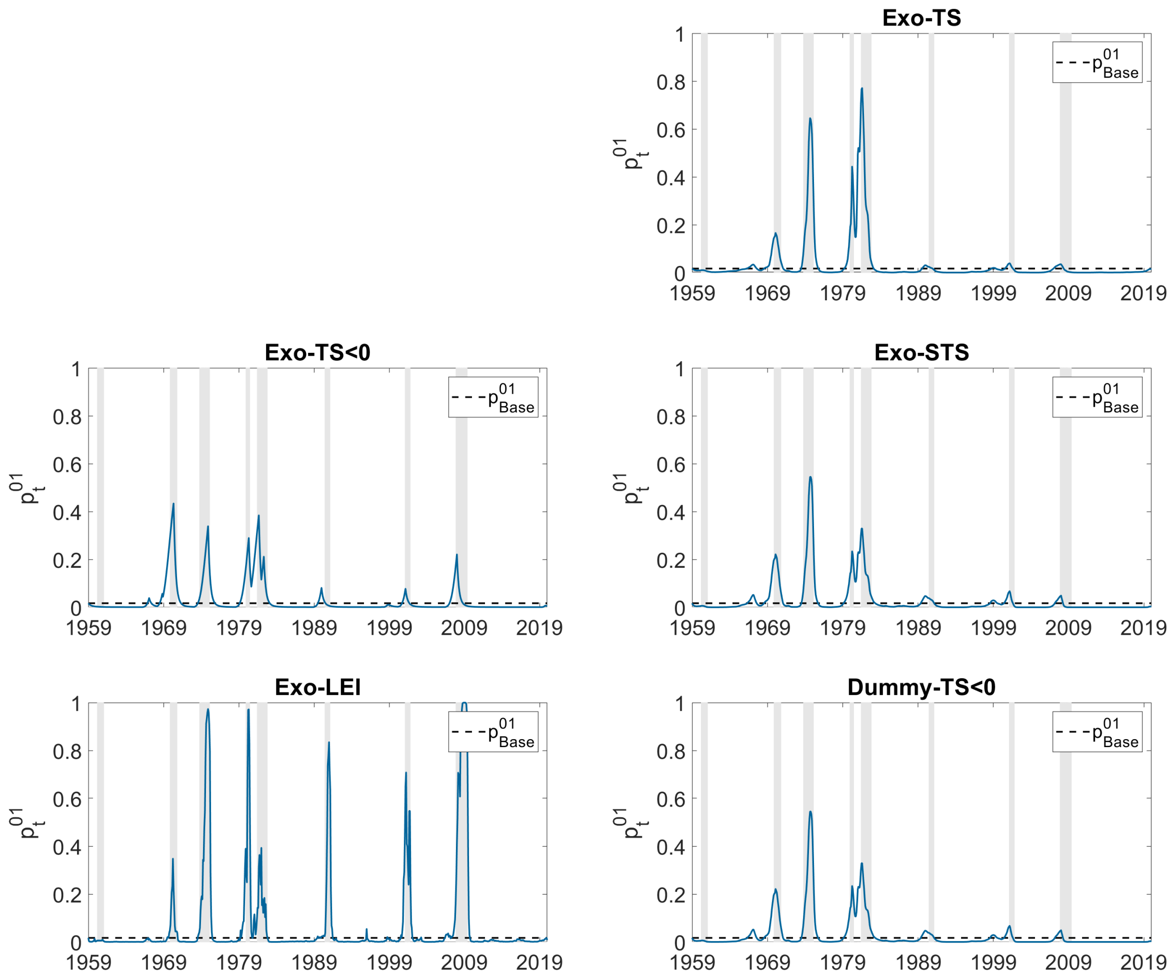
In the analysis of the paper we use a dummy for a negative term spread (TS) as an exogenous driver of the time-varying peak probability p_t^{01}. Figure C.6 and Figure C.7 contain results of the ex-post analysis where the TS is used differently. They also contains results when using TCB's Leading Economic Index (LEI). In particular, we observe that using the TS without transformation produces large movement only around the period of elevated interest rates in the 1980s. A transformation, such as the dummy for a negative sign, therefore appears necessary to balance this exogenous input. Using the LEI also appears an effective choice but appears to signal slightly later than the TS.

Figure C.6: Robustness exogenous variable choice, filtered state probabilities.



Note: Filtered contraction state probabilities ($\Pr(S_t = 1|\mathcal{I}_t)$) for the base model are depicted in the top left panel. The remaining figures depict the differences with the base model, denoted by $\Delta\Pr(S_t = 1|\mathcal{I}_t)$. Exo-($\cdot$) denotes the autoregressive specification that only makes use of ($\cdot$) to drive p_t^{01}, whereby STS denotes the standardized TS 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-TS<0 denotes a specification without autoregressive dynamics that simply considers two different levels of p^{01} depending on the sign of the TS. Finally, the shaded areas reflect the recession periods as determined by the NBER.

Figure C.7: Robustness exogenous variable choice, transition probability p_t^{01}.

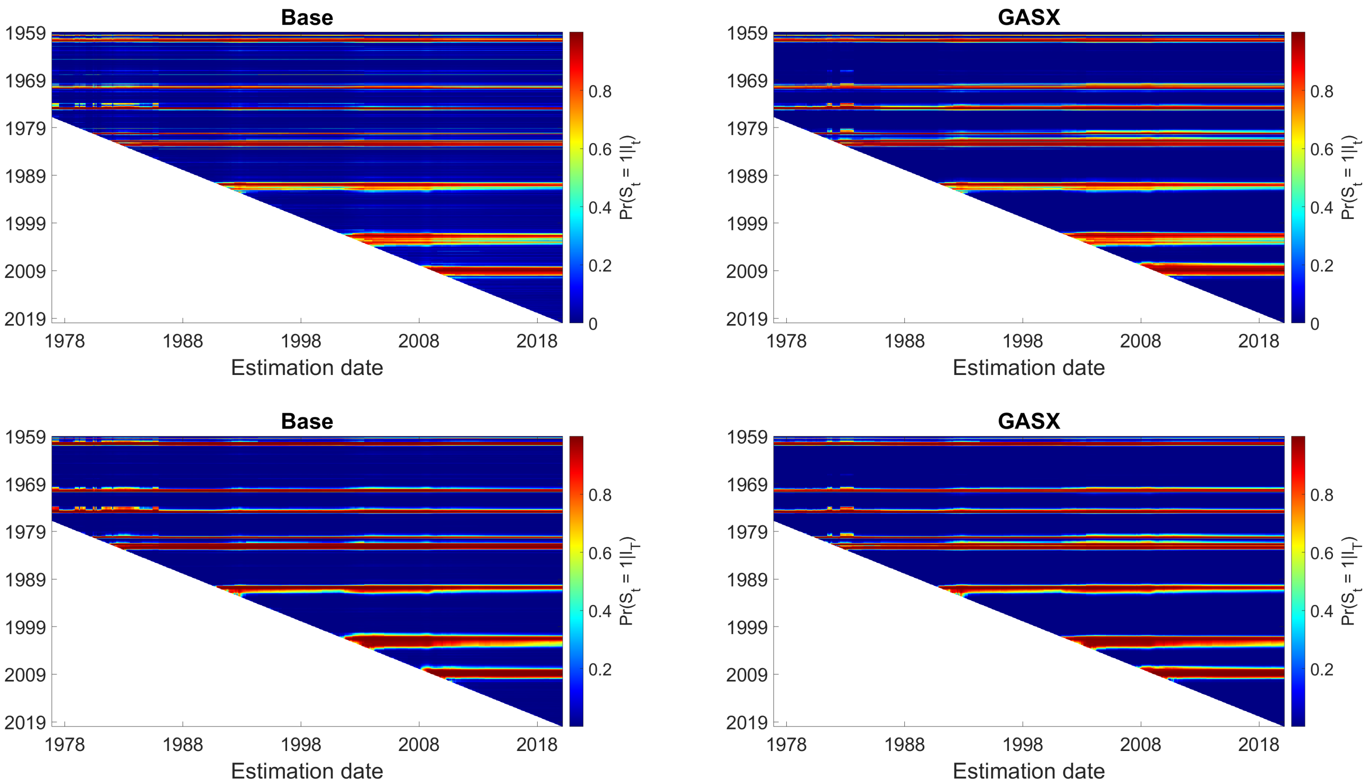


Note: This figure displays the evolution of the transition probability p_t^{01} over time for the alternative DFMS model specifications. The dotted line represents the constant p^{01} estimated by the base model. Exo-$(\cdot)$ denotes the autoregressive specification that only makes use of $(\cdot)$ to drive p_t^{01}, whereby STS denotes the standardized TS 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-TS<0 denotes a specification without autoregressive dynamics that simply considers two different levels of p^{01} depending on the sign of the TS. Finally, the shaded areas reflect the recession periods as determined by the NBER.

D Real-time

D.1 Complete history of state probabilities

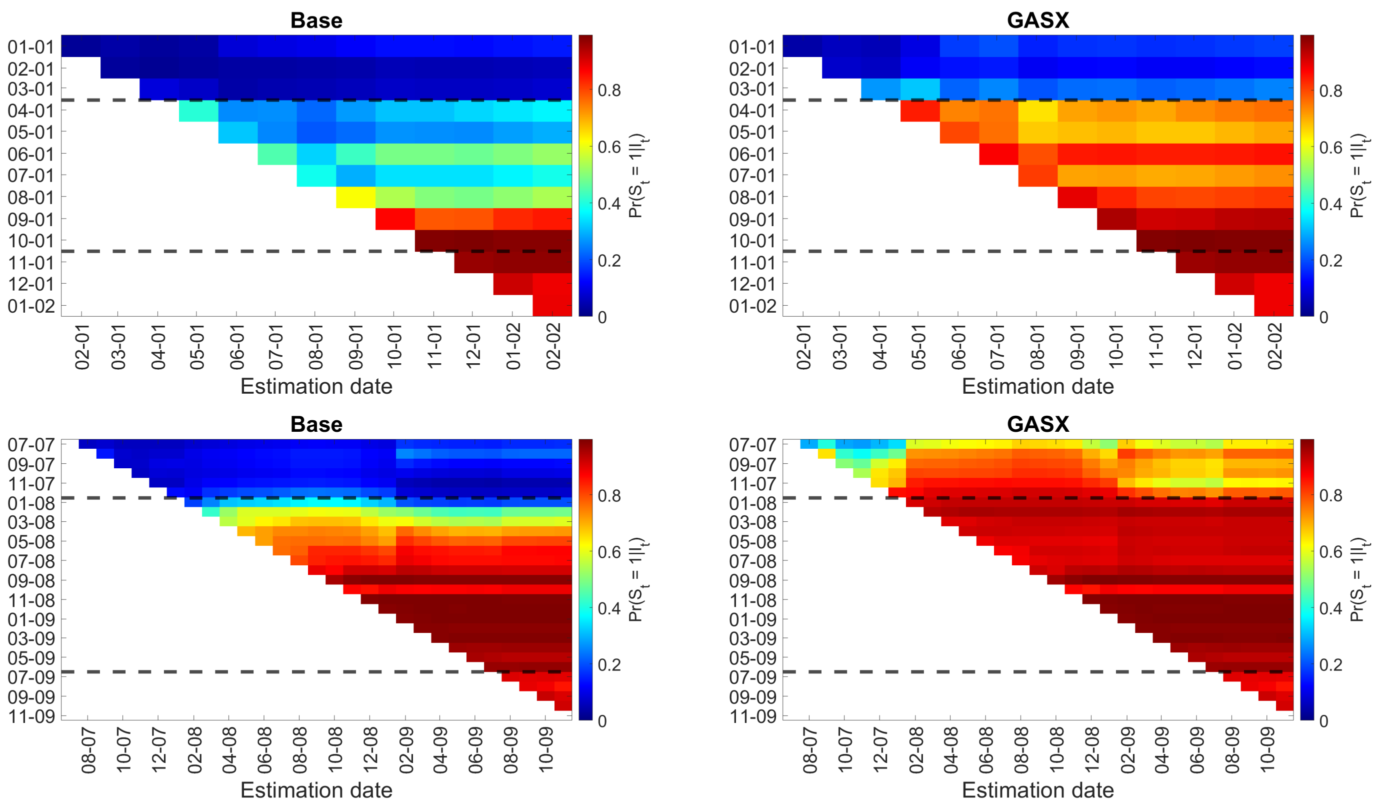
Figure D.1: Complete history of real-time filtered and smoothed state probabilities for the base DFMS model and the GASX extension.



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

D.2 History of filtered state probabilities around 2001 and 2008

Figure D.2: History of real-time filtered state probabilities for the base DFMS model and the GASX 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 GASX 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.

D.3 Troughs

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

$\tau = 0.65$

Trough date			*Announcement date*		
Base	**GASX**	**NBER**	**Base**	**GASX**	**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)	21 (9)	Nov 2001	5	5	Jul 2003
5 (4)	5 (4)	Jun 2009	-5	-5	Sep 2010

$\tau = 0.8$

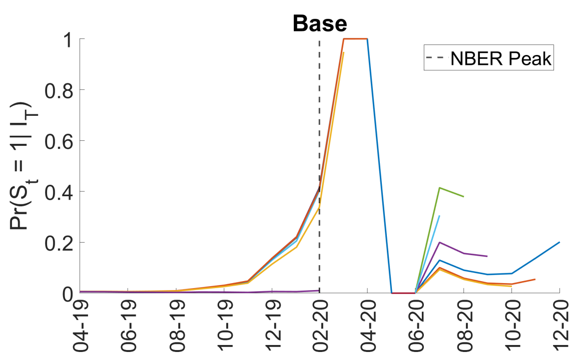
Trough date			*Announcement date*		
Base	**GASX**	**NBER**	**Base**	**GASX**	**NBER**
-1 (-1)	-1 (-1)	Jul 1980	-9	-9	Jul 1981
-1 (-1)	-1 (-1)	Nov 1982	-5	-5	Jul 1983
1 (0)	1 (0)	Mar 1991	-16	-16	Dec 1992
1 (2)	4 (3)	Nov 2001	-15	-12	Jul 2003
4 (3)	4 (3)	Jun 2009	-7	-5	Sep 2010

Note: This table contains the monthly differences in obtained initial trough dates of the base model and GASX extension with the NBER database. The numbers 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.

E COVID-19

We investigate the ability of our models to describe the COVID-19 recession that starts in March 2020, immediately after the end of the sample period considered so far. Figure E.1 depicts the real-time evolution of the smoothed recession state probabilities for the base and GASX models for the period March 2020 until December 2020. Because of the enormous magnitude of the movement in the coincident variables in this period, particularly in April 2020, we opt to use data up to and including February 2020 to estimate the model parameters. Including this period would dominate the estimation and yield a degenerate model almost solely designed to fit April 2020.

Figure E.1: Real-time smoothed recession probabilities for the base DFMS model and the GASX extension for the COVID-19 recession period, March 2020 until December 2020.



Note: The left and right plot depict the real-time evolution of the smoothed recession state probabilities for the base and GASX model respectively. Parameters are estimated using data up to and including February 2020. Finally, the black vertical dashed line is the peak date as established by the NBER.

In Figure E.1, we observe that both models are completely caught off guard, with an abrupt large increase in recession probabilities when data for March 2020 becomes available. Note that the heightened recession probabilities before March 2020 for later estimation moments is simply the result of smoothing; neither model is able to anticipate the COVID-19 period. This is not too surprising given the unusual cause and nature of this recession. The large rebound of the coincident variables in June 2020 in combination with the persistence of the factor yield a small artificial recession signal in July 2020. The GASX model in addition provides a secondary mild signal in December 2020 due to a negative employment growth.

We conclude that the magnitude and timing of the COVID-19 shocks demand special treatment as suggested among others by Ng (2021). Leaving these aberrant observations out of the parameter estimation procedure and running our model as is, such as in Figure E.1, presents an easy practical solution to do so. A more technical solution would be to use a robustified Kalman filter such as the one based on a t-distribution by Meinhold and Singpurwalla (1989). In view of also the Markov-switching and score-driven components of our setup, this is however non-trivial and left for future research.

References

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

Blasques, F. et al. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics* **227**, 325–346.

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.

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.

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

Meinhold, R.J. and N.D. Singpurwalla (1989). Robustification of Kalman filter models. *Journal of the American Statistical Association* **84**, 479–486.

Ng, S. (2021). COVID-19 and Estimation of Macroeconomic Factors. *Available at: http://arxiv.org/pdf/2103.02732*.