

Previsão das exportações com seleção de previsores

Francisco Dias
Banco de Portugal

Nuno Lourenço
Banco de Portugal e Nova SBE

António Rua
Banco de Portugal e Nova SBE

Resumo

Este artigo aplica modelos de fatores para prever as exportações mensais portuguesas, recorrendo a uma base de dados internacional que inclui os principais parceiros comerciais do país. Documentam-se ganhos de previsão assinaláveis até um horizonte de doze meses de distância quando são considerados conjuntamente indicadores qualitativos para estes países e os previsores são pré-selecionados antes da estimação dos fatores. Recorrendo apenas a dados nacionais e sem pré-seleção de previsores resulta em ganhos de previsão superiores para efeitos de *nowcasting*. Constata-se, assim, que os dados dos principais parceiros de Portugal são mais informativos para prever a h -períodos de distância. Por sua vez, quando os indicadores quantitativos e qualitativos são agrupados, os ganhos de previsão, em geral, não aumentam. (JEL: C38, C55, F47)

Introdução

A previsão de séries temporais macroeconómicas é de extrema importância para os decisores de política monetária e orçamental monitorizarem ou avaliarem desenvolvimentos em qualquer economia. Os avanços recentes na previsão de curto prazo têm sido ancorados no uso de bases de dados de grande dimensão, onde os desenvolvimentos na tecnologia de informação permitem hoje em dia aceder e manusear centenas de séries temporais em tempo real. Os indicadores quantitativos e qualitativos estão no centro deste contexto rico de informação. Enquanto os primeiros são baseados em informação quantitativa, os últimos assentam em inquéritos de conjuntura de atividade económica que são caracterizados pela natureza qualitativa das suas questões (por exemplo, inquéritos harmonizados regularmente divulgados pela Comissão Europeia para diferentes setores na União Europeia). O interesse em recorrer a dados qualitativos para prever variáveis macroeconómicas tem sido enfatizado na literatura (veja-se, por exemplo, Bańbura e Rünstler (2011) e Hansson *et al.* (2005) para uma aplicação para prever o crescimento do PIB). Uma vantagem central dos indicadores qualitativos reside na sua rápida disponibilidade, dado que a maioria destes inquéritos são publicados apenas alguns dias após o período de referência.

O elevado rácio sinal-ruído destes indicadores fornece conteúdo informativo substantivo sobre o estado da economia e a sua natureza abrangente permite uma vasta cobertura setorial. Para além disso, dado que algumas questões se referem a desenvolvimentos futuros, estes indicadores fornecem informação atempada sobre a possível evolução prospetiva da economia. Como estas séries qualitativas não estão sujeitas a revisões, a fiabilidade em tempo real também é garantida.

O uso de dados qualitativos a par de dados quantitativos para a previsão de curto prazo tem sido muito atrativo na Europa, onde estes inquéritos se encontram amplamente disponíveis para um período longo (veja-se, por exemplo, Schumacher (2007) para uma aplicação para prever o PIB para a Alemanha, Rünstler *et al.* (2009) para prever o PIB para vários países europeus e Angelini *et al.* (2011) para a área do euro). Enquanto a exploração de dados num contexto rico de informação tem sido amplamente documentada na literatura, poucos autores se têm focado no papel explícito de adicionar dados de outros países para prever variáveis macroeconómicas nacionais. Nesta área, sublinhe-se as contribuições de Brisson *et al.* (2003), onde os autores levam a bordo o conteúdo preditivo de variáveis dos Estados Unidos assim como de outros países para prever o PIB e a inflação para o Canadá. Da mesma forma, Schumacher (2009) considera o papel da informação da área do euro e restantes países do G7 para prever o PIB alemão.

Num contexto rico de informação, prever variáveis macroeconómicas implica extrair informação útil de um grande número de previsores. Os modelos de fatores têm sido bastante populares em tais exercícios, onde o conteúdo informativo de um vasto painel de séries temporais é sumariado num número pequeno de fatores que são então usados para efeitos de previsão. Entre as aplicações com modelos de fatores destacam-se as contribuições seminais de Stock e Watson (1999, 2002a,b) para prever algumas das principais variáveis macroeconómicas dos Estados Unidos, Marcellino *et al.* (2003) para prever a inflação e a atividade real para a área do euro e Rünstler *et al.* (2009) para um estudo que inclui vários países europeus.

Contudo, o uso de um conjunto de informação mais alargado para a previsão com modelos de fatores pode não melhorar a qualidade preditiva. De facto, estender uma base de dados para a estimação de fatores pode conduzir a piores resultados de previsão se as séries adicionais forem muito ruidosas ou se o poder preditivo for fornecido por um fator que é dominante numa base de dados menor, mas acaba por ser dominado numa base de dados mais extensa (veja-se Boivin e Ng (2006)). Reduzir a influência de previsores não informativos para prever uma variável macroeconómica deu origem a uma nova linha de abordagem na literatura. Neste sentido, Bai e Ng (2008) propuseram técnicas de regressão penalizadas para selecionar previsores, em particular *Least-Angle Regression with Elastic Net* (daqui em diante LARS-EN), onde a seleção de um subconjunto de previsores prévia à estimação dos fatores é levada a cabo para prever a inflação nos Estados Unidos. Outros exemplos

da importância da pré-seleção de previsores antes da estimação de fatores incluem Schumacher (2007) que apresenta uma aplicação para prever o PIB alemão e Li e Chen (2014) que se focam em diversas variáveis relevantes no acompanhamento da economia e política monetária para os Estados Unidos.

Neste artigo investiga-se o papel da informação contida em dados de outros países para prever fluxos de comércio internacional, com ênfase nas exportações de bens de Portugal, recorrendo à extensão de uma base de dados para cobrir os principais parceiros comerciais do país.¹ Assim, contribui-se para a vertente de literatura que se baseia em dados internacionais para prever variáveis nacionais. O elevado grau de interrelação da economia portuguesa com o resto do mundo fundamenta o uso desta base de dados heterogênea. Dado o seu tamanho extenso, recorre-se à pré-seleção de previsores através do algoritmo LARS-EN e infere-se sobre a utilidade da seleção dos previsores antes da estimação dos fatores para aumentar a qualidade preditiva. Exploram-se dados mensais para *nowcasting* e para prever as exportações de bens de Portugal numa base mensal até doze meses de distância.

O artigo encontra-se organizado da seguinte forma. A Secção 2 fornece uma breve análise das exportações portuguesas. A Secção 3 introduz o procedimento analítico usado no exercício de previsão. A Secção 4 descreve a base de dados subjacente à aplicação empírica. Os resultados empíricos são discutidos na Secção 5 e a Secção 6 conclui.

Uma análise breve das exportações portuguesas

Portugal tem registado um progresso notável em aumentar a sua integração no comércio mundial nas últimas décadas, onde o acesso à Comunidade Económica Europeia em 1986 se revelou um grande passo em frente. A importância relativa das exportações na economia tem crescido gradualmente. Contudo, diminuiu nitidamente em 2009 na sequência do colapso do comércio mundial, ganhando ritmo desde então (Gráfico 1).

Em Portugal, o rácio das exportações de bens no PIB em termos nominais aumentou cerca de 11 pontos percentuais desde 1993, situando-se perto de 27 por cento em 2016. Na sequência da crise económica e financeira recente, Portugal apresentou uma realocação gradual de *inputs* direcionados para a produção de bens para mercados estrangeiros. Como consequência, a importância relativa das exportações de bens na economia aumentou marcadamente a seguir à Grande Recessão.

1. Uma aplicação anterior de Cardoso e Duarte (2006) para a previsão de exportações de bens portuguesas baseia-se na utilização de um reduzido número de indicadores qualitativos num contexto de *bridge models*.

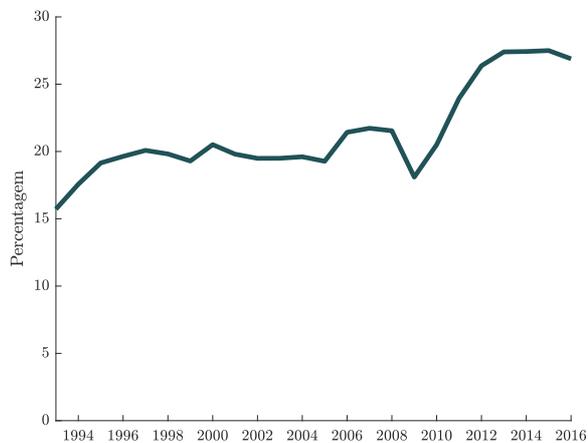


GRÁFICO 1: Rácio das exportações de bens no PIB em termos nominais.

Fonte: Instituto Nacional de Estatística.

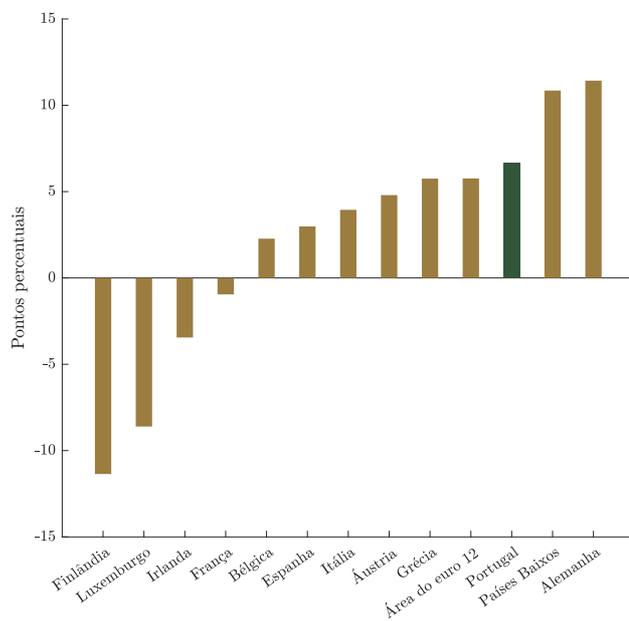


GRÁFICO 2: Variação do rácio das exportações de bens no PIB em termos nominais entre 2000 e 2016.

Fonte: Eurostat.

A comparação com os países membros fundadores da área do euro e a Grécia (AE-12) mostra que Portugal se destaca como sendo um dos países que registou maior aumento do rácio das exportações de bens no PIB durante o período em estudo (Gráfico 2). Posiciona-se à frente da maioria dos estados membros da AE-12, sendo ultrapassado apenas pelos Países Baixos e Alemanha.

No que concerne aos principais destinos das exportações de bens de Portugal, os países da área do euro representam a maior percentagem. No Quadro 1 encontram-se listados os principais parceiros comerciais em 2016. Em particular, as exportações para a Espanha representam mais de um quarto das exportações totais de bens, enquanto a França e a Alemanha representam mais de 10 por cento das exportações.

Principais parceiros comerciais	Frações (em percentagem)
Espanha	26,2
França	12,6
Alemanha	11,6
Reino Unido	7,0
Estados Unidos	4,9
Países Baixos	3,7
Itália	3,4
Angola	3,0
Bélgica	2,4

QUADRO 1. Principais destinos das exportações portuguesas de bens em 2016.

Procedimento analítico

Modelos de fatores

Esta secção inicia com uma discussão da representação de modelos de fatores que sustenta a previsão num contexto rico de informação. Seja X_t um vector coluna de séries temporais de previsores de dimensão N , observado para $t = 1, \dots, T$. O objetivo do exercício consiste em representar estas variáveis num modelo de fatores e utilizar os fatores estimados para produzir previsões a h -períodos de distância da variável de interesse, y , i.e., y_{t+h} , onde h denota o horizonte de previsão. As variáveis em X_t são representadas pela soma de duas componentes ortogonais: a componente comum, explicada por um pequeno número de fatores comuns não observáveis e a componente idiosincrática, determinada por choques específicos de cada variável.

O processo gerador dos dados para X_t admite uma representação de fatores estática escrita como:

$$X_t = \Lambda F_t + \xi_t \quad (1)$$

onde $F_t = (f_{1t}, \dots, f_{rt})'$ é um vector de fatores não observáveis de dimensão $(r \times 1)$, Λ é uma matriz $(N \times r)$ de pesos dos fatores desconhecidos e ξ_t é um vector de dimensão N de termos idiossincráticos. Conforme foi estabelecido por Stock e Watson (2002b), os fatores não observados podem ser estimados de forma consistente através de componentes principais sob hipóteses bastante gerais.

A estimação de fatores através de componentes principais visa maximizar a variância explicada na base de dados. Tipicamente, as primeiras componentes principais capturam uma fração considerável do co-movimento entre as séries na base de dados. Uma vez selecionado o número de fatores, a variável a prever é projetada eventualmente sobre os seus desfasamentos e sobre o conjunto de r fatores estimados. Daqui resulta a seguinte equação de previsão para a variável de interesse:

$$y_{t+h} = \alpha_0 + \sum_{i=1}^r \alpha_i \hat{F}_{t,i} + \sum_{j=0}^p \delta_j y_{t-j} + \varepsilon_{t+h} \quad (2)$$

onde α_0 é o termo constante, α_i denota os coeficientes associados a $\hat{F}_{t,i}$, isto é, as estimativas das componentes principais dos fatores na Equação (1), y_{t-j} representa a componente auto-regressiva da regressão, em que δ_j são os coeficientes correspondentes e p o número de desfasamentos.

O algoritmo LARS-EN

Tipicamente, quando os fatores são estimados, o conteúdo informativo para além daquele sumariado no pequeno conjunto de r fatores é ignorado. Assim sendo, informação útil para a variável a prever ou o horizonte de previsão em questão pode ser ignorada. Seguindo Bai e Ng (2008), recorre-se à previsão usando previsores selecionados. Por outras palavras, a relação entre y_{t+h} e X_t é analisada para selecionar o subconjunto de previsores $X_{t,A} \subseteq X_t$ antes da estimação de fatores.

Descreve-se agora um método baseado em regressões penalizadas que seleciona subconjuntos de variáveis de forma a comprimir a base de dados, deixando cair regressores não informativos. Ou seja, os coeficientes de regressão das variáveis menos informativas para prever a variável de interesse são penalizados. Seguindo Zou e Hastie (2005), Bai e Ng (2008) sugerem o uso do problema de otimização EN que é dado por:

$$\min_{\beta} \left\{ RSS + \lambda_1 \sum_{j=1}^N |\beta_j| + \lambda_2 \sum_{j=1}^N \beta_j^2 \right\} \quad (3)$$

onde RSS é a soma do quadrado dos resíduos de uma regressão de y_{t+h} sobre todos os regressores disponíveis, e λ_1 e λ_2 penalizam com a norma L_1 e L_2 de β , respetivamente.

A penalidade L_1 por si só resolve

$$\min_{\beta} \left\{ RSS + \lambda_1 \sum_{j=1}^N |\beta_j| \right\} \quad (4)$$

onde o parâmetro λ_1 controla para a dimensão de compressão e, desta forma, para o número de parâmetros que são fixados em zero. O método adiciona a regularização λ_1 à regressão dos mínimos quadrados ordinários, originando soluções que são esparsas em termos dos coeficientes de regressão. Este método é também conhecido como a solução-tipo do *Least Absolute Shrinkage and Selection Operator (LASSO)* de Tibshirani (1996).

Por sua vez, a penalidade L_2 isoladamente soluciona

$$\min_{\beta} \left\{ RSS + \lambda_2 \sum_{j=1}^N \beta_j^2 \right\} \quad (5)$$

o que para $0 \leq \lambda_2 < \infty$ comprime para zero os coeficientes dos previsores não informativos. Esta técnica também é conhecida como a penalidade L_2 da *ridge regression*.

Através da combinação das duas penalidades, i.e., as virtudes do LASSO e da *ridge regression*, a EN na Equação (3) permite a compressão dos coeficientes, eliminação de regressores e seleção eficiente de variáveis de toda a base de dados.

O algoritmo LARS fornece uma forma eficiente para resolver o problema de otimização da EN (ver Zou e Hastie (2005)). O algoritmo estima β e seleciona o subconjunto de previsores $X_{t,\mathcal{A}} \subseteq X_t$ através da resolução do problema de otimização definido na Equação (3), dados os parâmetros λ_1 e λ_2 . Na prática, a calibração de λ_1 é reformulada como uma regra para o número máximo de variáveis com β_j diferentes de zero a incluir na análise, i.e., o número de regressores $N_{\mathcal{A}} \leq N$ a ser incluído em $X_{t,\mathcal{A}}$. O procedimento funciona como se descreve a seguir. Inicia com todos os coeficientes iguais a zero e começa por encontrar o predictor mais correlacionado com a variável a prever. Avança o máximo na direção deste predictor até encontrar um segundo predictor que tenha tanta correlação com o resíduo contemporâneo. Em vez de continuar na direção do primeiro predictor, LARS prossegue numa direção equiangular entre os dois previsores até que uma terceira variável mais correlacionada entra no conjunto. O algoritmo prossegue de forma equiangular entre os três previsores, isto é, ao longo da direção de menor ângulo, até uma quarta variável entrar e por aí diante. O algoritmo formula estimativas $\hat{\mu} = X\hat{\beta}$ em passos sucessivos, onde em cada passo um regressor

é adicionado ao modelo, para que depois de k passos apenas k dos $\hat{\beta}_j$ sejam diferentes de zero (ver Efron *et al.* (2004) para detalhes).

Dados

O exercício de previsão compreende a previsão da taxa de crescimento de uma variável macroeconómica central, as exportações nominais de bens de Portugal. Esta variável é publicada sem qualquer ajustamento sazonal ou de efeitos de calendário numa base mensal 40 dias após o período de referência pelo Instituto Nacional de Estatística (INE).

O foco é na previsão da taxa de variação homóloga da série. Para além de ultrapassar o problema da sazonalidade determinística, esta escolha é motivada por várias outras razões, tais como a elevada volatilidade subjacente às taxas de variação em cadeia dos dados nominais de comércio ou a maior similitude entre variáveis medidas em variações homólogas e o perfil de vários indicadores qualitativos. Ao considerar taxas de variação homólogas, o ruído nos dados é reduzido e as irregularidades são atenuadas. Para uma discussão mais elaborada, veja-se Esteves e Rua (2012). Porém, mesmo quando se modeliza a variável dependente como uma taxa de variação homóloga, os efeitos de calendário ou de feriados móveis podem ser consideráveis e são expectáveis terem um impacto na variável de interesse. Estes efeitos são levados em conta recorrendo a variáveis determinísticas a serem descritas no exercício de previsão.

Os dados para Portugal consideram a base de dados mensal descrita em detalhe em Dias *et al.* (2015, 2018) que compreende inquéritos de opinião aos consumidores e às empresas, volume de negócios no comércio a retalho, produção industrial, volume de negócios na indústria e serviços, dormidas em Portugal, vendas de automóveis, vendas de cimento, ofertas de emprego e desemprego, consumo de energia, importações de bens, taxa de câmbio efetiva real, índice do mercado de ações para Portugal e uma série relativa a ATM/POS. Além disso, esta base de dados é estendida para incluir dados desagregados de preços no consumidor e no produtor.

Para além de dados nacionais, a base de dados foi estendida com dados mensais internacionais para abarcar informação dos principais parceiros comerciais de Portugal. Neste conjunto figuram Espanha, França, Alemanha, Reino Unido, Estados Unidos, Países Baixos, Itália e Bélgica. Os dados para Angola são escassos, daí este país não ter sido considerado na base de dados. Os dados mensais cobrem o período de janeiro-2000 a dezembro-2016.

Para cada parceiro comercial, o painel de variáveis inclui as principais variáveis de atividade económica, assim como variáveis qualitativas, totalizando, em média, 80 séries por país e 766 séries no total. As séries foram selecionadas para representar de um modo geral inquéritos de opinião aos consumidores e às empresas, preços, comércio a retalho, indústria e serviços

e mercado de trabalho. A repartição do número de variáveis entre dados quantitativos e qualitativos é apresentada no Quadro 2.²

	Número de séries	Dados qualitativos	Dados quantitativos
Portugal	145	39	106
Espanha	82	41	41
França	81	40	41
Alemanha	80	39	41
Reino Unido	80	39	41
Estados Unidos	75	20	55
Países Baixos	77	39	38
Itália	80	39	41
Bélgica	66	41	25
Total	766	337	429

QUADRO 2. Composição da base de dados.

No caso de Portugal e Espanha, para um número limitado de séries foi necessário recorrer ao algoritmo *Expectation-Maximization* sugerido por Stock e Watson (2002a) para balancear a base de dados no início do período amostral, dado que algumas séries estavam disponíveis para um período temporal mais curto. Em geral, com exceção dos dados qualitativos, tomaram-se logaritmos para todas as séries não negativas que não estavam em taxas ou unidades percentuais. A maioria das séries foi diferenciada para assegurar estacionaridade. Adicionalmente, seguindo Stock e Watson (2005), as séries foram corrigidas de *outliers*, onde o ajustamento correspondeu a substituir observações das séries transformadas com desvios absolutos excedendo seis vezes a amplitude interquartil pelo valor mediano das cinco observações imediatamente precedentes.

Exercício de previsão

Esta secção inicia com uma descrição detalhada do desenho do exercício de previsão. O procedimento recorre à estimação recursiva de parâmetros e fatores após a seleção dos previsores usando o algoritmo LARS-EN. Assim, não se restringe o conjunto de previsores selecionados a ser o mesmo para cada período temporal. Pelo contrário, os previsores são pré-selecionados em cada momento no tempo e para cada horizonte e a equação de previsão é re-estimada depois dos novos fatores serem estimados. Também foi considerado o caso onde não se faz pré-seleção de previsores, i.e., usa-se a abordagem usual dos modelos de fatores.

2. A lista de todas as séries e fontes de dados pode ser obtida dos autores mediante pedido.

Como modelo de referência considera-se como é habitual na literatura o modelo auto-regressivo de ordem p , $AR(p)$, em que o número p de termos auto-regressivos é determinado pelo critério BIC. Este modelo é alargado com três variáveis determinísticas adicionais para controlar os efeitos de calendário: o número de dias úteis em cada mês e duas variáveis binárias para os dois feriados móveis, Páscoa e Carnaval.

Um exercício fora da amostra de estimação é conduzido para avaliar o desempenho relativo do modelo de fatores com seleção de previsores face ao modelo de referência. O número de fatores estimados a ser incluído na equação de previsão é determinado pela minimização de uma versão modificada do critério BIC sugerido por Stock e Watson (1998).

O período fora da amostra de estimação estende-se desde janeiro-2009 a dezembro-2016, correspondendo a metade do período amostral disponível e o exercício de previsão é baseado na estimação por janela deslizante (*rolling window*) em que a amplitude da janela corresponde a 96 observações mensais (8 anos), o que coincide com a duração média típica do ciclo económico. A estimação através da janela deslizante aumenta a flexibilidade do modelo e a capacidade dos parâmetros lidarem com conteúdo preditivo da base de dados potencialmente variável. Todos os previsores candidatos estão disponíveis para o período t quando as exportações de bens são publicadas pelo INE. No caso dos dados qualitativos, quando as exportações para $t - 1$ são publicadas, os dados para o período t já estão disponíveis. Daí, quando se consideram apenas os dados qualitativos, considera-se o caso de *nowcasting* para além de previsões de um a doze meses de distância.

O desempenho do modelo é avaliado usando o Erro Quadrático Médio (EQM), e calcula-se o EQM relativo, usando o modelo auto-regressivo alargado como modelo de referência. Assim, um rácio inferior a um sinaliza que o modelo concorrente supera o modelo de referência. A significância estatística dos ganhos de previsão em relação ao modelo de referência é avaliada através do teste de Clark e West (2007).

Na análise empírica que se segue examinam-se dois conjuntos de previsores alternativos. Em primeiro lugar, analisam-se previsões baseadas apenas em dados qualitativos. Desta forma, exploram-se indicadores baseados em inquéritos de conjuntura para Portugal e para os seus principais parceiros comerciais. Em segundo lugar, alarga-se a base de dados para incluir em simultâneo indicadores qualitativos e quantitativos no exercício de previsão.

Previsões baseadas em dados qualitativos

O Quadro 3 reporta os resultados da previsão com dados qualitativos com previsores selecionados, i.e., com pré-seleção através de LARS-EN, definindo $\lambda_2 = 0.25$ como em Bai e Ng (2008) e Schumacher (2009), e sem pré-seleção de previsores. Caso seja aplicada a pré-seleção, o número de previsores

escolhidos é discretizado em cada linha, $N_{\mathcal{A}} = \{30, 40, \dots, 150\}$. As entradas no fundo da tabela correspondem ao caso sem pré-seleção. As bases de dados subjacentes compreendem apenas dados qualitativos totalizando 337 séries quando se consideram todos os países e 39 séries considerando apenas dados referentes a Portugal. As entradas na tabela reportam o EQM relativo do modelo de fatores *vis-à-vis* o modelo de referência para diferentes horizontes de previsão. As entradas a sombreado destacam o EQM relativo mínimo para cada horizonte de previsão.

	Horizonte de previsão												
	0	1	2	3	4	5	6	7	8	9	10	11	12
Previsores selecionados (N_A)													
30	1,12	1,06	0,89	0,81	0,83	0,90	0,89	1,04	0,81	0,90	1,23	1,04	1,17
40	1,15	1,03	0,71	0,73	0,71	1,06	0,66	0,91	0,85	0,93	0,89	0,71	1,01
50	0,95	0,96	0,64	0,64	0,73	1,11	0,53	0,81	0,90	0,83	0,76	0,69	1,17
60	0,92	0,83	0,63	0,70	0,73	1,08	0,49	0,70	0,83	0,77	0,81	0,65	1,04
70	0,91	0,84	0,63	0,66	0,73	0,97	0,50	0,69	0,92	0,67	0,85	0,53	0,90
80	0,94	0,87	0,66	0,69	0,76	1,17	0,51	0,77	0,97	0,76	0,96	0,59	0,82
90	0,87	0,88	0,68	0,72	0,78	0,91	0,56	0,83	1,10	0,82	0,97	0,63	0,97
100	0,93	0,94	0,65	0,77	0,65	0,81	0,69	0,81	1,09	0,96	0,88	0,66	0,99
110	0,96	0,84	0,69	0,75	0,51	0,91	0,63	0,82	1,03	0,98	0,79	0,69	0,99
120	0,89	0,94	0,70	0,73	0,50	0,94	0,72	0,96	0,95	1,04	0,83	1,03	0,92
130	0,95	0,92	0,74	0,70	0,52	1,07	0,84	1,09	0,97	0,88	0,86	1,15	0,81
140	0,96	0,89	0,73	0,71	0,52	1,10	0,93	0,96	1,06	0,97	1,03	1,21	0,87
150	0,91	0,85	0,72	0,71	0,58	1,14	1,02	0,90	1,05	0,98	1,13	1,35	0,93
Sem pré-seleção													
<i>Todas as séries</i>	0,89	0,82	0,71	0,68	0,83	1,39	1,60	1,85	2,19	1,39	1,43	1,54	1,42
<i>Apenas séries para PT</i>	0,78	0,86	0,87	0,78	0,84	0,97	0,96	1,07	0,94	1,07	1,05	2,14	2,45

QUADRO 3. EQM relativo das previsões baseadas em dados qualitativos *vis-à-vis* o modelo de referência.

Uma visão geral do conjunto de resultados revela que a maioria das entradas são inferiores a um, mostrando que há, em geral, ganhos de previsão em usar modelos de fatores com pré-seleção de previsores *vis-à-vis* o modelo de referência. Para efeitos de *nowcasting* os maiores ganhos de previsão são obtidos usando apenas dados qualitativos nacionais e sem pré-seleção de previsores, e estes excedem 20 por cento. Tal pode refletir o facto de os dados dos parceiros comerciais terem conteúdo informativo de carácter mais prospetivo. Neste sentido, os ganhos de previsão são notórios quando os indicadores qualitativos para os parceiros comerciais de Portugal são explorados e estes são maiores ainda quando se aplica pré-seleção de previsores através do algoritmo LARS-EN. Para prever a um mês de distância, embora o ganho máximo seja próximo de 20 por cento com todas as séries, i.e., sem pré-seleção, um valor semelhante pode ser obtido considerando apenas 60 previsores pré-selecionados. Para prever para o caso de horizontes de dois até doze meses, o uso de previsores pré-selecionados é consistentemente uma estratégia dominante, com ganhos que vão desde quase 20 por cento até aproximadamente 50 por cento *vis-à-vis* o modelo de referência. Em geral, o melhor desempenho preditivo é obtido escolhendo não mais de 70 variáveis entre 337 e os ganhos de previsão são estatisticamente significativos ao longo dos horizontes de previsão.

Para ilustrar a composição do conjunto de previsores selecionados usados para a estimação dos fatores, os Gráficos 3 e 4 mostram a fração média de previsores selecionados de cada país e inquérito setorial, respetivamente, para diferentes números de previsores (N_A) e horizontes de previsão (h). No Gráfico 3 constata-se que a fração média de séries selecionadas de Portugal aumenta com o horizonte de previsão e com o número de previsores selecionados, variando desde menos de 5 por cento até mais de 15 por cento. Focando a análise nos parceiros comerciais mais importantes, o mesmo se verifica globalmente para Espanha. No caso da França, a fração média de séries escolhida diminui com o horizonte de previsão e com o número de previsores, reduzindo-se de cerca de 30 por cento a menos de 10 por cento. Por sua vez, para a Alemanha a fração média é particularmente importante para horizontes mais curtos (cerca de 15 por cento) e menos relevante para horizontes próximos de um ano, mostrando-se relativamente estável com a variação do número de previsores selecionados.

No Gráfico 4 pode-se constatar que as variáveis associadas aos inquéritos à indústria são muito relevantes para prever para horizontes mais curtos, atingindo cerca de 50 por cento para um número pequeno de previsores, e a sua importância diminui com o horizonte de previsão e número de previsores. Em contraste, quando se consideram os inquéritos aos consumidores, a fração aumenta com o horizonte de previsão e, em menor grau, com o número de previsores, atingindo cerca de 40 por cento. As frações nos restantes inquéritos são relativamente estáveis, com os serviços a representar cerca de 10 por cento e o comércio a retalho e construção cerca de 15 por cento.

GRÁFICO 3: Fração média de previsores selecionados de cada país para diferentes números de previsores e horizontes de previsão.

GRÁFICO 4: Fração média de previsores selecionados de cada inquérito de conjuntura para diferentes números de previsores e horizontes de previsão.

O algoritmo EN envolve a escolha de λ_2 , que penaliza com a norma L_2 de β . À semelhança de Bai e Ng (2008), considerou-se $\lambda_2 = \{0.5, 1.5\}$ como análise de robustez. Os resultados não são muito sensíveis a esta escolha, o que está em linha com as conclusões de Bai e Ng (2008) e Schumacher (2009).

Agrupando dados quantitativos e qualitativos

De seguida estende-se a análise ao caso em que se juntam dados quantitativos e qualitativos. Por isso, explora-se o conteúdo informativo dos dados quantitativos para Portugal e seus parceiros comerciais, para além da base de dados assente em dados qualitativos anteriormente usada. Os resultados encontram-se reportados no Quadro 4. Uma breve análise mostra que a inclusão de dados quantitativos não parece trazer poder preditivo adicional para os horizontes mais longos *vis-à-vis* o modelo de referência. Por seu turno, os ganhos de previsão para horizontes mais curtos são, em geral, similares àqueles obtidos quando se recorre apenas a dados qualitativos. Note-se que a pré-seleção de previsores através do algoritmo LARS-EN aumenta o desempenho preditivo comparativamente ao caso sem pré-seleção e parece ter um papel relevante na previsão para horizontes mais curtos.

	Horizonte de previsão												
	1	2	3	4	5	6	7	8	9	10	11	12	
Previsores selecionados (N_A)													
30	1,12	0,90	1,17	0,88	1,25	1,31	1,33	1,50	0,94	1,18	1,82	1,16	
40	1,10	0,92	1,12	0,76	1,18	0,84	1,18	1,50	0,95	1,22	1,12	1,15	
50	0,98	0,87	1,10	0,77	1,17	0,82	1,27	1,05	1,36	1,34	1,36	1,13	
60	0,98	0,77	0,99	0,76	1,16	0,95	0,78	1,24	1,20	1,33	1,44	1,25	
70	0,84	0,67	0,98	0,72	1,52	0,87	0,92	1,30	1,24	1,56	1,37	1,27	
80	0,91	0,56	1,05	0,64	1,47	0,89	0,86	1,40	1,32	1,67	1,37	1,42	
90	0,83	0,57	1,06	0,58	1,60	0,96	0,89	1,44	1,45	1,88	1,57	1,48	
100	0,98	0,53	1,15	0,53	1,68	1,08	1,04	1,32	1,28	1,64	1,69	1,61	
110	0,86	0,61	1,14	0,49	1,49	1,26	1,01	1,42	1,48	1,67	1,62	1,50	
120	0,98	0,69	1,00	0,57	1,40	1,30	1,19	1,56	1,44	1,63	1,68	1,61	
130	0,97	0,68	1,01	0,62	1,37	1,46	1,30	1,67	1,39	1,75	1,76	1,72	
140	0,90	0,74	1,12	0,64	1,36	1,58	1,44	1,62	1,43	1,87	1,97	2,01	
150	0,92	0,72	1,07	0,66	1,47	1,69	1,46	1,80	1,72	2,08	2,05	1,97	
Sem pré-seleção													
<i>Todas as séries</i>	0,86	0,82	1,18	1,46	2,26	2,43	2,73	3,78	4,06	3,57	4,28	4,64	
<i>Apenas séries para PT</i>	0,92	0,87	1,15	1,52	1,96	1,40	2,57	2,53	2,10	3,22	2,00	1,46	

QUADRO 4. EQM relativo das previsões baseadas em dados qualitativos e quantitativos *vis-à-vis* o modelo de referência.

Considerações finais

Este artigo explora o papel de bases de dados internacionais para previsão, num contexto rico de informação, das exportações portuguesas de bens numa base mensal. Recorrendo ao conteúdo informativo dos principais parceiros comerciais do país, documentam-se ganhos de previsão assinaláveis até doze meses de distância quando dados qualitativos para estes países são agrupados e os previsores são pré-selecionados através do algoritmo LARS-EN antes da estimação dos fatores. Em geral, o melhor desempenho preditivo é conseguido com não mais de 70 variáveis escolhidas. Os ganhos de previsão de modelos de fatores usando previsores selecionados são estatisticamente significativos. Para além disso, agrupando dados quantitativos com dados qualitativos não parece trazer poder preditivo adicional para prever as exportações de bens.

Referências

- Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, e G. Rünstler (2011). "Short-term forecasts of euro area GDP growth." *Econometrics Journal*, 14, C25–C44.
- Bañbura, M. e G. Rünstler (2011). "A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP." *International Journal of Forecasting*, 27, 333–346.
- Bai, J. e S. Ng (2008). "Forecasting economic time series using targeted predictors." *Journal of Econometrics*, 146, 304–317.
- Boivin, J. e S. Ng (2006). "Are more data always better for factor analysis?" *Journal of Econometrics*, 132, 169–194.
- Brisson, M., B. Campbell, e J. W. Galbraith (2003). "Forecasting Some Low-predictability Time Series Using Diffusion Indices." *Journal of Forecasting*, 22, 515–531.
- Cardoso, F. e C. Duarte (2006). "A Utilização de Informação Qualitativa na Previsão das Exportações." *Banco de Portugal, Boletim Económico, Inverno 2006*, pp. 73–81.
- Clark, T. E. e K. D. West (2007). "Approximately normal tests for equal predictive accuracy in nested models." *Journal of Econometrics*, 138, 291–311.
- Dias, F., M. Pinheiro, e A. Rua (2015). "Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence." *Economic Modelling*, 44, 266–272.
- Dias, F., M. Pinheiro, e A. Rua (2018). "A bottom-up approach for forecasting GDP in a data-rich environment." *Applied Economics Letters*, 25, 718–723.
- Efron, B., T. Hastie, I. Johnstone, e R. Tibshirani (2004). "Least angle regression." *Annals of Statistics*, 32, 407–499.

- Esteves, P. S. e A. Rua (2012). "Previsões económicas de curto prazo para Portugal: uma síntese metodológica." *Banco de Portugal, Boletim Económico, Outono 2012*, pp. 143–155.
- Hansson, J., P. Jansson, e M. Löf (2005). "Business survey data: Do they help in forecasting GDP growth?" *International Journal of Forecasting*, 21, 377–389.
- Li, J. e W. Chen (2014). "Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models." *International Journal of Forecasting*, 30, 996–1015.
- Marcellino, M., J. Stock, e M. Watson (2003). "Macroeconomic forecasting in the euro area: country specific versus euro wide information." *European Economic Review*, 47, 1–18.
- Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. Den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, e C. Van Nieuwenhuyze (2009). "Short-Term Forecasting of GDP Using Large Datasets: A Pseudo Real-Time Forecast Evaluation Exercise." *Journal of Forecasting*, 28, 595–611.
- Schumacher, C. (2007). "Forecasting German GDP Using Alternative Factor Models Based on Large Datasets." *Journal of Forecasting*, 26, 271–302.
- Schumacher, C. (2009). "Factor forecasting using international targeted predictors: The case of German GDP." *Economic Letters*, 107, 95–98.
- Stock, J. e M. Watson (1998). "Diffusion indexes." *NBER Working Paper no. 6702*.
- Stock, J. e M. Watson (1999). "Forecasting inflation." *Journal of Monetary Economics*, 44, 293–335.
- Stock, J. e M. Watson (2002a). "Macroeconomic forecasting using diffusion indices." *Journal of Business and Economics Statistics*, 20, 147–162.
- Stock, J. e M. Watson (2002b). "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 97, 1167–1179.
- Stock, J. e M. Watson (2005). "Implications of dynamic factor models for VAR analysis." *NBER Working Paper Series 11467*.
- Tibshirani, R. (1996). "Regression Shrinkage and Selection via the LASSO." *Journal of Royal Statistical Society Series B*, 58, 267–288.
- Zou, H. e T. Hastie (2005). "Regularization and variable selection via the elastic net." *Journal of Royal Statistical Society Series B*, 67, 301–320.