

# Desajuste educacional no mercado de trabalho

Incidência e impacto salarial  
em Portugal



**FICHA TÉCNICA****Título**

Desajuste educacional no mercado de trabalho: incidência e impacto salarial em Portugal

**Autoria**

PLANAPP – Centro de Planeamento e de Avaliação de Políticas Públicas

**Data**

Abril 2026

**PLANAPP – Centro de Planeamento e de Avaliação de Políticas Públicas**

Campus XXI, Av. João XXI, n. 63

1000-300 Lisboa

[planapp@planapp.gov.pt](mailto:planapp@planapp.gov.pt)

[www.planapp.gov.pt](http://www.planapp.gov.pt)

# Índice

Índice de Figuras.....	3
Resumo Ilustrado .....	5
1. Enquadramento e conceitos .....	6
1.1. Relevância da análise sobre desajuste educacional.....	6
1.2. Definições operacionais.....	7
1.3. Medição do desajuste: principais abordagens e implicações.....	8
2. Dados e metodologia .....	10
2.1. Fonte de dados, população-alvo e âmbito.....	10
2.2. Métrica de desajuste vertical: âncora normativa e componente empírica .....	10
2.3 Efeitos salariais: medida, contrafactual e heterogeneidade .....	11
3. Prevalência da sobreeducação entre trabalhadores licenciados (2010–2021).....	12
3.1. Evolução agregada .....	12
3.2. Evolução por eixos de heterogeneidade .....	13
4. Efeitos da sobreeducação entre licenciados .....	28
4.1. Efeitos médios da sobreeducação .....	28
4.2. Heterogeneidade dos efeitos salariais da sobreeducação .....	30
5. Síntese e notas finais .....	42
Referências .....	45
Anexo 1 – Apêndice metodológico sobre deteção de ajuste educacional e inferência de efeitos salariais heterogêneos da sobreeducação .....	47

## Índice de Figuras

<b>Figura 1</b> – Evolução da prevalência de sobreeducação entre licenciados, 2010-2021 .....	13
<b>Figura 2</b> – Prevalência de sobreeducação entre licenciados, por sexo, 2010-2021 .....	14
<b>Figura 3</b> – Prevalência de sobreeducação entre licenciados, por idade e sexo, média 2010-2021 ...	15
<b>Figura 4</b> – Evolução da prevalência de sobreeducação: total <i>versus</i> novos entrantes (menos de 35 anos), 2010–2021 .....	16
<b>Figura 5</b> – Probabilidade de manutenção no mesmo posto de trabalho, para desajuste <i>versus</i> ajuste educacional, média 2010–2021 .....	17
<b>Figura 6</b> – Evolução da prevalência de sobreeducação entre licenciados, por nacionalidade, 2010-2021 .....	18
<b>Figura 7</b> – Sobreeducação por área de formação: escala (n.º de casos) e prevalência (taxa), 2021.	19
<b>Figura 8</b> – Evolução da prevalência de sobreeducação por área de formação: 2010 vs. 2021.....	20
<b>Figura 9</b> – Evolução do número de licenciados e da prevalência (taxa) de sobreeducação, 2012-2021 .....	22

<b>Figura 10</b> – Sobreeducação por setor: escala (n.º de casos) e intensidade (taxa), 2021 .....	23
<b>Figura 11</b> – Prevalência de sobreeducação entre licenciados do próprio setor, 2021 .....	24
<b>Figura 12</b> – Evolução do emprego nos setores de atividade com elevada prevalência de sobreeducação, 2010-2021 .....	25
<b>Figura 13</b> – Evolução da prevalência de sobreeducação por setor: 2010 vs. 2021 .....	26
<b>Figura 14</b> – Evolução do efeito médio da sobreeducação nos salários, 2014-2021 .....	28
<b>Figura 15</b> – Evolução do salário médio por hora observado e salário por hora contrafactual para trabalhadores sobreeducados (euros por hora), 2014-2021 .....	29
<b>Figura 16</b> – Distribuição (densidade) dos efeitos individuais da sobreeducação nos salários, 2014, 2018, 2021 .....	30
<b>Figura 17</b> – Efeito da sobreeducação nos salários por decis (média do decil), 2021 .....	31
<b>Figura 18</b> – Distribuição (densidade) dos efeitos individuais da sobreeducação nos salários, por sexo, 2021 .....	32
<b>Figura 19</b> – Evolução do efeito médio da sobreeducação nos salários ao longo da vida profissional, por sexo.....	33
<b>Figura 20</b> – Evolução do efeito médio da sobreeducação nos salários ao longo da vida profissional, por nacionalidade.....	34
<b>Figura 21</b> – Evolução do efeito mediano da sobreeducação no salário por área de formação: 2014 vs. 2021 .....	35
<b>Figura 22</b> – Salário médio por hora observado <i>versus</i> salário por hora contrafactual (salário esperado sem desajuste educacional), por área de formação, 2021 .....	36
<b>Figura 23</b> – Efeito médio da sobreeducação nos salários (%): matriz de interação idade <i>versus</i> setor de atividade .....	37
<b>Figura 24</b> – Efeito médio da sobreeducação nos salários (%): matriz de interação idade <i>versus</i> área de formação .....	38
<b>Figura 25</b> – Prevalência e efeito salarial mediano da sobreeducação por área de formação, 2021... 39	
<b>Figura 26</b> – Prevalência e efeito salarial mediano da sobreeducação por setor de atividade económica, 2021 .....	41

## Resumo Ilustrado

### Sobrieducação no mercado de trabalho em Portugal: um retrato heterogéneo e assimétrico

O fenómeno da sobrieducação em Portugal acarreta uma penalização salarial heterogénea e assimétrica, apresentando maior prevalência em certos segmentos demográficos, áreas de formação e setores de atividade.



Setores de atividade

Áreas de formação

# 1. Enquadramento e conceitos

No âmbito de uma série de Notas de Análise dedicadas à evolução dos salários em Portugal, a presente Nota centra-se no **desajuste (*mismatch*) educacional** no mercado de trabalho, isto é, na discrepância entre as qualificações formais dos trabalhadores e aquelas que são tipicamente esperadas nas ocupações que desempenham.

Em particular, esta Nota de Análise reúne evidência sobre o desajuste educacional entre trabalhadores licenciados em Portugal, no período compreendido entre **2010 e 2021**. O objetivo é separar (i) **prevalência** (quem está em sobreeducação e onde) e (ii) **impactos salariais** (quanto se perde/ganha e para quem), de modo a produzir informação para apoiar políticas de transição ensino–mercado de trabalho, integração de qualificações e desenvolvimento setorial.

Para esse efeito, recorre-se aos microdados administrativos dos Quadros de Pessoal, que permitem observar, com granularidade, perfis de remuneração, escolaridade, ocupações e setores de entre os trabalhadores do setor privado e trabalhadores com contratos individuais de trabalho no setor público, tornando possível caracterizar a dimensão do fenómeno e os seus potenciais efeitos salariais.

Como nota de leitura, salienta-se que os resultados apresentados devem ser interpretados como **associações ajustadas** estimadas com base nos referidos microdados administrativos e num indicador calibrado. Não se trata de estatísticas com grau elevado de precisão e certeza nem as estimativas de efeitos constituem inferência causal estrita. Ainda assim, a articulação entre prevalência da sobreeducação, penalização salarial e concentração por segmentos fornece evidência suficiente para orientar prioridades e desenho de instrumentos de política pública.

## 1.1. Relevância da análise sobre desajuste educacional

A educação é uma determinante central do bem-estar individual e coletivo: associa-se a melhores oportunidades de emprego, rendimentos mais elevados, maior mobilidade social e ganhos de produtividade. Contudo, quando o capital humano disponível não é plenamente utilizado – por exemplo, porque trabalhadores altamente qualificados têm ocupações onde desempenham funções que exigem qualificações inferiores (i.e., trabalhadores estão em situação de sobreeducação) – podem emergir ineficiências de alocação, perdas de produtividade e trajetórias salariais menos favoráveis. Para além do impacto económico, o desajuste pode também estar associado a custos não monetários, como menor satisfação profissional e menor qualidade do ajuste trabalhador–posto de trabalho.

Do ponto de vista empírico, a literatura mostra de forma consistente que, em particular, a sobreeducação tende a associar-se a penalizações salariais e a resultados menos favoráveis no mercado de trabalho, embora as magnitudes estimadas dependam do método de medição e do contexto institucional. Estudos clássicos mostram que a escolaridade “em excesso” relativamente ao requisito ocupacional se traduz, em média, em retornos marginais inferiores aos retornos da escolaridade “requerida” (Rumberger, 1987; Verdugo & Verdugo, 1989). Evidência mais recente

também liga a sobreeducação a insatisfação no trabalho e a alterações na dispersão dos retornos do ensino superior (Green & Zhu, 2010). Ao nível microeconómico, a persistência do desajuste pode ainda ser transitória ou duradoura, variando com ciclos económicos, trajetórias individuais e mobilidade ocupacional (Frei & Sousa-Poza, 2012).

Ao nível das empresas e do sistema económico, a persistência de desajustes pode refletir falhas de alocação de capital humano, com desperdício de investimento educativo e perdas de produtividade quando o desajuste implica que competências são subutilizadas ou mal distribuídas. Nesta linha, a evidência baseada no PIAAC mostra que medidas de *skills mismatch* (proveniência e uso de competências) têm relações sistemáticas com resultados laborais e produtividade (Allen et al., 2013; McGowan & Andrews, 2017).<sup>1</sup> Do ponto de vista de política pública, estes custos tornam-se particularmente relevantes quando a estrutura produtiva não acompanha o ritmo de expansão das qualificações e quando certas fricções de integração (por exemplo, reconhecimento de credenciais) ampliam a probabilidade de desajuste em certos grupos (Quintini, 2011).

De facto, nas últimas décadas, verificou-se uma expansão significativa do acesso ao ensino superior, tornando mais premente a questão da capacidade de absorção do mercado de trabalho. Quando o crescimento da oferta de diplomados excede a criação de postos de trabalho com requisitos compatíveis, pode aumentar a incidência de sobreeducação e o recurso a ocupações historicamente classificadas como não exigindo ensino superior.<sup>2</sup>

Em termos conceptuais, estas tensões são consistentes com mecanismos discutidos na literatura sobre medição e dinâmica do desajuste: mudanças tecnológicas, reorganização do trabalho e ajustamentos lentos entre oferta e procura de competências podem gerar desfasamentos persistentes (Mendes de Oliveira *et al.*, 2000). Por outro lado, a evidência sugere que uma parte do fenómeno pode ser transitória, associada a fases iniciais de carreira, fricções de procura de emprego e mobilidade subsequente para ocupações mais ajustadas (Frei & Sousa-Poza, 2012; Quintini, 2011).

## 1.2. Definições operacionais

O desajuste educacional assume duas formas conceptuais principais, com implicações distintas para a análise e para a política pública.

**Desajuste vertical** ocorre quando o nível de escolaridade do trabalhador é inferior ou superior ao nível tipicamente exigido para a ocupação, podendo, assim, existir:

- **subeducação** (*undereducation*), quando o nível de escolaridade é inferior ao esperado;
- **sobreeducação** (*overeducation*), quando o nível de escolaridade é superior ao esperado.

<sup>1</sup> PIAAC refere-se ao *Programme for the International Assessment of Adult Competencies*, um estudo internacional periódico da OCDE que avalia as competências cognitivas e profissionais de adultos (16-65 anos).

<sup>2</sup> Esta dinâmica surge de forma clara em medidas normativas usadas em estatísticas europeias, como se verá abaixo: o EUROSTAT define como sobrequalificados os trabalhadores com ISCED 5–8 empregados em ocupações de baixo ou médio nível de qualificação, com base no *Labour Force Survey*.

**Desajuste horizontal** refere-se ao desalinhamento entre a *área de estudo* (ou domínio de formação) do trabalhador e a área/ocupação exercida, mesmo quando o nível de escolaridade é adequado.

Nesta Nota de Análise, o foco analítico recairá sobre o fenómeno da sobreeducação e, especificamente, no caso dos trabalhadores com licenciatura, dado tratar-se da situação mais premente em Portugal.

### 1.3. Medição do desajuste: principais abordagens e implicações

Não existe apenas uma única abordagem para a medição do desajuste educacional e as conclusões sobre prevalência e impactos dependem do método adotado.

Uma primeira família de medidas baseia-se em procedimentos de **autoavaliação**: recolhe perceções dos trabalhadores sobre se o seu nível de educação/competências corresponde ao requerido ou qual o nível necessário para desempenhar as tarefas. Esta abordagem capta dimensões subjetivas relevantes e pode abranger desajuste vertical e horizontal, mas é vulnerável a enviesamentos de perceção e a comparabilidade limitada entre grupos e países. A literatura discute amplamente estas limitações e mostra que diferentes formas de perguntar e de operacionalizar a autoavaliação podem produzir prevalências e associações salariais distintas (Chevalier, 2003; Frei & Sousa-Poza, 2012; Green & Zhu, 2010; Verhaest & Omey, 2006).

Uma segunda família, frequentemente designada por **realized matches**, infere o requisito típico de nível de educação a partir da tendência central da distribuição de escolaridade observada dentro de cada ocupação (por exemplo, média ou moda), classificando como desajustados os trabalhadores que se afastam dessa referência. É operacionalizável com dados administrativos, mas sensível a escolhas paramétricas (métrica de centralidade, limiares) e pode confundir mudanças de coorte de trabalhadores com mudanças reais de requisitos se não incorporar dinâmica temporal (Kiker et al., 1997; Mendes de Oliveira et al., 2000; Quinn & Rubb, 2006; Verdugo & Verdugo, 1989).

Uma terceira família é **normativa** (também designada por *job analysis*): define *a priori* o nível de educação requerido por ocupação, com base em análise especializada. É transparente e comunicável, mas pode ser dispendiosa e tornar-se rapidamente obsoleta com a transformação das ocupações e das tarefas (Flisi et al., 2017; McGoldrick & Robst, 1996; Rumberger, 1987).

No plano internacional, o EUROSTAT publica estatísticas regulares sobre **sobreeducação** entre trabalhadores com ensino superior. A métrica utilizada é normativa e define como sobreeducados os indivíduos com ISCED 5–8 empregados em ocupações identificadas *a priori* como não requerendo ensino superior, tipicamente associadas aos grandes grupos 4 a 9 da ISCO (com base no *Labour Force Survey*). Importa, no entanto, explicitar uma limitação estrutural desta abordagem: a correspondência ISCO–ISCED é **fixa** e, por construção, não reflete mudanças graduais nos requisitos de qualificação dentro de ocupações ao longo do tempo. Em contextos de transformação tecnológica e reorganização do trabalho, ocupações que se tornaram mais complexas podem continuar classificadas como não

exigindo ensino superior, inflacionando a medição de sobreeducação por via de uma âncora normativa rígida.

O exercício empírico conduzido nesta Nota de Análise recorre a uma solução híbrida (ver Secção 2). Neste contexto, os microdados administrativos dos Quadros de Pessoal oferecem uma oportunidade de inserir dinâmica na análise e aprofundá-la com elevada granularidade por ocupação, setor, região e perfil do trabalhador.

A heterogeneidade demográfica e institucional também deve ser considerada: por exemplo, diferenças por idade podem refletir fricções na transição escola–mercado de trabalho e evolução ao longo da carreira, enquanto diferenças por nacionalidade podem refletir fricções no reconhecimento de qualificações e barreiras linguísticas (Allen et al., 2013; McGowan & Andrews, 2017; Quintini, 2011). Estas dimensões são exploradas na análise descritiva de prevalências e na heterogeneidade de efeitos salariais.

A análise do desajuste educacional nesta Nota de Análise separa dois planos:

- **Prevalência:** quem está em desajuste de sobreeducação, onde e como evolui no tempo (por perfis demográficos, áreas de educação e setores de atividade);
- **Impacto salarial:** qual a diferença salarial associada ao desajuste de sobreeducação (em percentagem e em euros/hora) e como esse impacto varia entre grupos (heterogeneidade).

Esta separação é relevante para identificar segmentos onde a combinação de prevalência e penalização sugere maior custo económico e, por isso, maior prioridade de política.

A secção seguinte descreve as escolhas de dados e a estratégia metodológica adotada para operacionalizar a medição do desajuste em dados administrativos e para estimar os respetivos impactos salariais.

## 2. Dados e metodologia

A análise baseia-se em microdados administrativos longitudinais dos Quadros de Pessoal (QP), do Ministério do Trabalho, Solidariedade e Segurança Social (MTSSS), cobrindo o setor privado e contratos individuais de trabalho do sector público, o que permite uma caracterização granular por ocupação, setor, região, demografia e remuneração. Como os QP não incluem requisitos educacionais para cada ocupação, a identificação de desajuste educacional é operacionalizada através de uma estratégia híbrida: (i) uma **âncora normativa** alinhada com a prática do EUROSTAT para trabalhadores com ensino superior e (ii) **detetores empíricos** calibrados para capturar a evolução temporal dos requisitos ocupacionais. A metodologia completa de calibração e validação do indicador encontra-se documentada no Anexo 1 – Apêndice Metodológico, que complementa esta Nota de Análise.

### 2.1. Fonte de dados, população-alvo e âmbito

A fonte principal são os ficheiros anuais dos QP (GEP–MTSSS), com observações ao nível trabalhador–empresa. O âmbito desta Nota de Análise incide sobre trabalhadores licenciados no período compreendido entre 2010 e 2021. Esta delimitação é, em parte, forçada pela âncora normativa utilizada, cuja definição só é apropriada para trabalhadores com ensino superior. A escolha de licenciados mitiga riscos associados a níveis educacionais com maior instabilidade de codificação e maior substituíbilidade de requisitos (como é o caso do mestrado).

### 2.2. Métrica de desajuste vertical: âncora normativa e componente empírica

O desajuste vertical é medido classificando cada trabalhador em três estados operacionais: *em subeducação* (escolaridade inferior ao esperado), *com educação adequada* (alinhado) e *em sobreeducação* (escolaridade superior ao esperado). A componente normativa funciona como referência de calibração, enquanto a componente empírica estima o “esperado” dentro de **células contextuais** (por exemplo, ocupação da Classificação Portuguesa de Profissões a 3 dígitos ou ocupação-setor), permitindo que esse padrão evolua ao longo do tempo, acompanhando a transformação ocupacional. A escolha de desenho das células, regras mínimas de dimensão e opções de detetor são tratadas como decisões metodológicas avaliadas face à âncora normativa, evitando a existência de decisões arbitrárias do utilizador face aos parâmetros e hiperparâmetros a aplicar.

A calibração dos detetores e a seleção do método são realizadas com validação cruzada aninhada e separação temporal e das unidades entre fases (*calibration*, *validation* e *test*). A função-objetivo de calibração privilegia a correta identificação da categoria *educação adequada*, refletindo a maior

fiabilidade da âncora normativa na categoria de adequação para níveis ISCED 5-8.<sup>3</sup> O procedimento completo (incluindo definição de janelas, *Nested CV* e famílias de detetores) encontra-se no Apêndice Metodológico.

## 2.3 Efeitos salariais: medida, contrafactual e heterogeneidade

Os efeitos salariais associados à sobreeducação são estimados em salário por hora, comparando o salário observado com um salário contrafactual esperado na situação de *educação adequada* (mantendo características observáveis relevantes). A estimação privilegia uma leitura de associação ajustada e reporta heterogeneidade por dimensões (por exemplo, área de educação e setor), permitindo identificar segmentos onde coexistem elevada prevalência e maior penalização. As opções de estimação estão descritas no Apêndice Metodológico.

Os resultados são apresentados nas secções seguintes separados em dois blocos: (i) prevalência e perfil do desajuste vertical (sobreeducação) entre trabalhadores licenciados; (ii) impactos salariais e a sua heterogeneidade.

---

<sup>3</sup> ISCED refere-se à classificação internacional *International Standard Classification of Education*.

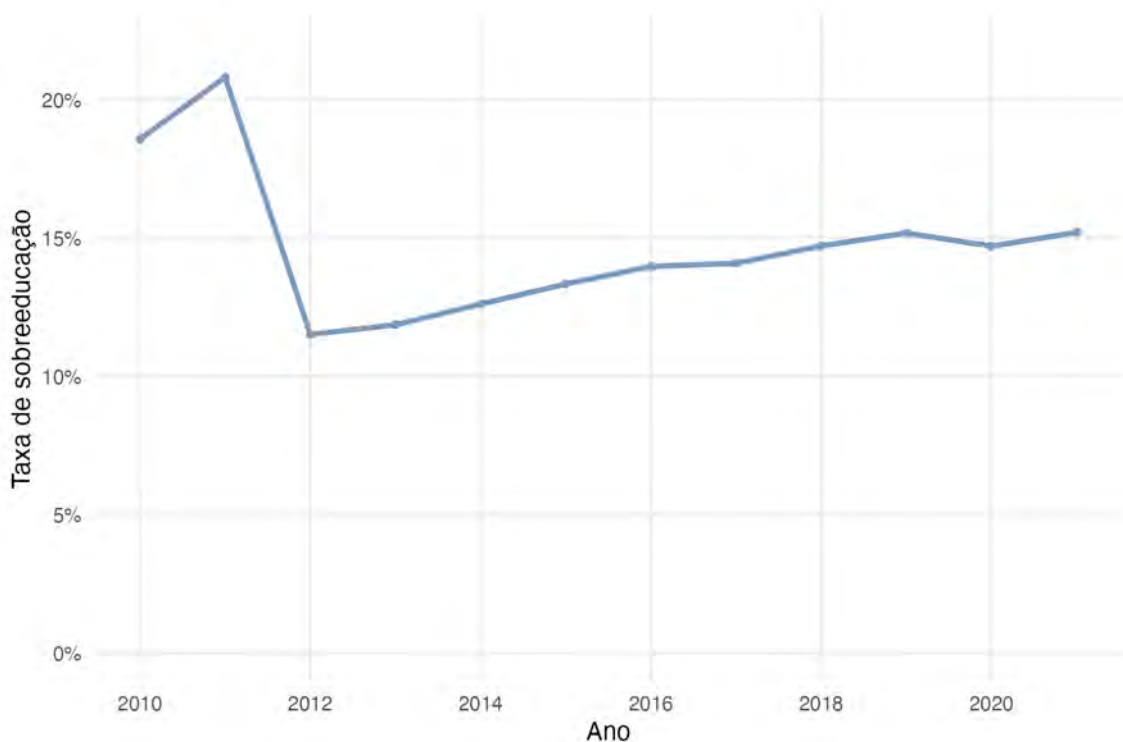
### 3. Prevalência da sobreeducação entre trabalhadores licenciados (2010–2021)

Esta secção quantifica a evolução da prevalência da sobreeducação entre trabalhadores licenciados no setor privado e em contratos individuais de trabalho no setor público, no período 2010–2021. O objetivo é duplo. Primeiro, fornecer o retrato agregado e identificar se existem tendências de crescimento, estabilidade ou reversão. Segundo, decompor esse retrato ao longo de eixos de heterogeneidade – fase de entrada no mercado de trabalho, idade, sexo, nacionalidade, setor de atividade e área de formação – para distinguir (i) efeitos de composição (quem está no mercado em cada momento) de (ii) fenómenos de persistência (manutenção do desajuste educacional ao longo da carreira). Esta decomposição é substantivamente relevante para a discussão de política pública: a sobreeducação é um indicador de subutilização de competências e, portanto, portador de sinal de ineficiências no mercado de trabalho e, por essa via, de uma economia a funcionar abaixo do seu potencial. Esta decomposição leva a entender melhor a capacidade de absorção de qualificações do tecido produtivo e sobre a qualidade da transição educação–mercado de trabalho, permitindo sinalizar prioridades de resposta por parte da política pública e adequá-las a setores e áreas de formação distintas.

#### 3.1. Evolução agregada

A Figura 1 apresenta a evolução da taxa de sobreeducação (prevalência) entre trabalhadores licenciados ao longo do período 2010-2021. Observa-se uma dinâmica não monótona, marcada por uma quebra inicial significativa e, posteriormente, por uma escalada gradual.

**Figura 1** – Evolução da prevalência de sobreeducação entre licenciados, 2010-2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

Entre 2010 e 2012, verifica-se uma queda acentuada: a taxa inicia-se em torno de 20% em 2010, mantém-se elevada em 2011 e reduz-se abruptamente em 2012, atingindo o mínimo da série (cerca de 12%). A partir de 2013, a trajetória inverte-se, registando-se um crescimento lento, mas persistente ao longo de 2013–2019, aproximando-se de 15% no final da década.

Do ponto de vista agregado, a sobreeducação não é um fenómeno residual: apesar do mínimo atingido em 2012, situou-se sistematicamente em patamares quantitativamente relevantes e com tendência moderadamente crescente, como já referido. Esta evidência justifica a análise desagregada que se segue, com especial atenção à fase de entrada no mercado de trabalho e à concentração em certos setores e áreas de formação.

### 3.2. Evolução por eixos de heterogeneidade

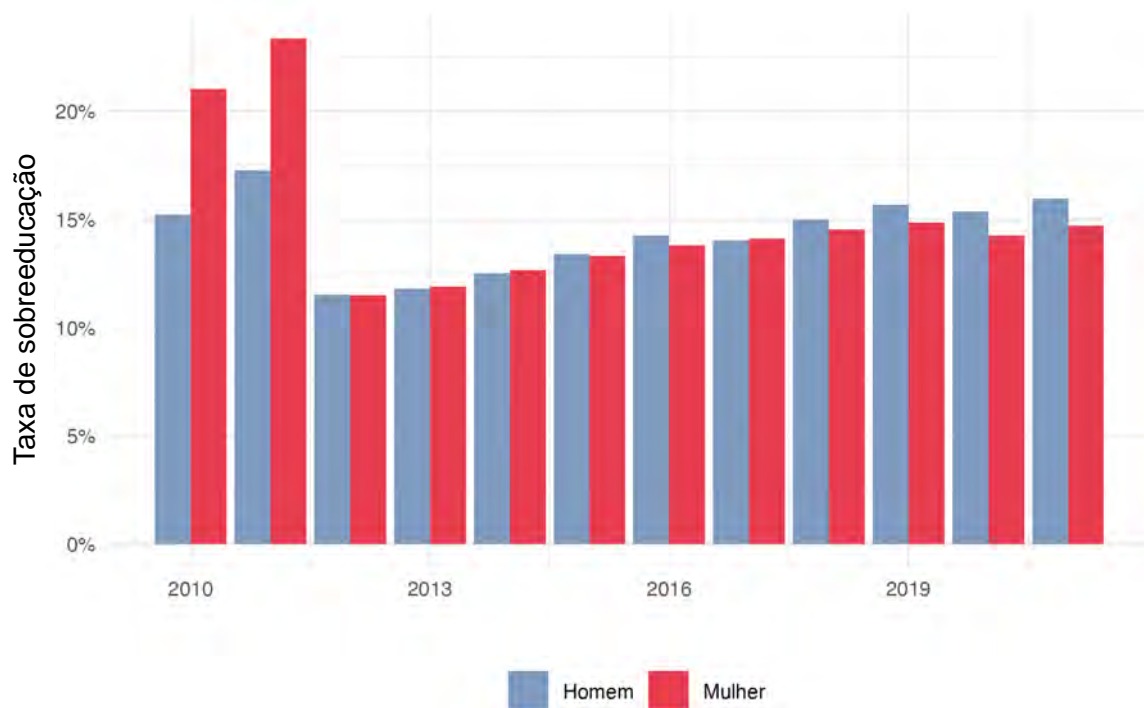
Nesta subsecção, procura-se identificar que tipo de trabalhador estará mais exposto à sobreeducação, analisando perfis de risco e respetiva dinâmica ao longo da carreira.

#### ***Diferenças por sexo***

A Figura 2 desagrega a prevalência de sobreeducação por sexo no período 2010-2021. O resultado central pode ser sintetizado em dois padrões: (i) um elevado diferencial inicial desfavorável às mulheres

e (ii) convergência rápida após a quebra de 2012, com o surgimento de um diferencial marginal desfavorável aos homens nos anos mais recentes.

**Figura 2 – Prevalência de sobreeducação entre licenciados, por sexo, 2010-2021**



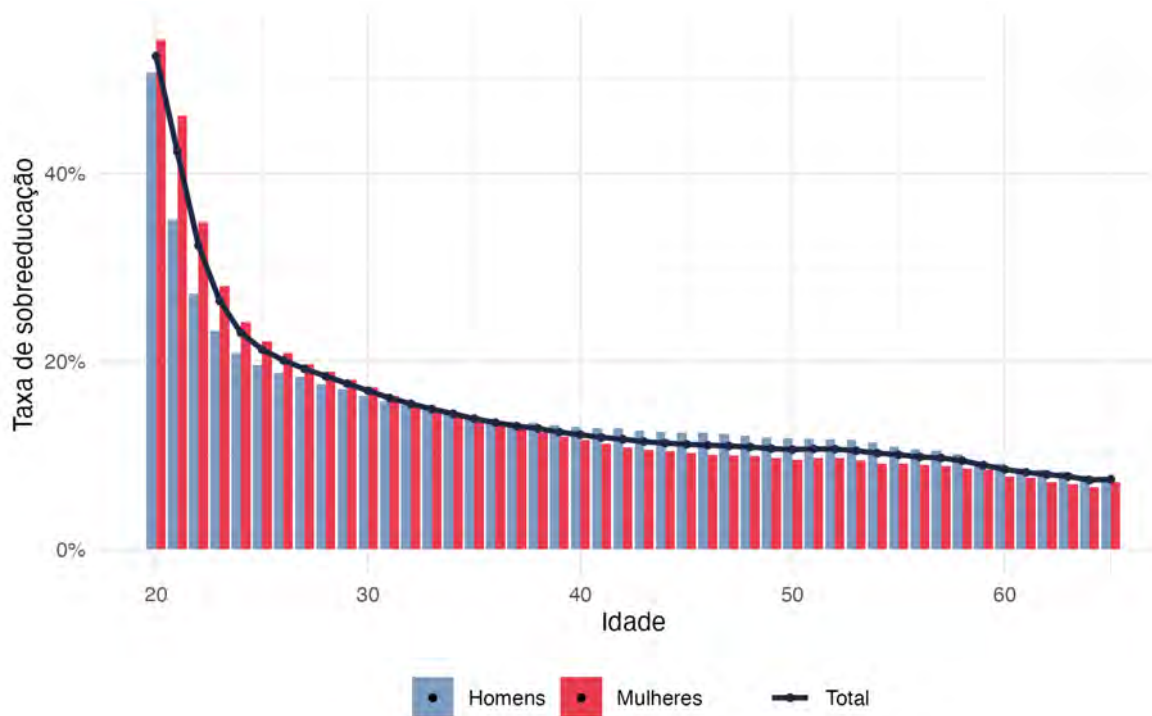
**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

Em 2010–2011, a prevalência de sobreeducação de licenciados era superior entre mulheres (valores acima de 20%) do que entre homens (cerca de 16%). Tal parece sugerir que, durante o período de subida acentuada do desemprego que marcou estes anos, as mulheres foram substancialmente afetadas pelo ajustamento, provavelmente escolhendo enfrentar os custos do desajuste educacional para mitigar os custos do desemprego. Em 2012, ocorreu uma queda pronunciada para ambos os sexos, aproximando-se de 11-12%, eliminando praticamente o diferencial observado nos anos anteriores.

***Diferenças por idade***

A Figura 3 evidencia um gradiente etário pronunciado, sugerindo que a sobreeducação está fortemente associada à fase inicial da carreira, embora o risco desta não desapareça totalmente ao longo do ciclo de vida. Em idades muito jovens (início dos 20 anos), a prevalência é particularmente elevada e decresce rapidamente nos primeiros anos de inserção profissional; a partir da segunda metade da década dos 20, a redução torna-se mais gradual, aproximando-se de um patamar mais baixo e relativamente estável nas idades intermédias e avançadas.

**Figura 3** – Prevalência de sobreeducação entre licenciados, por idade e sexo, média 2010-2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

A desagregação por sexo mostra diferenças tanto no nível como na dinâmica. No início da vida ativa, as mulheres apresentam prevalências substancialmente superiores às dos homens, com um diferencial marcado nos 20–24 anos, revelando dificuldades adicionais na transição educação–mercado de trabalho face aos homens. Com o aumento da idade, observa-se convergência: o diferencial diminui progressivamente e aproxima-se de zero por volta das idades intermédias (aproximadamente 35 anos). Nas idades posteriores, o padrão tende a inverter-se, com prevalência masculina ligeiramente superior.

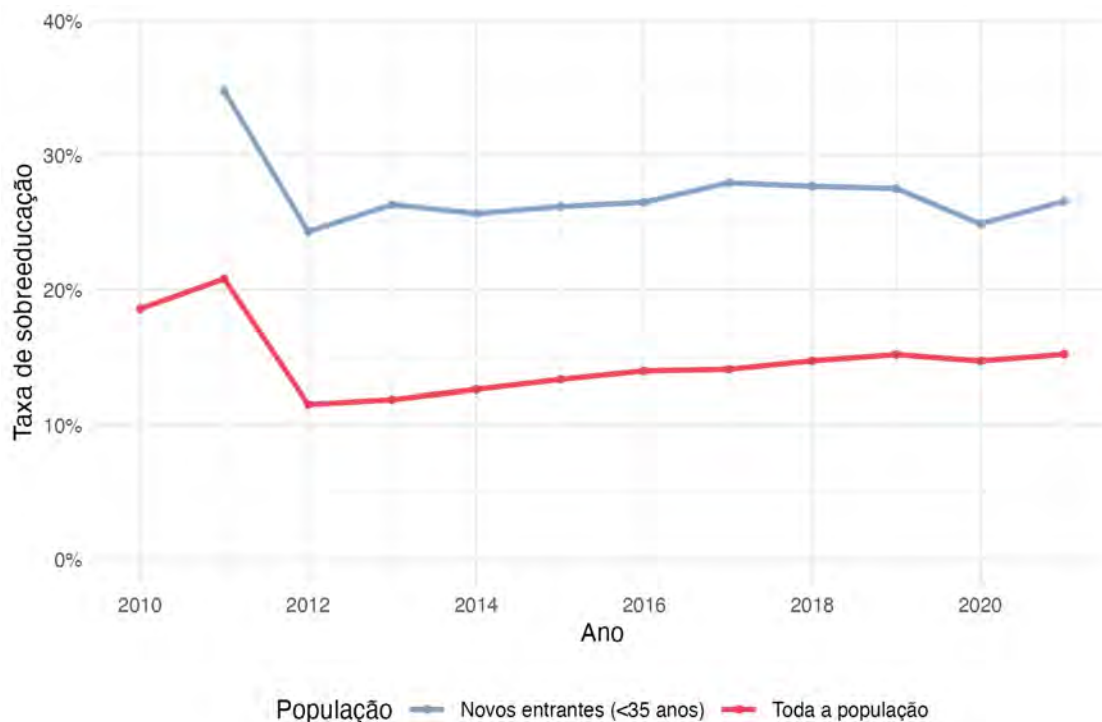
Em suma, as diferenças de género parecem concentrar-se sobretudo na entrada no mercado de trabalho, tornando a transição ensino–mercado de trabalho um ponto crítico para análise das diferenças salariais entre homens e mulheres, merecendo estas atenção específica nas políticas de inserção profissional. Esta análise suporta uma leitura em que a sobreeducação é predominantemente um fenómeno de transição (forte incidência na entrada e rápida correção).

**Diferenças por fase da carreira.**

A Figura 4 reforça as leituras anteriores sobre o período de transição ensino–mercado de trabalho, ao comparar a evolução agregada com a prevalência observada entre *novos entrantes* no mercado de trabalho (definidos como indivíduos com menos de 35 anos). Retiram-se duas notas: (i) os novos entrantes exibem níveis sistematicamente superiores aos do agregado; e (ii) a tendência de

crescimento a partir de 2013 é ligeiramente mais pronunciada na população total que entre novos entrantes.

**Figura 4** – Evolução da prevalência de sobreeducação: total versus novos entrantes (menos de 35 anos), 2010–2021

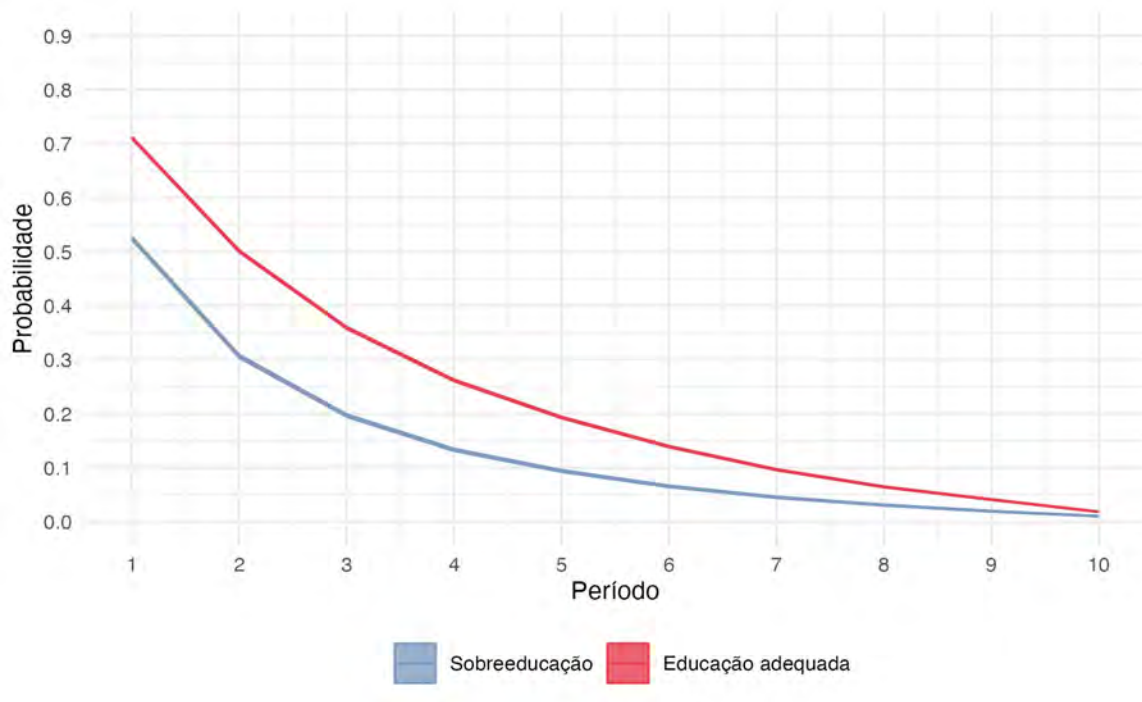


**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

Esta evidência mostra que o ponto de partida – a exposição inicial na entrada no mercado de trabalho – ser muito desvantajoso tem vindo a tornar-se ligeiramente menos desfavorável, sugerindo uma atenuação relativa do ajuste na transição ensino–mercado de trabalho.

A Figura 5 apresenta curvas de sobrevivência de Kaplan–Meier para a duração na mesma ocupação, distinguindo indivíduos em situação de sobreeducação e indivíduos com educação adequada. A curva associada à sobreeducação encontra-se sistematicamente abaixo da curva de educação adequada, implicando uma menor probabilidade de permanência na mesma ocupação quando existe desajuste educacional desta natureza. Ou seja, os trabalhadores sobreeducados exibem maior propensão a transitar para outro emprego. Esta maior taxa de saída é compatível com um processo de ajustamento: perante um emprego abaixo do nível de qualificação, os indivíduos tendem a procurar e/ou aceitar alternativas que potencialmente melhorem o ajuste.

**Figura 5** – Probabilidade de manutenção no mesmo posto de trabalho, para desajuste *versus* ajuste educacional, média 2010–2021

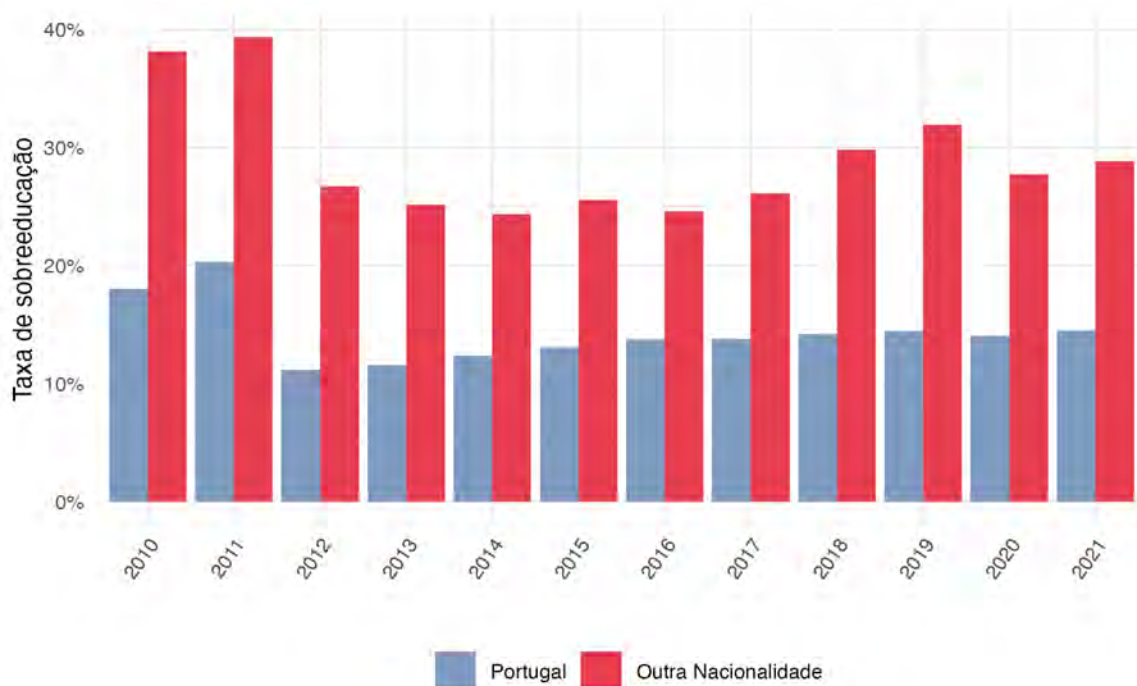


**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

***Diferenças por nacionalidade.***

A Figura 6 evidencia diferenças por nacionalidade ao longo de 2010–2021. Em todos os anos, trabalhadores licenciados com *outra nacionalidade* (estrangeiros) apresentam níveis substancialmente superiores de sobreeducação, configurando um eixo central de desigualdade no acesso a postos compatíveis com o nível de qualificação.

**Figura 6** – Evolução da prevalência de sobreeducação entre licenciados, por nacionalidade, 2010-2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

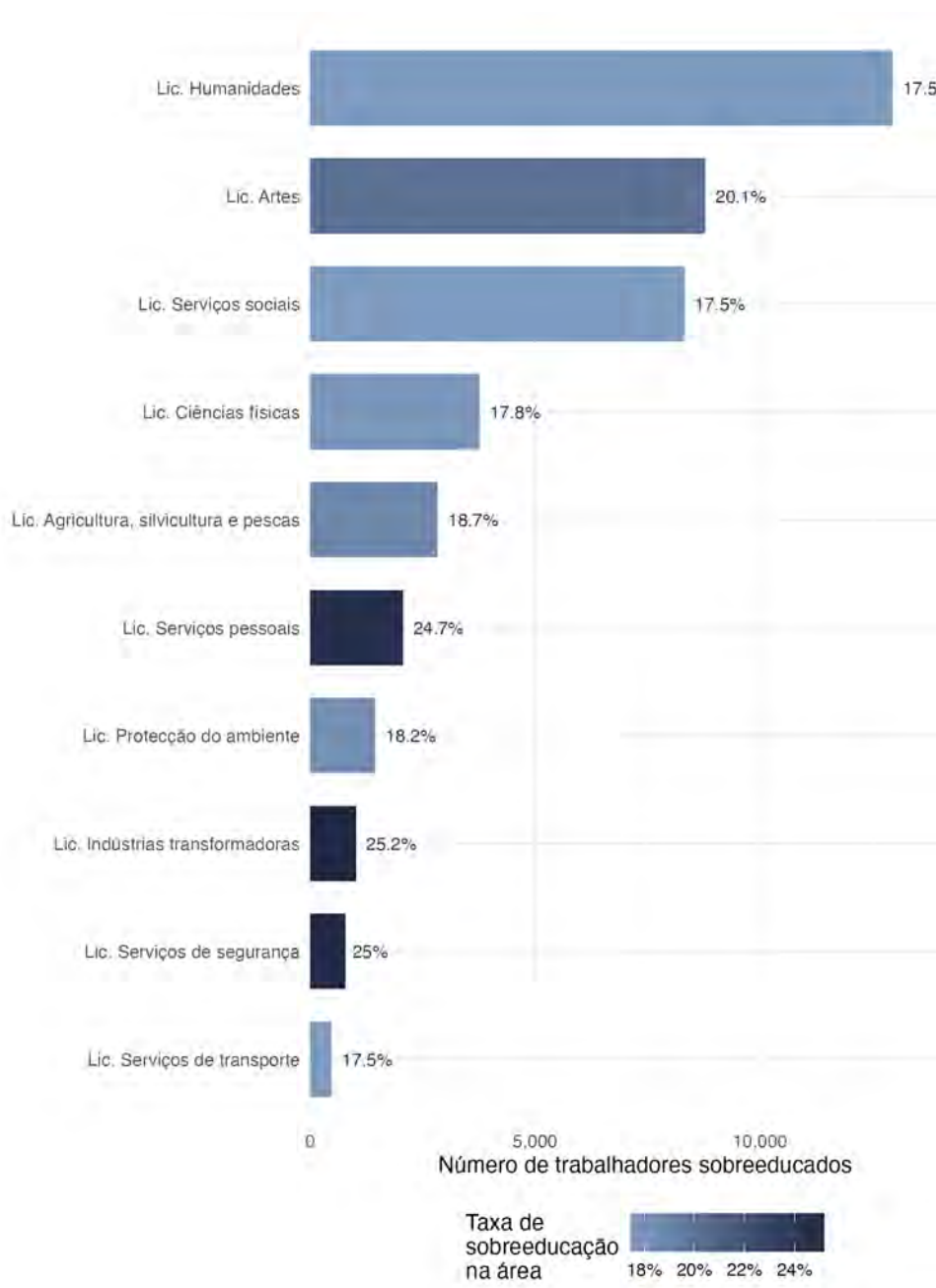
Enquanto a prevalência entre nacionais se situa, em geral, entre 11–15% após 2012, entre estrangeiros encontra-se tipicamente entre 25-32%, atingindo valores próximos de 40% no início do período. Mesmo após a quebra de 2012, a distância mantém-se estrutural ao longo da década, sem se observar qualquer tendência de convergência.

Este padrão é compatível com fricções na integração laboral de trabalhadores estrangeiros altamente qualificados (por exemplo, devido a dificuldades de reconhecimento de credenciais, qualidade das redes de intermediação e segmentação ocupacional). Do ponto de vista substantivo, a conclusão principal é a existência de uma subutilização relevante de capital humano neste grupo, com custos potenciais para indivíduos (nível e progressão de rendimentos, estabilidade laboral e inclusão) e para a economia (ineficiência alocativa).

**Diferenças por área de formação.**

A Figura 7 apresenta a desagregação da sobreeducação por área de formação em 2021, combinando (i) escala (número absoluto de trabalhadores sobreeducados nos dados disponíveis) e (ii) intensidade (taxa de sobreeducação na área de formação).

Figura 7 – Sobreeducação por área de formação: escala (n.º de casos) e prevalência (taxa), 2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. O gráfico representa as 10 áreas de formação com maior número de casos de sobreeducação.

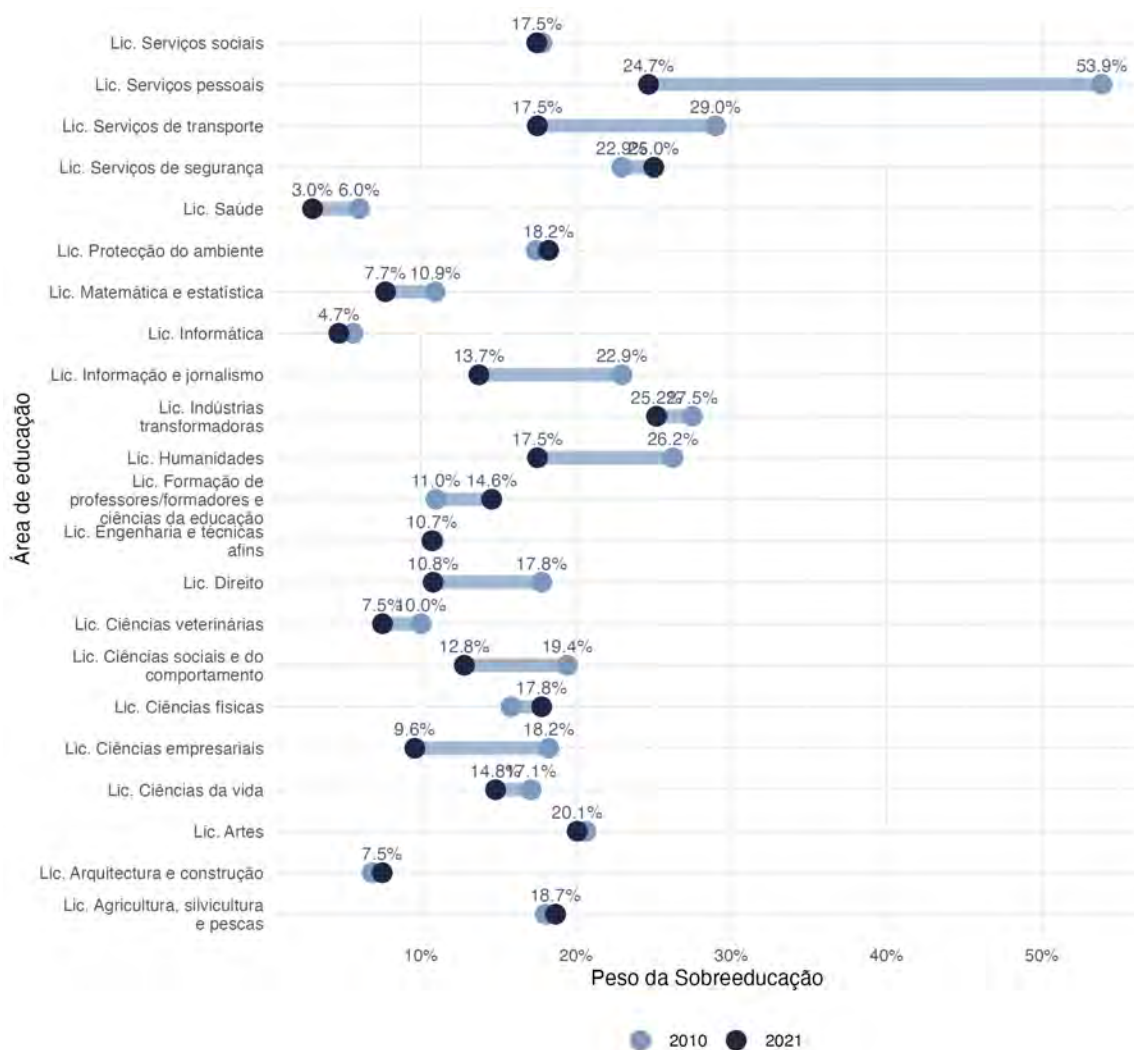
Do ponto de vista dos volumes, sobressaem áreas de formação com maior massa de diplomados e, portanto, maior número absoluto de casos: *Humanidades*, *Artes* e *Serviços sociais*. Em particular, a área de *Humanidades* concentra o maior número absoluto de trabalhadores em situação de sobreeducações, apesar de uma taxa moderada (17,3%).

Em contraste, quando se observa a intensidade, emergem áreas mais vulneráveis: *Serviços de segurança* (26,3%), *Serviços pessoais* (24,4%) e *Indústrias transformadoras* (24,3%). Nestes casos,

os volumes podem ser menores, mas o risco relativo é elevado, sugerindo maior probabilidade de colocação abaixo do nível de qualificação.

A Figura 8 introduz uma perspetiva temporal na intensidade (taxa) de sobreeducação, ao comparar 2010 e 2021. Verifica-se um padrão global de redução, ainda que com magnitudes distintas, entre áreas de formação. As áreas de formação que registaram agravamento são minoritárias e, em geral, com variações diminutas.

**Figura 8 – Evolução da prevalência de sobreeducação por área de formação: 2010 vs. 2021**



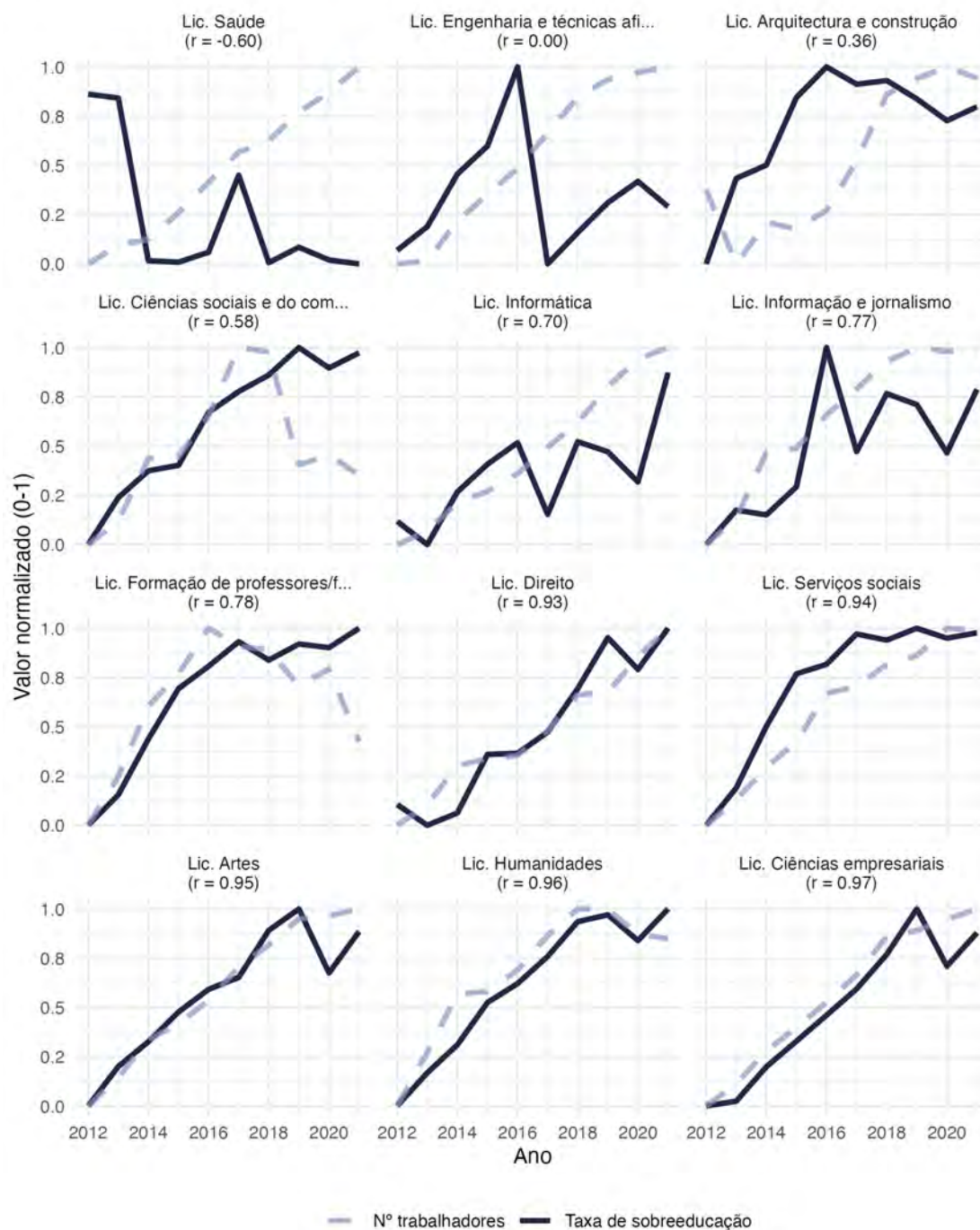
**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

Do lado das dinâmicas mais preocupantes, destaca-se a *Formação de professores/formadores e ciências da educação*, onde a prevalência aumenta de cerca de 11% (2010) para 14,6% (2021), sugerindo deterioração do ajuste neste domínio.

Em contraste, observam-se reduções acentuadas em várias áreas, incluindo *Serviços pessoais* (53,9% para 24,7%), *Serviços de transporte* (29,0% para 17,5%), *Humanidades* (26,2% para 17,5%), *Informação e jornalismo* (22,9% para 13,7%) e *Ciências empresariais* (18,2% para 9,6%). Adicionalmente, áreas de mais forte correspondência ocupacional exibem níveis baixos em 2021 e trajetórias descendentes, como a *Saúde* (6% para 3%).

Praticamente todas as áreas de formação têm visto aumentar o número de licenciados. Este aumento aparenta estar associado a maior risco de desajuste, embora com algumas exceções. A Figura 9 permite conduzir esta análise para as áreas de educação com maior número de licenciados, ainda que apenas a partir da observação de comovimentos contemporâneos simples. As exceções à existência de uma correlação forte entre as dinâmicas das duas séries, número de trabalhadores e taxa de sobreeducação, são a Saúde, a Engenharia e a Arquitetura. Tal parece indicar que, nestas áreas, o mercado ainda não aparenta sinais de saturação. Nas Ciências Sociais, também se verifica uma correlação menos intensa entre as duas séries, mas, neste caso, porque a quebra de número de trabalhadores a partir de 2019 é acompanhado por uma manutenção do crescimento da taxa de sobreeducação (a área de Formação de professores parece estar a seguir o mesmo padrão, mas com a clara discrepância entre as duas séries a revelar-se apenas a partir de 2020)

**Figura 9 –** Evolução do número de licenciados e da prevalência (taxa) de sobreeducação, 2012-2021



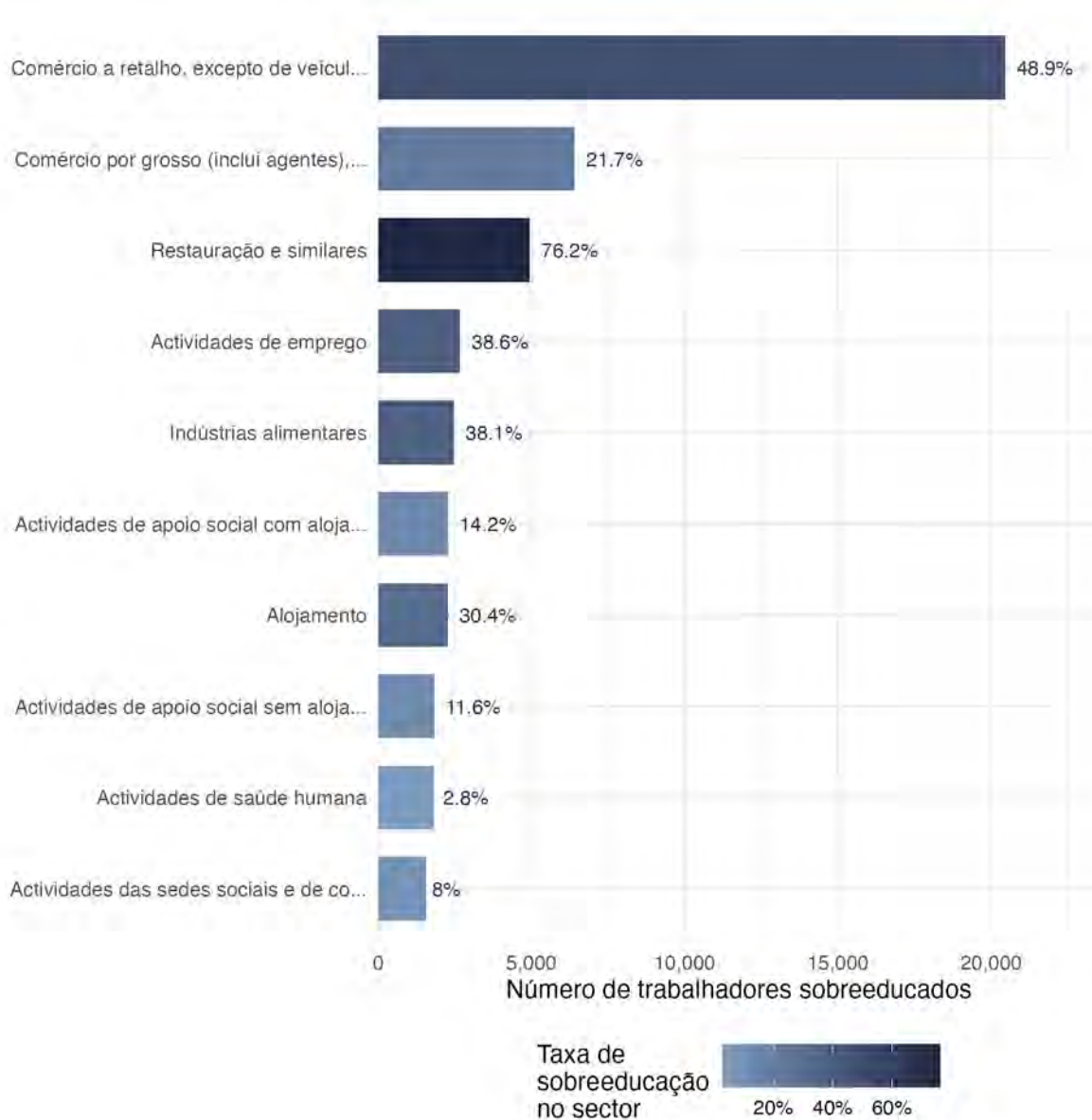
**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. r – Coeficiente de correlação absoluta.

**Diferenças por setor de atividade económica.**

A Figura 10 apresenta a prevalência de sobreeducação por setor, em 2021. A leitura setorial é informativa porque evidencia que a sobreeducação está associada a segmentos produtivos com elevada capacidade de absorção de emprego, mas com estrutura ocupacional frequentemente pouco intensiva em funções associadas a nível de ensino superior, o que pode ser desajustado face à

emergência de um perfil de competências cada vez mais qualificado. Ou seja, importa que essa capacidade de absorção exista em setores que se tornam mais sofisticados.

**Figura 10** – Sobreeducação por setor: escala (n.º de casos) e intensidade (taxa), 2021



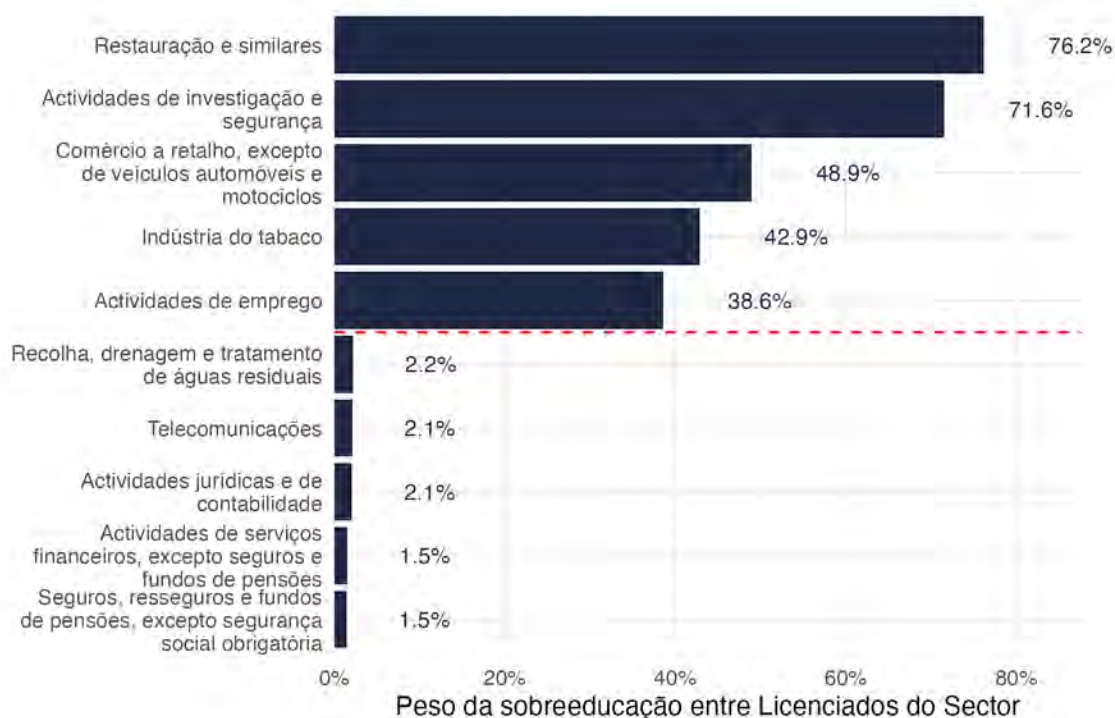
**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. O gráfico representa os 10 setores com maior número de casos de sobreeducação.

Em termos de escala, destaca-se o *Comércio a retalho*, com cerca de 20 mil licenciados sobreeducados e uma taxa de sobreeducação no setor elevada (48,9%). Este resultado é substantivamente relevante por combinar grande volume com risco elevado. O *Comércio por grosso* surge também com volume importante, mas com taxa inferior (21,7%).

Quanto à intensidade, o caso mais extremo é a *Restauração e similares* (76,2%), sugerindo ser um setor onde a presença de licenciados ocorre frequentemente em funções que não requerem ensino superior. Outros setores com taxas elevadas incluem *Atividades de investigação e segurança* (71,6%), *Comércio a retalho* (48,9%) (já referido acima), *Indústria do tabaco* (42,9%) e *Atividades de emprego* (38,6%). Em sentido oposto, setores intensivos em competências específicas, como *Seguros* (1,5%) e *Atividades de serviços financeiros* (1,5%), exibem sobreeducação residual, consistente com maior correspondência credencial/ocupacional.

A Figura 11 ordena os setores segundo a prevalência de sobreeducação entre os licenciados que trabalham em cada setor, destacando os casos de maior e de menor prevalência.

**Figura 11 – Prevalência de sobreeducação entre licenciados do próprio setor, 2021**

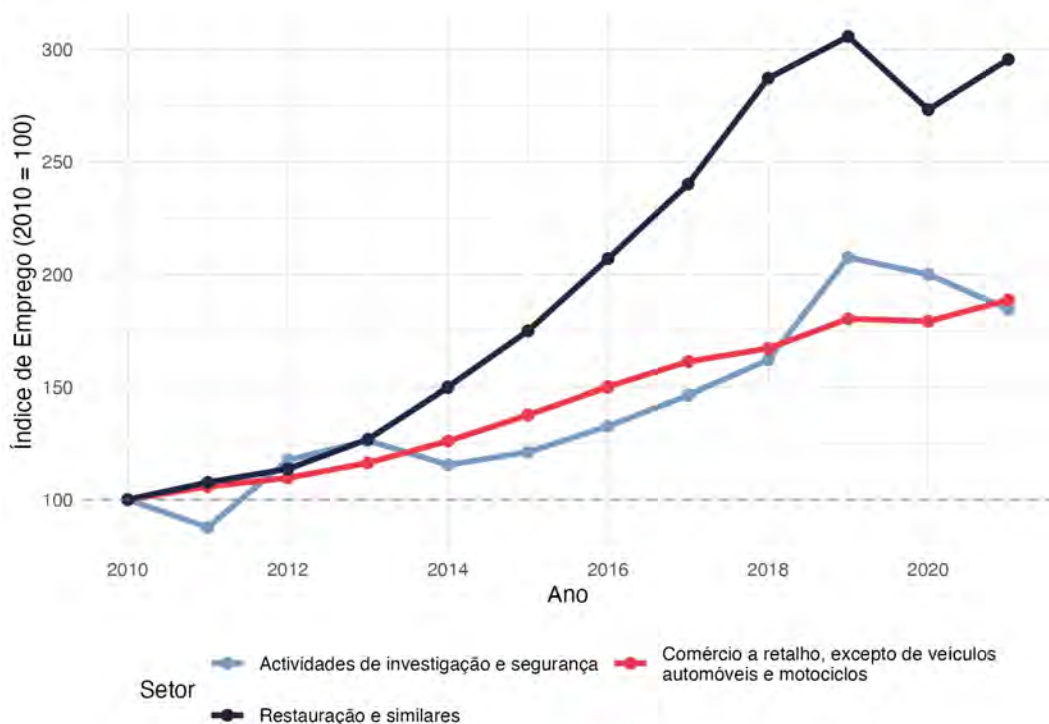


**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. O gráfico retrata os setores de maior e de menor prevalência da sobreeducação.

Em termos interpretativos, esta dispersão sugere uma desadequação do tipo de especialização produtiva face ao tipo de qualificações que estão a redefinir o mercado de trabalho. As atividades relacionadas com o turismo (*Restauração e similares*) e, ainda que em menor grau, o *Comércio a retalho* e as *Atividades de investigação e segurança* sugerem especial preocupação, por serem setores de elevado crescimento no emprego nos anos em causa (ver Figura 12). Este crescimento poderá ter criado um efeito de realocação de recursos que canibalizará, a prazo, o crescimento de outros setores mais compatíveis com as qualificações da população e que possibilitariam alocações mais eficientes. No curto prazo, o crescimento rápido destes setores pode ter-se traduzido em ganhos que levaram a

que, momentaneamente, alguns licenciados produzam maior valor nesses segmentos do que em alternativas disponíveis. Mas esta realocação poderá ter custos de eficiência dinâmica: ao concentrar oferta de trabalho qualificado em setores cuja estrutura ocupacional é, em média, menos intensiva em funções graduadas, limita-se a acumulação de competências e aprendizagem organizacional em setores com maior potencial de produtividade e de crescimento a longo prazo.

**Figura 12** – Evolução do emprego nos setores de atividade com elevada prevalência de sobreeducação, 2010-2021



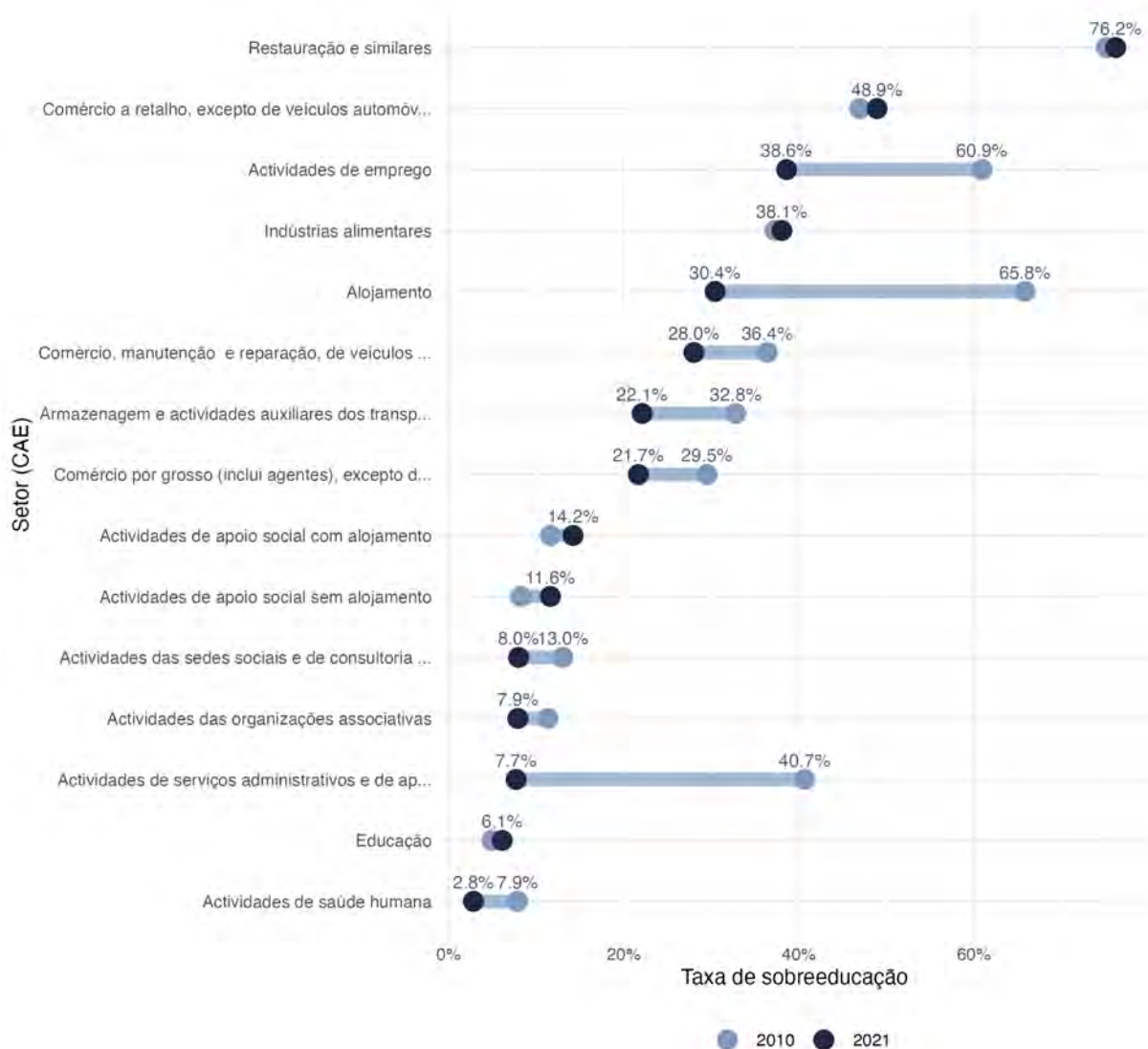
**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

O mecanismo subjacente pode ser interpretado como um efeito de “desvio” de talento. Os salários e condições de contratação praticados em segmentos turísticos podem funcionar como um preço de reserva que dificulta a sua mobilização por setores mais inovadores, mas ainda em fase de consolidação, os quais não conseguem oferecer remunerações competitivas no curto prazo. O resultado é uma fricção adicional: não apenas há desajuste no setor que absorve, como se reduz a probabilidade de que setores potencialmente mais intensivos em conhecimento consigam escalar e atrair massa crítica de capital humano. Neste enquadramento, a sobreeducação em certas atividades não é apenas um indicador de subutilização estática de qualificações; pode também refletir uma especialização produtiva que condiciona a trajetória de desenvolvimento. Ao “fixar” trabalhadores licenciados em setores de crescimento rápido, mas de menor intensidade tecnológica e de conhecimento, a economia sacrifica oportunidades de expansão de setores de elevado valor

acrescentado, onde os licenciados tenderiam a criar e a apropriar-se de maior valor a prazo (através de ganhos de produtividade e de salários).

A Figura 13 compara 2010 com 2021 e mostra que coexistem setores onde a sobreeducação se agravou com outros onde diminuiu substancialmente.

**Figura 13 – Evolução da prevalência de sobreeducação por setor: 2010 vs. 2021**



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. O gráfico retrata os setores de maior prevalência de sobreeducação.

Entre as dinâmicas mais preocupantes, sobressaem, novamente, níveis muito elevados e que se têm agravado (ainda que marginalmente) em *Restauração e similares* e *Comércio a retalho*. Observam-se ainda aumentos moderados em *Actividades de apoio social sem alojamento* e em *Actividades de apoio social com alojamento*, que, embora em patamares absolutos mais baixos, são relevantes por ocorrerem em setores fundamentais socialmente. Nestes, pode estar a dar-se um processo de

qualificação que ainda não atingiu patamares suficientes para que os licenciados neste setor sejam identificados, empiricamente, como bem ajustados.

Em contraste, verificam-se reduções acentuadas em *Alojamento* (65,8% para 30,4%), *Atividades de emprego* (60,9% para 38,6%) e *Serviços administrativos e de apoio prestados a empresas* (40,7% para 7,7%). Em setores como *Atividades de saúde humana*, a descida para níveis muito baixos (8,2% para 2,8%) confirma maior eficiência da alocação entre trabalhadores e ocupações e, possivelmente, alguma escassez de licenciados.

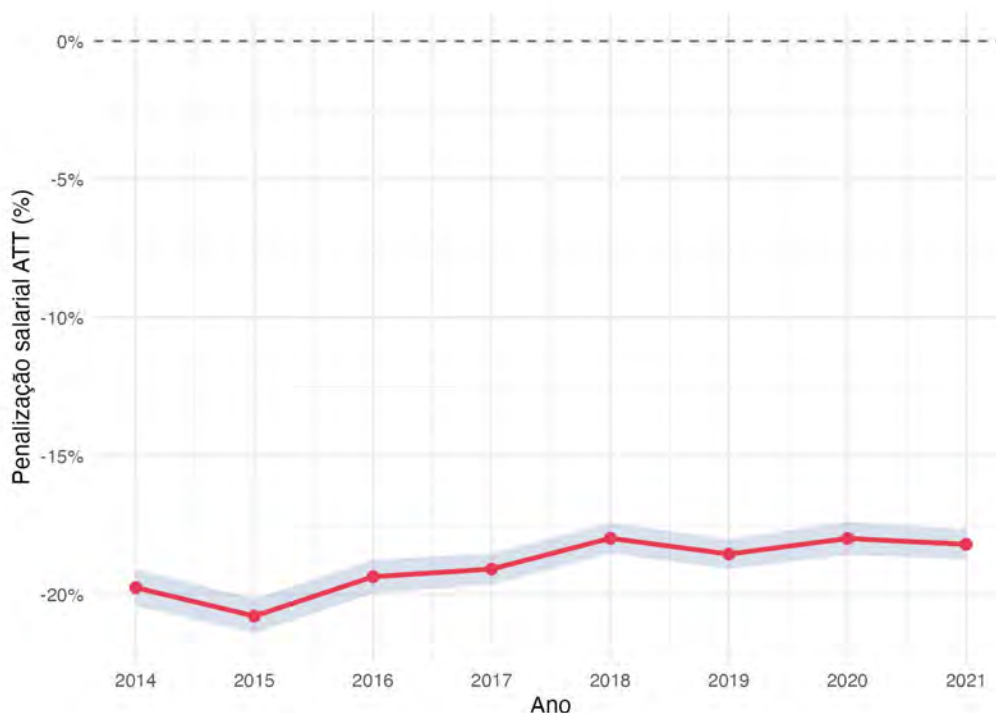
## 4. Efeitos da sobreeducação entre licenciados

Esta secção sintetiza evidência sobre a penalização salarial associada à sobreeducação entre licenciados em Portugal, enfatizando (i) a distribuição dos efeitos em 2021, (ii) a sua evolução ao longo do ciclo de vida do trabalhador, e (iii) padrões de heterogeneidade por nacionalidade, área de formação e setor de atividade. Em termos gerais, os resultados apontam para uma penalização robusta e maioritariamente negativa, mas com heterogeneidade substantiva entre grupos e contextos laborais, permitindo identificar perfis de maior vulnerabilidade.

### 4.1. Efeitos médios da sobreeducação

A Figura 14 apresenta a evolução temporal do efeito médio da sobreeducação nos salários. O padrão é consistente com uma penalização persistente e elevada ao longo de 2014-2021, apesar de alguma tendência de atenuação no período em análise. No final do período, esta penalização era de cerca de 17%, depois de ter atingido valores de cerca de 21% em 2015.

**Figura 14** – Evolução do efeito médio da sobreeducação nos salários, 2014-2021

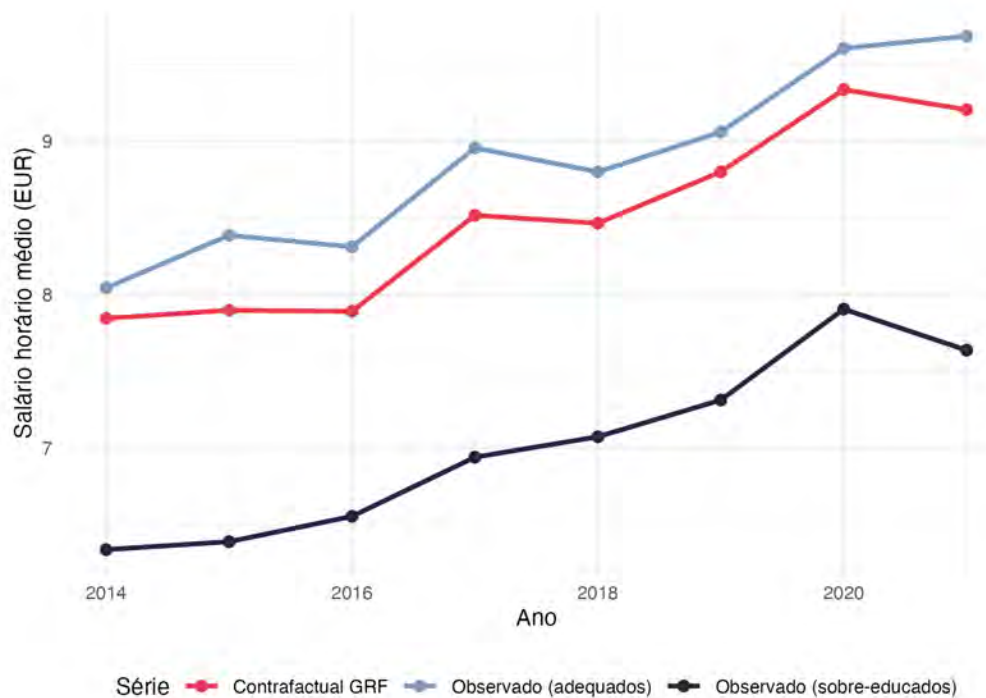


**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Valores negativos indicam penalização salarial.

A Figura 15 compara três séries: (i) salário médio por hora observado dos trabalhadores licenciados em situação de ajuste educacional (trabalhadores adequados), (ii) salário por hora observado dos

sobreeducados e (iii) salário por hora contrafactual para os sobreeducados (remuneração esperada, caso não estivessem sobreeducados, com a mesma idade, sexo e área de formação). A distância entre o observado e o contrafactual para os sobreeducados é o valor médio de penalização associada ao desajuste vertical. As séries revelam que o salário contrafactual permaneceu sistematicamente abaixo do salário dos trabalhadores adequados no período em análise, ainda que com uma tendência de gradual aproximação até 2020, o que indicia uma penalização salarial sistemática associada à sobreeducação, mas em trajetória de redução. Contudo, em 2021, este movimento de redução sofreu uma inversão assinalável.

**Figura 15** – Evolução do salário médio por hora observado e salário por hora contrafactual para trabalhadores sobreeducados (euros por hora), 2014-2021

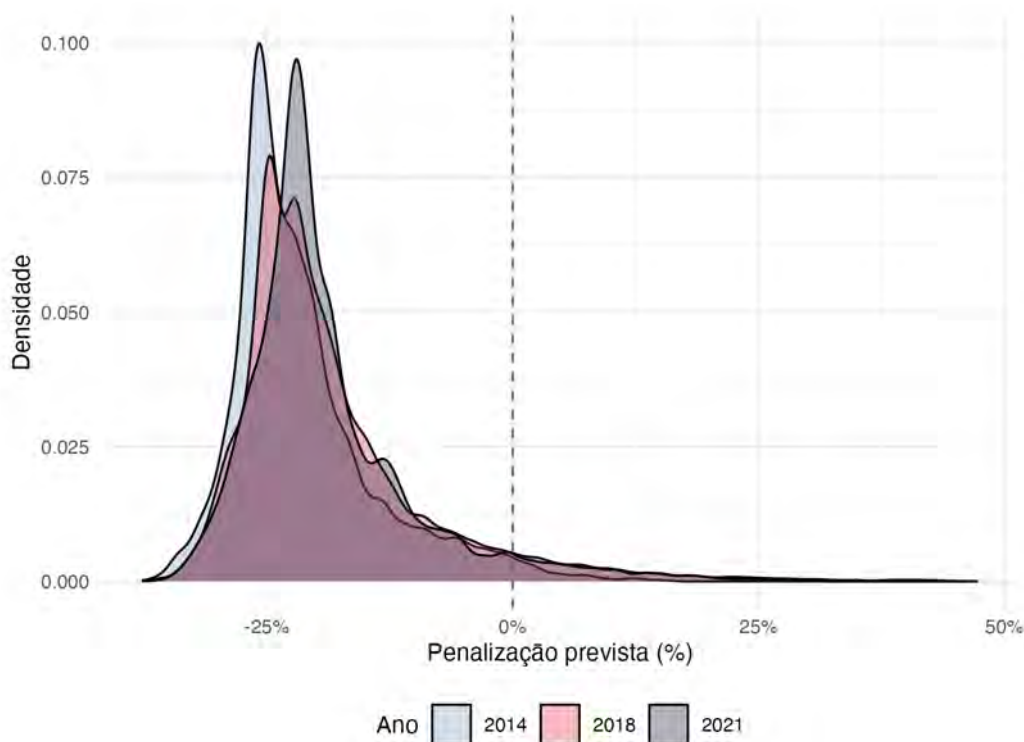


**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Salário contrafactual dos trabalhadores sobreeducados: salário esperado caso os trabalhadores não estivessem sobreeducados, para dada idade, sexo e área de formação.

## 4.2. Heterogeneidade dos efeitos salariais da sobreeducação

A penalização salarial média oculta uma heterogeneidade de efeitos considerável. A Figura 16 mostra a distribuição (densidade) dos efeitos individuais estimados para três anos selecionados. A massa da distribuição concentra-se em valores negativos, mas observa-se uma cauda direita que se aproxima de zero e pode ultrapassá-lo (i.e., indicando a ocorrência de prémio salarial associado ao desajuste educacional) para uma pequena fração de indivíduos.

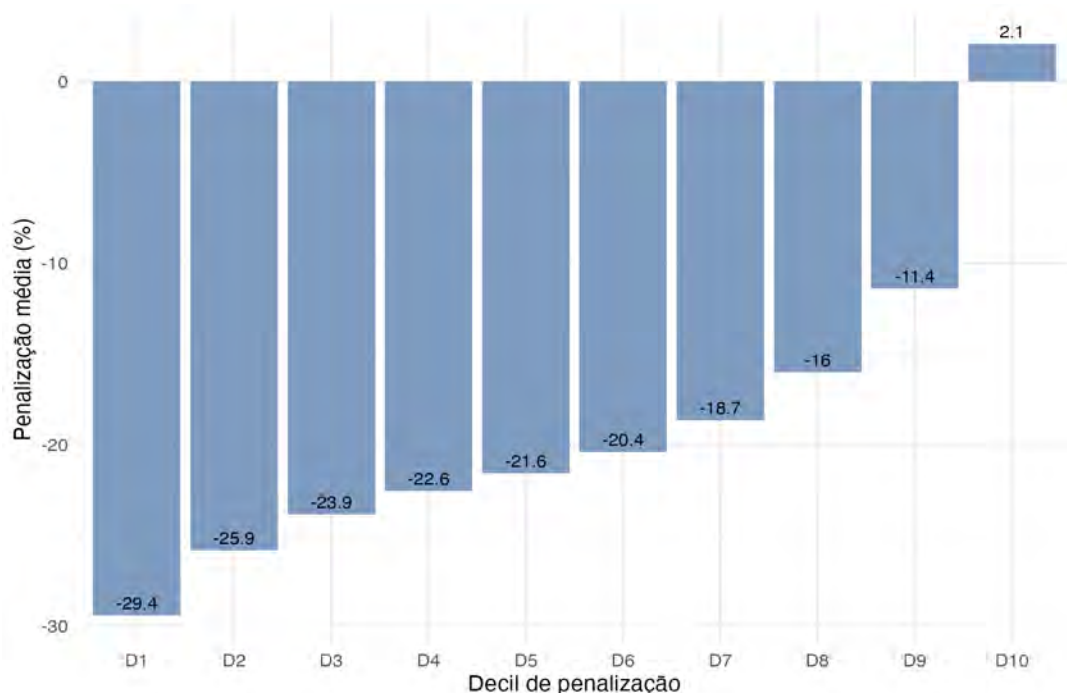
**Figura 16** – Distribuição (densidade) dos efeitos individuais da sobreeducação nos salários, 2014, 2018, 2021



Nota: Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Valores negativos indicam penalização salarial.

Na Figura 17, organizam-se os efeitos da sobreeducação nos salários por decis da distribuição dos efeitos para 2021. Os decis inferiores exibem penalizações (em média do decil) muito elevadas, com valores entre cerca de -30% e -20%, enquanto o decil superior pode aproximar-se de zero ou tornar-se ligeiramente positivo, como já referido. Em conjunto, os resultados sugerem que a maioria dos licenciados sobreeducados enfrenta perdas salariais substanciais, ainda que coexistindo com um pequeno subconjunto com penalização muito reduzida ou até com um ligeiro prémio salarial.

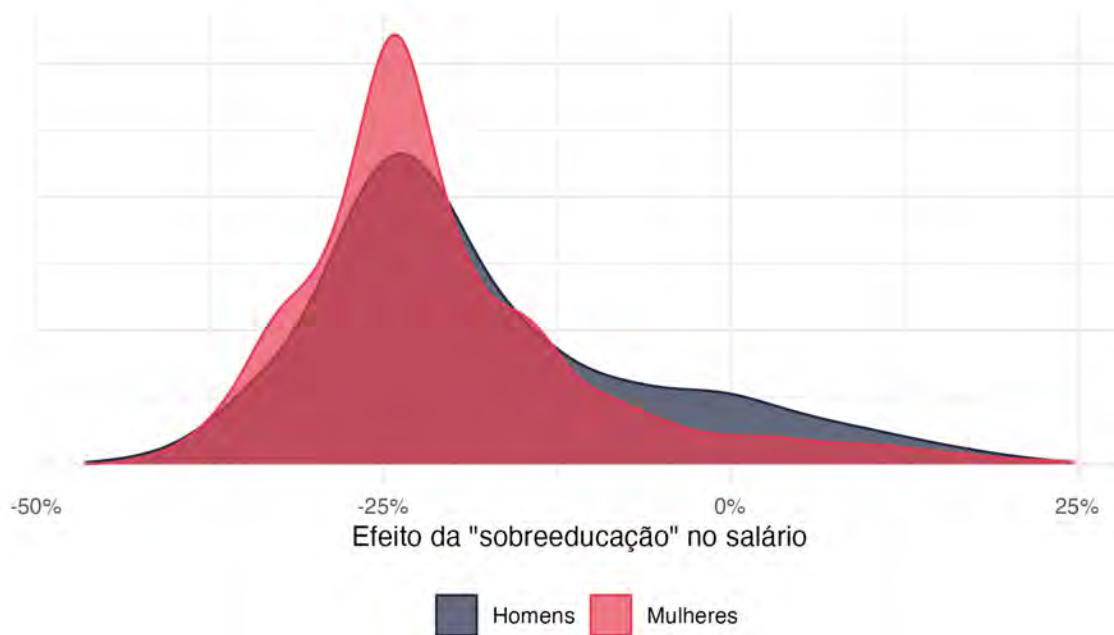
**Figura 17 – Efeito da sobreeducação nos salários por decis (média do decil), 2021**



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Valores negativos indicam penalização salarial.

A Figura 18 apresenta as distribuições (densidades) do efeito estimado da sobreeducação no salário em 2021, por género, entre licenciados sobreeducados. Observa-se que a penalização é predominantemente negativa para ambos os grupos, com maior concentração de massa nos valores entre cerca de -35% e -15% e uma moda próxima de -25%. Ainda assim, a distribuição das mulheres é mais concentrada em torno desta penalização típica (moda), sugerindo menor dispersão dos efeitos, enquanto a dos homens é mais espalhada e com cauda direita mais pronunciada, indicando maior heterogeneidade e maior probabilidade de penalizações menos severas (valores próximos de zero) e mais casos de efeitos positivos.

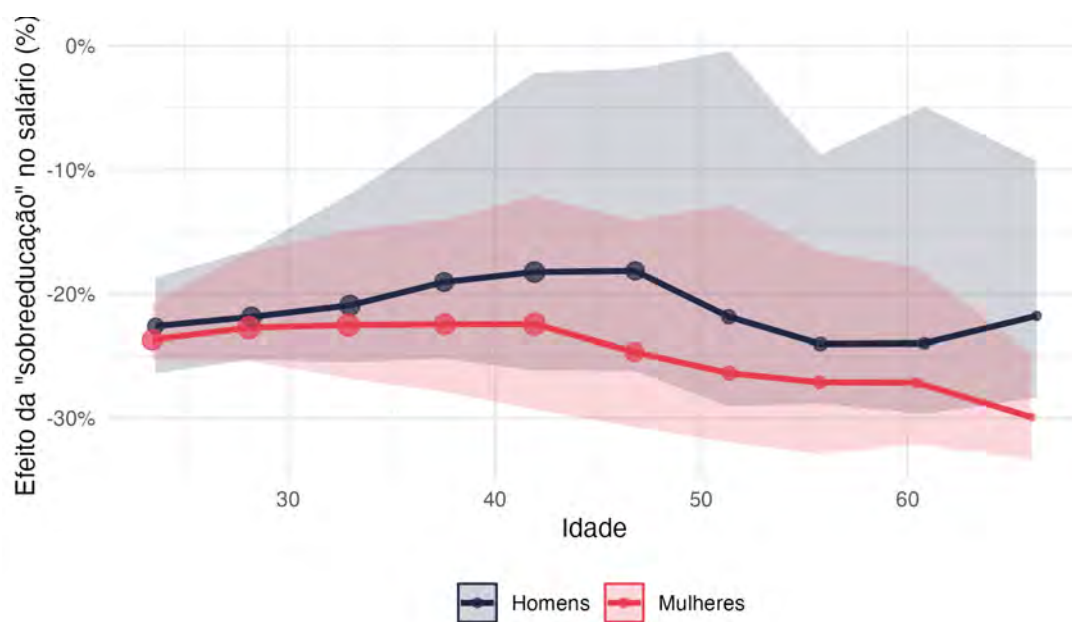
Figura 18 – Distribuição (densidade) dos efeitos individuais da sobreeducação nos salários, por sexo, 2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Valores negativos indicam penalização salarial.

A evolução da penalização ao longo da vida profissional sugere que o desajuste não é estático, podendo intensificar-se em fases mais avançadas do ciclo de vida e diferir entre grupos. A Figura 19 mostra que, em todas as idades observadas, o efeito médio é negativo (i.e., existe penalização salarial) para homens e mulheres. Até meados dos 30 anos de idade, as trajetórias são relativamente próximas e tendem ambas a diminuir. A partir dos 40 anos, observa-se uma divergência: para as mulheres, a penalização torna-se progressivamente mais severa, aproximando-se de 30% nas idades mais avançadas; para os homens, a penalização agrava-se mais tarde no ciclo de vida (continua a atenuar-se até perto dos 50 anos) e de forma mais moderada, mantendo-se tipicamente entre cerca de 20% e 25% – e mostrando até alguma atenuação – nas idades mais avançadas. Os intervalos de confiança alargam-se nas idades mais elevadas, sugerindo maior incerteza e heterogeneidade nos grupos etários mais velhos.

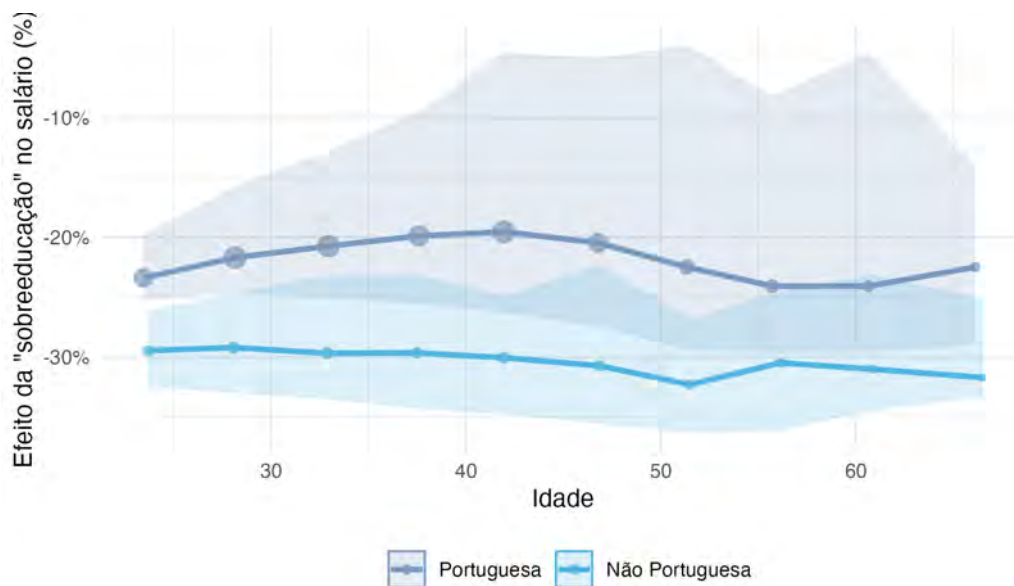
Figura 19 – Evolução do efeito médio da sobreeducação nos salários ao longo da vida profissional, por sexo



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. As linhas representam a penalização média estimada para cada faixa etária, com intervalos de confiança sombreados. Valores negativos indicam penalização salarial.

A Figura 20 complementa esta leitura ao distinguir trabalhadores portugueses e estrangeiros. Em todas as idades, as estimativas permanecem negativas para ambos os grupos. Contudo, verifica-se um diferencial sistemático por nacionalidade: para trabalhadores estrangeiros, a penalização é mais elevada e relativamente estável, situando-se tipicamente em torno de 30% com agravamento ligeiro a partir dos 45 anos de idade. Para trabalhadores portugueses, a penalização é menos severa, oscilando aproximadamente entre 20% e 25%, com alguma deterioração em torno dos 45 anos e recuperação parcial nas idades mais avançadas. Tal como no caso do sexo, os intervalos de confiança alargam-se nas idades mais elevadas, sugerindo maior incerteza/heterogeneidade de efeitos nos dados aquando do fim do ciclo de vida profissional.

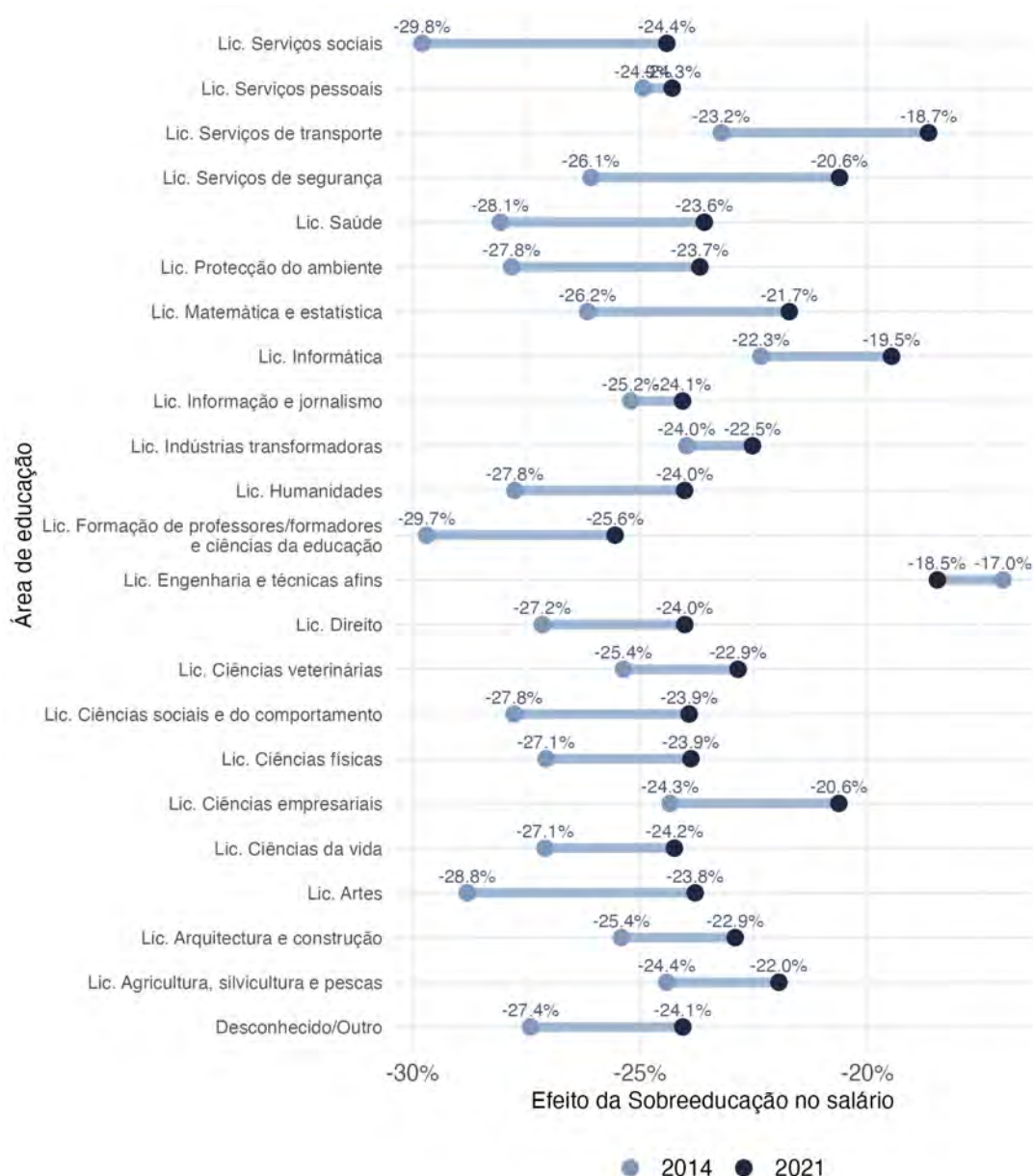
**Figura 20** – Evolução do efeito médio da sobreeducação nos salários ao longo da vida profissional, por nacionalidade



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Valores negativos indicam penalização salarial.

A Figura 21 compara o efeito salarial mediano da sobreeducação por área de formação, entre 2014 e 2021. Em todas as áreas de formação, a mediana do efeito permanece negativa, confirmando penalizações salariais generalizadas. Ainda assim, observa-se uma melhoria transversal: na maioria das áreas de formação, a penalização tornou-se menos severa em 2021 (valores menos negativos), sugerindo redução da penalização mediana ao longo do período; apenas a área da *Engenharia e técnicas afins* registou um agravamento da penalização no período em análise. Mas heterogeneidade entre áreas manteve-se: em 2021, as penalizações medianas menos acentuadas concentravam-se em *Engenharia e técnicas afins*, *Serviços de transporte* e *Informática*, com valores entre -19,5% e -18,5%, enquanto as áreas de *Formação de professores/ciências da educação*, *Serviços sociais* e *Serviços Pessoais* apresentavam as perdas medianas mais elevadas, com valores entre -25,6% e -24,3%. Entre 2014 e 2021, a *Formação de professores/ciências da educação* e *Serviços sociais* mantiveram-se como os mais penalizados.

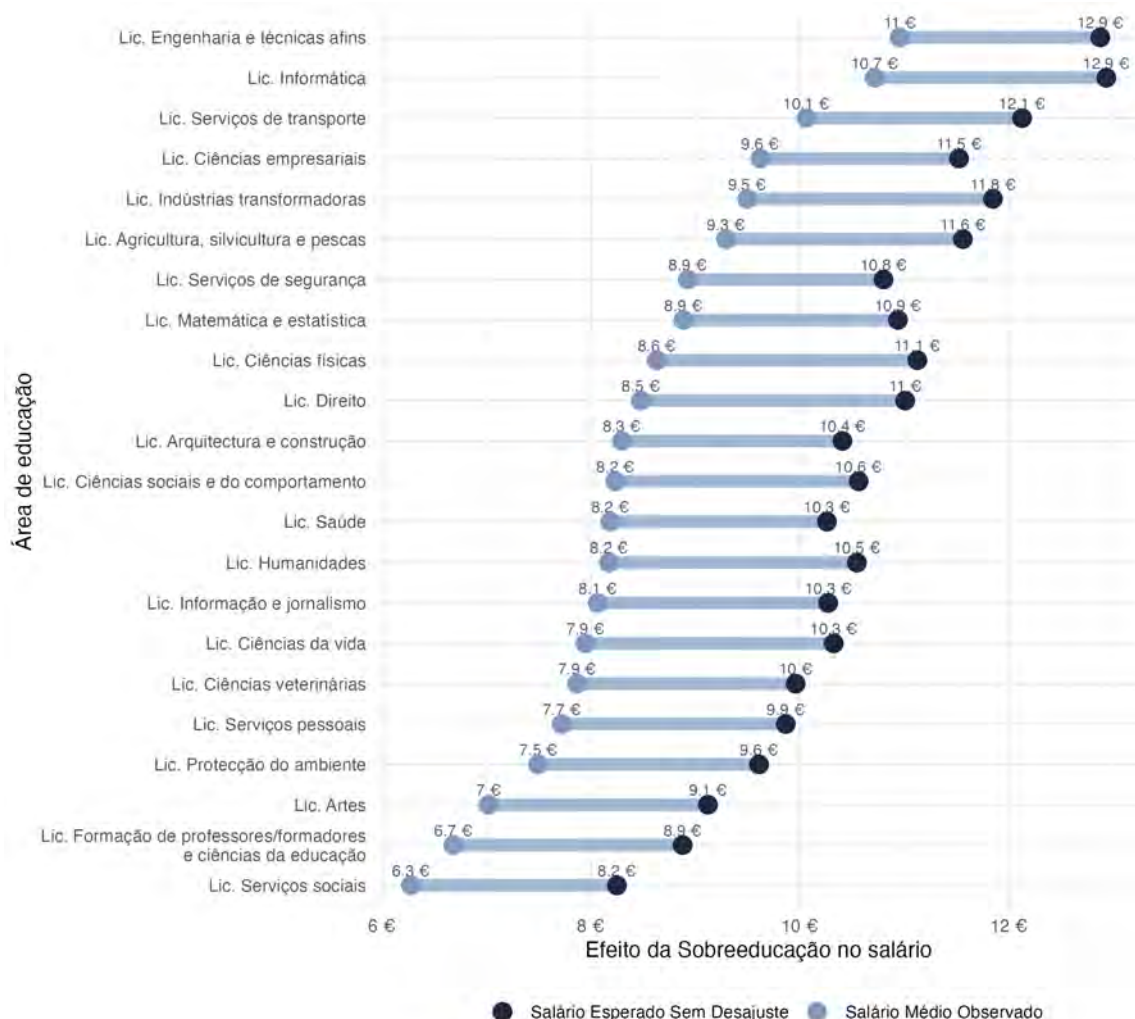
Figura 21– Evolução do efeito mediano da sobreeducação no salário por área de formação: 2014 vs. 2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Valores negativos indicam penalização salarial.

Para tornar a análise da penalização mais tangível em termos de níveis salariais, a Figura 22 contrapõe, por área de formação, o salário médio por hora observado em 2021 e o salário por hora contrafactual, ou seja, o salário esperado no cenário sem desajuste educacional. Em todas as áreas, o salário esperado sem desajuste educacional é superior ao valor médio observado em pelo menos 1,9 euros, evidenciando que a penalização se traduz numa perda sistemática de rendimento dentro de cada área de formação.

**Figura 22 – Salário médio por hora observado versus salário por hora contrafactual (salário esperado sem desajuste educacional), por área de formação, 2021**



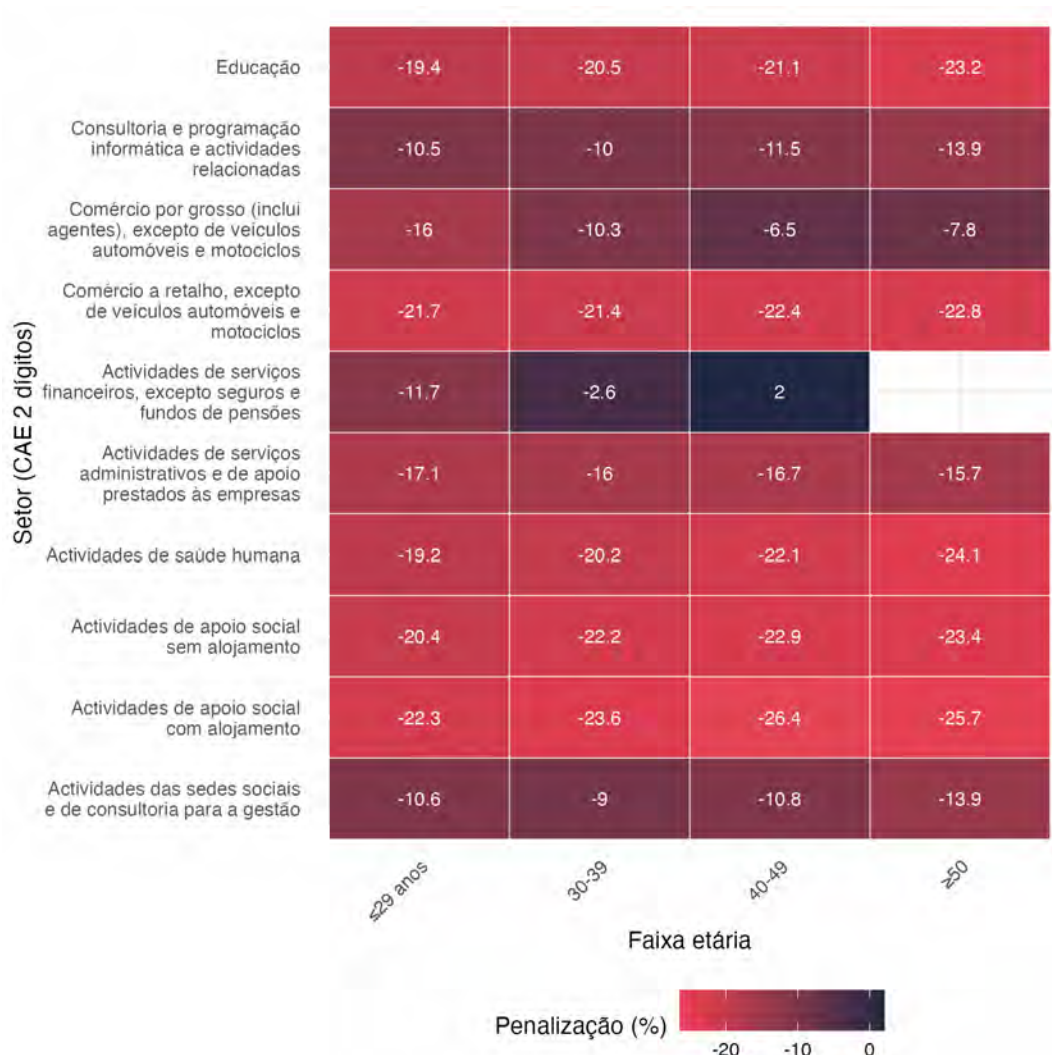
**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal.

Para além de médias globais, a política pública beneficia de evidência capaz de localizar onde a penalização salarial associada à sobreeducação é mais intensa. As matrizes de interação, nas Figuras 23 e 24, permitem identificar combinações de características com risco acrescido, isto é, perfis em que o ciclo de vida e o contexto produtivo (setor de atividade) ou educacional (área de formação) se conjugam para agravar perdas salariais.

A Figura 23 reporta a penalização média por combinação de faixa etária e setor de atividade (CAE a 2 dígitos), considerando os setores mais relevantes. Observam-se penalizações especialmente severas nas *Atividades de apoio social*, sobretudo *com alojamento*, onde a penalização atinge valores muito elevados (valor de -26,4% na faixa 40-49 e -25,7% após os 50 anos), e também nas *Atividades de saúde humana*, com agravamento até cerca de 24% no grupo etário mais velho. É também muito penalizada a sobreeducação, com agravamento ao longo da idade, nos setores do *Comércio a retalho* e da *Educação*. Em contraste, setores como *Consultoria e programação informática* e *Atividades das*

*sedes sociais e consultoria para a gestão* apresentam penalizações mais moderadas (valores aproximadamente entre -9% e -14%) e menor variação com a idade. Destaca-se ainda um padrão atípico nas *Atividades de serviços financeiros*, onde a penalização diminui com a idade e se aproxima de zero, chegando a efeitos salariais com valores ligeiramente positivos na faixa 40-49 anos, sugerindo uma interação específica entre progressão na carreira e desajuste educacional nesse setor. O setor do *Comércio por grosso* é o outro onde a penalização mais se reduz com a idade. No conjunto, a matriz evidencia que a penalização depende simultaneamente do estágio do ciclo de vida do trabalhador e do contexto setorial, com maior gravidade concentrada em setores de cuidados/apoio social e, em vários casos, agravamento nas idades mais avançadas.

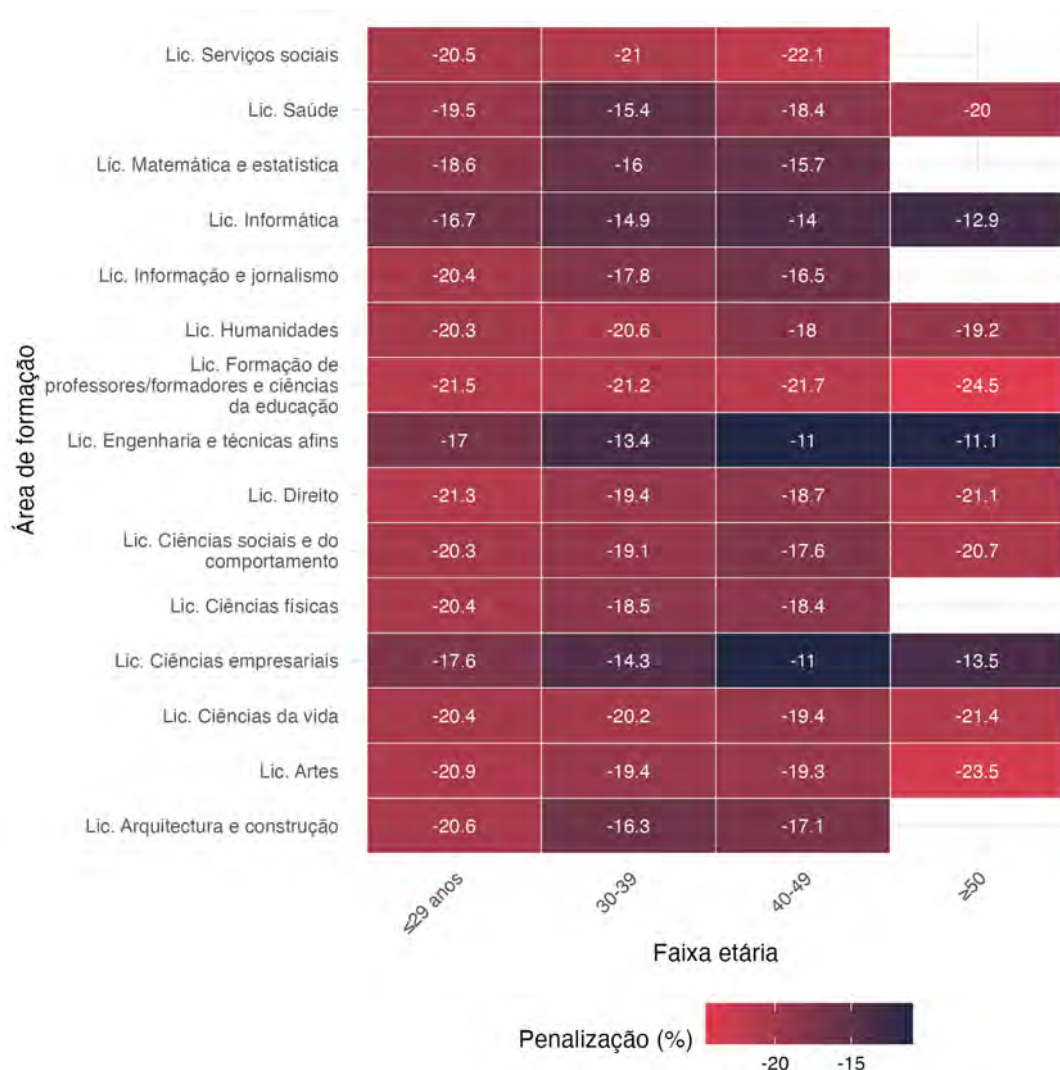
**Figura 23** – Efeito médio da sobreeducação nos salários (%):  
matriz de interação idade *versus* setor de atividade



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Efeito salarial médio da sobreeducação por combinação de faixa etária e setor de atividade. As células mais claras identificam combinações de elevada vulnerabilidade, onde trabalhadores enfrentam penalizações particularmente severas.

A Figura 24 apresenta a matriz de interação entre idade e área de formação (mais relevantes por número de trabalhadores), permitindo identificar perfis de dupla vulnerabilidade associados ao campo académico de formação e ao ciclo de vida. As penalizações são maioritariamente negativas, mas com diferenças relevantes: a *Engenharia e técnicas afins* e a *Informática e Ciências empresariais* exibem penalizações comparativamente mais moderadas, sobretudo a partir dos 40 anos (valores de -11% e de -14% a -13%, respetivamente). No polo oposto, a *Formação de professores/ciências da educação*, as *Artes*, os *Serviços sociais*, as *Ciências da vida* e o *Direito* registam penalizações elevadas e com agravamento nas idades mais avançadas (atingindo entre -21,1% a -24,5% no grupo acima dos 50 anos de idade).

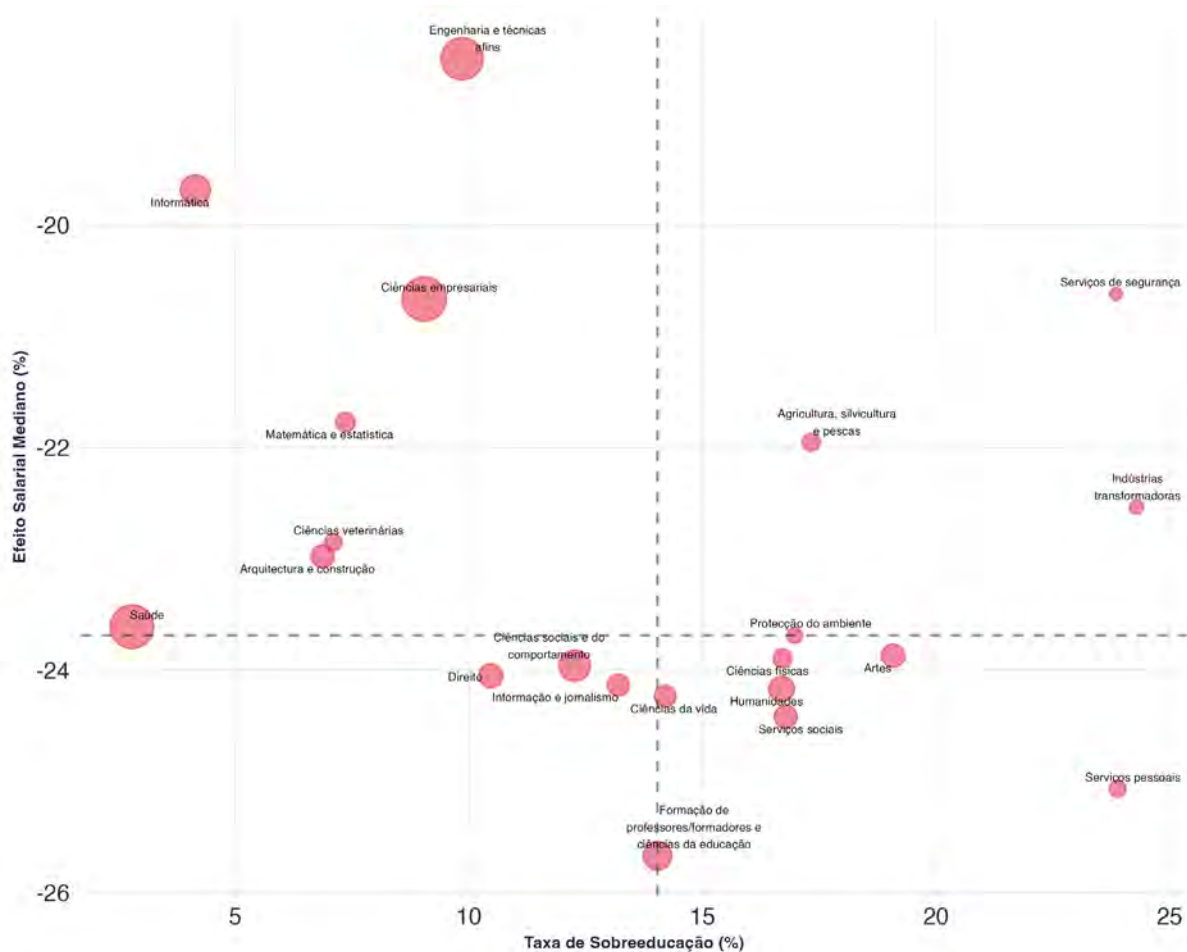
**Figura 24** – Efeito médio da sobreeducação nos salários (%):  
matriz de interação idade *versus* área de formação



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. Efeito salarial médio da sobreeducação por combinação de faixa etária e área de formação. As células mais claras identificam combinações de elevada vulnerabilidade, onde trabalhadores enfrentam penalizações particularmente severas.

A Figura 25 relaciona, por área de formação, a prevalência da sobreeducação (eixo horizontal) com o efeito salarial mediano (eixo vertical). Regista-se uma relação tendencialmente negativa entre prevalência e efeito salarial (i.e., uma relação positiva entre prevalência e penalização salarial), sobretudo se analisarmos esta relação ponderando o peso de cada área de formação no emprego total. Em termos gerais, observa-se que as áreas STEM (por exemplo, *Informática* e *Engenharia e técnicas afins*) combinam baixa sobreeducação (prevalência entre cerca de 4 e 10%) com penalizações salariais menos severas em termos relativos (valores acima de -20%). Por sua vez, um conjunto de áreas com prevalência de sobreeducação acima da média (aproximadamente 16–20%) concentra penalizações próximas ou abaixo de 24%, nomeadamente *Humanidades*, *Serviços sociais*, *Ciências físicas* e *Artes*. Destaca-se ainda um caso particularmente desfavorável em *Formação de professores/formadores e ciências da educação*, com penalização muito elevada (quase -26%) para uma taxa de sobreeducação próxima da média. Por fim, entre os domínios com prevalência de sobreeducação muito elevada (24–25%), observa-se heterogeneidade: aos *Serviços pessoais* associa-se uma penalização forte (perto de 25%), enquanto *Serviços de segurança* e *Indústrias transformadoras* apresentam penalizações mais moderadas (valores entre -21% e -22,5%).

Figura 25 – Prevalência e efeito salarial mediano da sobreeducação por área de formação, 2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. O diâmetro de cada circunferência representa o peso da respetiva área de formação no emprego total. As linhas a tracejado representam o valor médio do eixo correspondente.

A Figura 26 estabelece a mesma relação entre prevalência da sobreeducação e efeito salarial mediano, mas agora por setor de atividade económica. Observa-se uma clara relação negativa entre prevalência e efeito salarial (i.e., prevalências mais elevadas estão associadas a penalizações salariais mais gravosas).

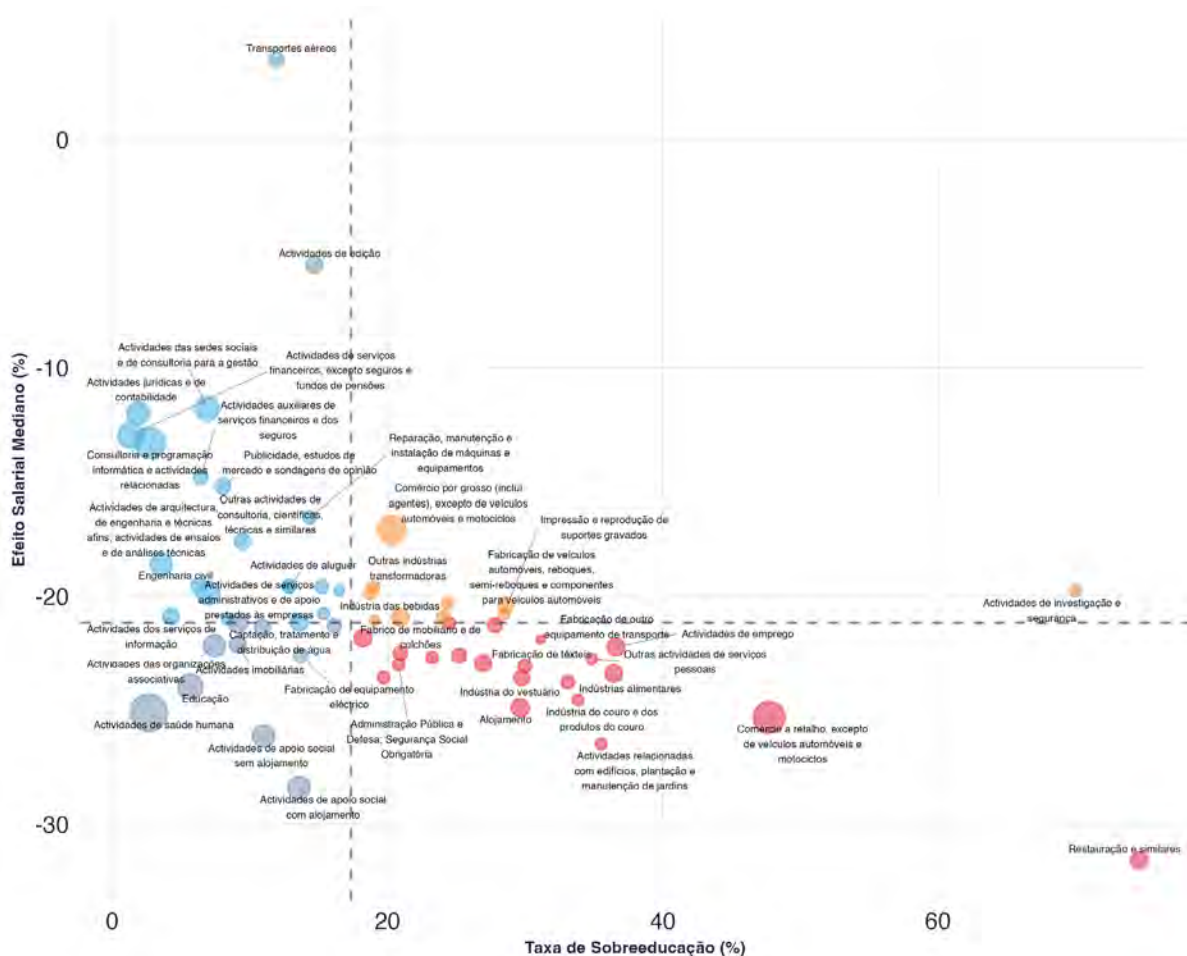
Em termos de grandes padrões, destacam-se três blocos. Um deles reúne setores com baixa prevalência de sobreeducação e penalizações mais moderadas (com valores tipicamente entre cerca de -10% e -18%), onde surgem atividades mais qualificadas, como *Consultoria e programação informática*, *Atividades de serviços financeiros*, *Atividades das sedes sociais e consultoria para a gestão*, *Atividades jurídicas e de contabilidade*, *Publicidade e estudos de mercado*, entre outras.

Outro bloco diz respeito a setores com penalizações muito severas, ainda que com taxas de sobreeducação não particularmente altas, próximas ou até abaixo do valor de referência. Destacam-se aqui as *Atividades de apoio social (com e sem alojamento)*, *Atividades de saúde humana e Educação*, onde os efeitos salariais se aproximam frequentemente de valores entre -25% e -30%.

O terceiro bloco reúne setores com incidência de sobreeducação muito elevada, associados a perdas salariais também elevadas em vários casos. O exemplo mais extremo é a *Restauração e similares*, que combina uma taxa de sobreeducação muito alta (próxima do topo da distribuição) com a penalização mais severa (de cerca de 32%). Também o Comércio a retalho e alguns serviços pessoais e indústrias surgem com incidências elevadas e penalizações acentuadas, embora nunca tão extremas como na *Restauração*.

Por fim, há casos atípicos com efeitos pouco negativos ou mesmo positivos. Destaca-se o setor dos *Transportes aéreos*, que aparece com valor um pouco acima de zero no eixo vertical, sugerindo ausência de penalização salarial mediana.

Figura 26 – Prevalência e efeito salarial mediano da sobreeducação por setor de atividade económica, 2021



**Nota:** Cálculos dos autores a partir das tabelas dos Quadros de Pessoal. O diâmetro de cada circunferência representa o peso do respetivo setor de atividade no emprego total. As linhas a tracejado representam o valor médio do eixo correspondente.

## 5. Síntese e notas finais

Esta Nota de Análise procurou caracterizar a prevalência e os perfis de risco da sobreeducação entre trabalhadores licenciados no setor privado e com contrato individual de trabalho no setor público em Portugal, no período 2010-2021, assim como reunir evidência sobre a penalização salarial associada ao desajuste educacional, considerando heterogeneidade por sexo, idade, nacionalidade, área de formação e setor de atividade.

Os resultados apresentados devem ser interpretados como associações ajustadas com base em microdados administrativos e num indicador calibrado e não como inferência causal estrita. Ainda assim, a combinação entre (i) prevalência, (ii) efeitos salariais e (iii) concentração por grupos de trabalhadores fornece informação útil para possível orientação de prioridades de política pública e desenho de instrumentos.

A evidência empírica da Secção 3 aponta para o seguinte conjunto de conclusões sobre a prevalência, dinâmica e heterogeneidade da sobreeducação:

1. **A sobreeducação tem uma prevalência quantitativamente relevante.** Após atingir um mínimo em 2012, a prevalência evidenciou uma tendência moderadamente crescente durante o resto da década, posicionando-se em patamares médios relativamente elevados no conjunto do período em análise, o que sinaliza uma subutilização persistente de qualificações no mercado de trabalho.
2. **O fenómeno é fortemente concentrado na transição ensino–mercado de trabalho, com correção rápida nos primeiros anos de carreira.** A evidência transversal por idade e, sobretudo, o acompanhamento por coortes revelam incidências elevadas no início da vida ativa e uma redução pronunciada nos primeiros anos, compatível com acumulação de experiência e mobilidade ocupacional.
3. **A exposição inicial dos jovens licenciados é sistematicamente superior à dos trabalhadores mais velhos ao longo do período em análise.** Os novos entrantes exibem níveis persistentemente mais elevados do que o agregado dos trabalhadores licenciados, ainda que com uma dinâmica de crescimento um pouco menos pronunciada que os restantes trabalhadores no período em análise, sugerindo uma atenuação do problema de ajustamento educacional dos trabalhadores às ocupações na entrada no mercado de trabalho.
4. **A sobreeducação associa-se a uma menor permanência do trabalhador no mesmo posto de trabalho.** As curvas de sobrevivência indicam maior probabilidade de transição laboral no que toca os trabalhadores sobreeducados, o que é consistente com a tendência para correção do desajuste educacional via mobilidade entre postos de trabalho, sendo evidência de um mercado de trabalho com alguma flexibilidade.
5. **A heterogeneidade é substantiva entre grupos e segmentos do mercado e revela assimetrias estruturais.** Trabalhadores estrangeiros apresentam prevalências muito superiores às dos nacionais; as mulheres tendem a registar uma transição ensino-mercado de

trabalho com mais atritos que os homens; coexistem áreas de formação e setores de atividade com risco de sobreeducação elevados e áreas e setores com risco baixo; em particular, por setor, observam-se bolsas críticas de absorção de licenciados em sobreeducação (sobretudo em comércio e atividades associadas ao turismo).

Na sua globalidade, estes resultados indicam que a sobreeducação combina (i) uma componente transitória, concentrada na entrada no mercado de trabalho e frequentemente (parcialmente) corrigida por mobilidade e progressão, com (ii) uma componente estrutural, refletida na concentração persistente em certos grupos e em segmentos produtivos específicos.

A evidência empírica da Secção 4 permite o seguinte conjunto de conclusões sobre os efeitos salariais e a sua heterogeneidade associados à sobreeducação:

1. **A sobreeducação está associada a penalizações salariais elevadas e persistentes.** Em 2021, os efeitos individuais típicos (mediana) da sobreeducação nos salários situavam-se em valores em torno de -25%, apontando o desajuste educacional vertical como uma determinante quantitativamente relevante de perdas remuneratórias.
2. **A penalização varia ao longo do ciclo de vida profissional e acumula desvantagens em segmentos específicos.** Os perfis etários indicam que a penalização não é constante com a idade do trabalhador, sendo especialmente desfavorável para as mulheres em idades mais avançadas.
3. **Existe um diferencial sistemático por nacionalidade, com penalizações mais severas para trabalhadores estrangeiros.** Os resultados apontam para penalizações mais elevadas nos trabalhadores estrangeiros, com efeito médio da sobreeducação no salário em torno de -30%, compatível com a existência de barreiras estruturais de integração.
4. **A heterogeneidade por área de formação e setor de atividade é substantiva e permite identificar perfis de elevada vulnerabilidade.** Observam-se penalizações salariais particularmente severas em setores como as *Atividades de apoio social* e *Atividades de saúde humana* e em áreas de formação como *Formação de professores/ciências da educação* e *Serviços sociais*, contrastando com contextos relativamente mais favoráveis como as áreas de formação da *Engenharia*, dos *Serviços de transporte* e da *Informática* e setores como a *Consultoria e programação informática* e as *Atividades das sedes sociais e consultoria para a gestão*, onde as penalizações tendem a ser mais moderadas.
5. **A incidência (taxa) de sobreeducação e a penalização salarial não se alinham de forma simples por área de formação.** Uma elevada incidência tende a estar associada a uma penalização mais severa, mas com significativa dispersão entre áreas de formação, tal que, por exemplo, penalizações muito severas podem ocorrer mesmo quando a incidência é apenas moderada. Isto aponta para a presença de mecanismos adicionais (segmentação, estruturas salariais, regimes de progressão e características das tarefas nos postos de trabalho) para além da mera frequência do desajuste educacional.

6. **As áreas de formação técnico-científicas apresentam, em geral, melhor desempenho relativo, enquanto áreas ligadas a educação, humanidades e serviços sociais exibem maior vulnerabilidade.** Os primeiros domínios tendem a combinar menor exposição à sobreeducação com penalizações menos negativas, sugerindo maior capacidade de absorção de competências e maior transferibilidade para ocupações valorizadas. Nos segundos domínios, as penalizações são frequentemente elevadas e, em vários casos, coexistem com taxas de sobreeducação acima da média, o que sugere maior risco de subutilização persistente de qualificações.
7. **Os setores de atividade associados a serviços de baixa remuneração exibem perdas particularmente acentuadas.** Atividades como restauração, comércio a retalho e certos serviços pessoais tendem a combinar incidência elevada com penalizações severas, refletindo contextos em que a presença de licenciados em funções abaixo do nível de qualificação é mais estrutural. Mas **alguns segmentos de cuidados e serviços coletivos evidenciam penalizações muito severas mesmo sem incidência extrema.** Isto é compatível com estruturas salariais comprimidas e com limitações institucionais/organizacionais que reduzem a capacidade de transformar qualificações em remuneração, tornando o desajuste mais oneroso. Porém, também **existem exceções, com penalizações reduzidas ou praticamente nulas.** Certos setores específicos apresentam efeitos pouco negativos (ou até próximos de zero), sugerindo a ocorrência de elevada seleção de trabalhadores, prémios de qualificação independentes da adequação estrita ao posto de trabalho ou percursos de carreira que recompensam credenciais mesmo em situações de desajuste educacional.

No seu conjunto, estes resultados revelam que as penalizações salariais da sobreeducação combinam (i) uma componente distributiva, que penaliza de forma desigual os trabalhadores ao longo do ciclo de vida e entre grupos, com (ii) uma componente estrutural e segmentada. Este padrão sustenta a necessidade de respostas de política pública mais finamente direcionadas, articulando instrumentos de transição ensino–mercado de trabalho, mitigação de desigualdades persistentes e intervenção setorial em segmentos críticos de absorção de licenciados em sobreeducação.

## Referências

- Allen, J., Levels, M., & Velden, R. van der. (2013). *Skill mismatch and skill use in developed countries: evidence from the PIAAC study* (Research Memorandum No. 2013/17). Research Centre for Education; the Labour Market (ROA). <https://doi.org/10.26481/umaror.2013017>
- Chevalier, A. (2003). Measuring over-education. *Economica*, 70(279), 509–531. <https://doi.org/10.1111/1468-0335.t01-1-00296>
- Flisi, S., Goglio, V., Meroni, E. C., Rodrigues, M., & Vera-Toscano, E. (2017). Measuring occupational mismatch: overeducation and overskill in Europe—evidence from PIAAC. *Social Indicators Research*, 131, 1211–1249. <https://doi.org/10.1007/s11205-016-1292-7>
- Frei, C., & Sousa-Poza, A. (2012). Overqualification: permanent or transitory? *Applied Economics*, 44(14), 1837–1847. <https://doi.org/10.1080/00036846.2011.554380>
- Green, F., & Zhu, Y. (2010). Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. *Oxford Economic Papers*, 62(4), 740–763. <https://doi.org/10.1093/oep/gpq002>
- Kiker, B. F., Santos, M. C., & Mendes de Oliveira, M. M. (1997). Overeducation and undereducation: Evidence for Portugal. *Economics of Education Review*, 16(2), 111–125. [https://doi.org/10.1016/S0272-7757\(96\)00040-4](https://doi.org/10.1016/S0272-7757(96)00040-4)
- McGoldrick, K., & Robst, J. (1996). Gender differences in overeducation: A test of the theory of differential overqualification. *American Economic Review*, 86(2), 280–284.
- McGowan, M. A., & Andrews, D. (2017). *Skills mismatch, productivity and policies: Evidence from the second wave of PIAAC* (Economics Department Working Paper No. 1403). Organisation for Economic Co-operation; Development (OECD). <https://doi.org/10.1787/65dab7c6-en>
- Mendes de Oliveira, M. M., Santos, M. C., & Kiker, B. F. (2000). The role of human capital and technological change in overeducation. *Economics of Education Review*, 19(2), 199–206. [https://doi.org/10.1016/S0272-7757\(99\)00020-5](https://doi.org/10.1016/S0272-7757(99)00020-5)
- Quinn, M. A., & Rubb, S. (2006). Mexico's labor market: The importance of education-occupation matching on wages and productivity in developing countries. *Economics of Education Review*, 25(2), 147–156. [https://doi.org/10.1016/S0272-7757\(05\)00036-1](https://doi.org/10.1016/S0272-7757(05)00036-1)
- Quintini, G. (2011). *Over-qualified or under-skilled: A review of existing literature* (Social, Employment and Migration Working Paper No. 121). Organisation for Economic Co-operation; Development (OECD). <https://doi.org/10.1787/5kg58j9d7b6d-en>
- Rumberger, R. W. (1987). The impact of surplus schooling on productivity and earnings. *Journal of Human Resources*, 22(1), 24–50. <https://doi.org/10.2307/145865>
- Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings: some additional findings. *Journal of Human Resources*, 24(4), 629–643. <https://doi.org/10.2307/145998>

Verhaest, D., & Omey, E. (2006). The impact of overeducation and its measurement. *Social Indicators Research*, 77(3), 419–448. <https://doi.org/10.1007/s11205-005-4276-6>

# Anexo 1 – Apêndice metodológico sobre detecção de ajuste educacional e inferência de efeitos salariais heterogêneos da sobreeducação

## Introdução

Este apêndice metodológico descreve o procedimento usado para calibrar uma métrica de desajuste educacional baseada em detetores de *outliers*, ancorada numa referência normativa inspirada na prática do EUROSTAT para trabalhadores com ensino superior. O texto deve ser lido como complemento técnico à Nota de Análise principal: clarifica escolhas, apresenta o racional de cada decisão e fornece um guião operacional para reprodução do processo com dados administrativos, em particular os Quadros de Pessoal.

A motivação de base é conhecida na literatura e na prática estatística: métricas normativas fixas são transparentes e comparáveis, mas podem tornar-se progressivamente desajustadas quando as ocupações se transformam; métricas puramente empíricas são adaptativas, mas são muito sensíveis a escolhas paramétricas, não existindo critérios adequados para a sua escolha. O procedimento aqui descrito procura conciliar estes objetivos, utilizando uma âncora normativa para calibrar os hiperparâmetros do detetor e escolher a família de detetores. Para estabilidade numérica e para manter o poder estatístico de cada detetor, os parâmetros de cada um destes é apurado numa amostra de *bootstrap* de tamanho fixo. É aplicada uma estratégia de validação cruzada (*Nested CV*) para garantir que a avaliação do detetor empírico não é enviesada e permitir estabelecer a comparação de vários detetores e hiperparâmetros com a âncora normativa.

O apêndice está organizado em quatro secções principais:

- **Secção 1:** Enquadramento conceptual e metodológico (o que é a abordagem, porquê, quando usar, vantagens e limitações).
- **Secção 2:** Abordagem metodológica (definições, dados, detetores, função-objetivo e validação cruzada aninhada).
- **Secção 3:** Resultados da aplicação aos Quadros de Pessoal (resultados utilizados na nota de análise).

## 1. Enquadramento Conceptual

### Definição e âmbito

A metodologia apresentada consiste num procedimento de **seleção e calibração** de detetores empíricos de desajuste educacional, tratando o desajuste como um problema de *outlier detection* em *rolling window* em distribuições condicionais por contexto (células ocupacionais ou pares ocupação – setor). A componente empírica é **ancorada** numa referência normativa compatível com a prática do

EUROSTAT para trabalhadores com ensino superior (ISCED 5–8 correspondendo a ISCO 1 a 3) e é calibrado com **validação cruzada aninhada** (Nested CV), garantindo separação entre: (i) afinação de hiperparâmetros por método, (ii) seleção de método, e (iii) avaliação final em teste.

#### Principais características:

- **Contextualização:** o padrão de referência é definido dentro de células (por exemplo, ocupação–setor), reduzindo heterogeneidade não observada.
- **Flexibilidade temporal:** a definição empírica do “esperado” pode evoluir ao longo do tempo, sendo definida em janelas, mitigando rigidez de correspondências normativas fixas.
- **Base em dados:** a âncora normativa que serve de base à avaliação em Nested CV e a separação temporal permite deixar claros e reproduzíveis os critérios de seleção do método utilizando métricas finais não enviesadas.

O âmbito deste apêndice é a medição do desajuste educacional com especial enfoque nos trabalhadores com licenciatura), por duas razões: (i) é nesse domínio que existe uma referência normativa externa amplamente usada (EUROSTAT), e (ii) a separação das competências é mais notória na passagem para este nível;

Esta metodologia é particularmente adequada quando:

1. Existe necessidade de **monitorização contínua** do desajuste educacional, com dados anuais e com elevada granularidade (ocupações, setores, regiões), e é importante capturar mudança estrutural nas ocupações.
2. A referência normativa disponível é **útil mas incompleta**: pretende-se compatibilidade conceptual (por exemplo com EUROSTAT), mas reconhece-se que correspondências fixas podem tornar-se desajustadas ao longo do tempo.

A abordagem continua a depender de carga normativa (enviesamento indutivo) nomeadamente na definição da âncora e da métrica de otimização; este enviesamento indutivo pode ser estendido a outras dimensões justificavelmente, como obrigar os detetores a considerarem complexidades diferentes entre empresas e setores. Contudo, esta prática metodológica permite que estes enviesamentos impostos pelo utilizador sejam apresentados de forma clara, justificada, replicável e auditável, não dependendo do utilizador o resto do procedimento.

## 2. Abordagem metodológica

### Princípios orientadores e relação com a Nota de Análise

A metodologia descrita neste apêndice operacionaliza a proposta apresentada na Nota de Análise: medir desajuste educacional no setor privado com base em dados administrativos (QP). O ponto de partida é reconhecer que o desajuste educacional é relacional (depende do contexto ocupacional/produtivo) e dinâmico (os requisitos de qualificação evoluem). Assim, em vez de impor uma correspondência fixa entre ocupações e níveis educacionais, definimos um padrão de

“normalidade” a partir do comportamento empírico dos próprios dados dentro de contextos comparáveis (células). A âncora normativa inspirada no EUROSTAT é usada como base de verdade comparável: ajuda a calibrar o sistema para que não se torne arbitrário e garante que o indicador continua interpretável em termos conhecidos por utilizadores e decisores.

## Dados, população-alvo e definição de janelas temporais

A fonte principal é o ficheiro anual dos Quadros de Pessoal, com observações ao nível trabalhador–empresa. A calibração e a avaliação são conduzidas com **janelas trianuais** para reduzir ruído anual e refletir a natureza gradual da mudança ocupacional. Definimos três blocos com separação temporal e decisional:

- **Calibração (ano âncora 2011):** 2010–2012 (cal\_data).
- **Validação independente (ano âncora 2013):** 2013–2015 (val\_data).
- **Teste final (ano âncora 2015):** 2016–2018 (test\_data).

O foco substantivo recai na subpopulação com ensino superior (ISCED 5–8). Esta decisão tem duas implicações práticas: (i) a âncora normativa é mais diretamente aplicável e consensual neste domínio, e (ii) a leitura do indicador é mais robusta, tratando-se de uma separação mais vincada face a níveis de habilitações inferiores, evitando extrapolações para segmentos onde os mecanismos de mismatch podem ser estruturalmente distintos.

## Construção de contextos comparáveis

A metodologia exige definir um critério de agrupamento que particiona a amostra em células relativamente homogéneas do ponto de vista de tarefas e requisitos. Exemplos incluem *ocupação + setor*, *ocupação + intensidade tecnológica* ou somente *ocupação*. A função destas células é dupla. Em primeiro lugar, reduzem heterogeneidade não observada: a mesma ocupação pode implicar requisitos de escolaridade diferentes consoante o setor e o tipo de empresa. Em segundo lugar, permitem que o “esperado” seja atualizado ao longo do tempo dentro de cada célula, acomodando requalificação ocupacional e mudanças tecnológicas muito granulares. A escolha do agrupamento é tratada como componente do desenho experimental: deve ser documentada, testada quanto a robustez e acompanhada de regras mínimas de dimensão por célula, para evitar instabilidade estatística.

Nos exercícios conduzidos para a nota de análise em causa, consideraram-se alternativas de composição de células (contexto) como hiperparâmetro (reportados mais adiante) e definiu-se um mínimo de 100 observações por célula.

## Definição operacional das categorias: *under*, *fit* e *over*

O output central do sistema é uma classificação por trabalhador em três categorias operacionais:

- ***undereducation*** (*under*): escolaridade inferior ao esperado no contexto;
- ***fit*** (*fit*): escolaridade alinhada com o esperado;
- ***overeducation*** (*over*): escolaridade superior ao esperado.

## Famílias de detetores avaliadas e o porquê de cada escolha

A detecção empírica de desajuste é implementada com famílias de métodos de *outlier detection*. A comparação entre famílias é uma escolha metodológica para reduzir dependência de pressupostos sobre a forma da distribuição (simetria, normalidade, unimodalidade) e para aumentar robustez a cenários típicos em dados administrativos (células pequenas, caudas longas, multimodalidade e assimetria).

Incluem-se: (i) Distância de Mahalanobis, como baseline para a definição de regiões normais em espaços multidimensionais; (ii) MCD, como alternativa robusta quando outliers distorcem a covariância; (iii) regra Média  $\pm$  Desvio-Padrão, como baseline para a definição de regiões normais em espaços unidimensionais; (iv) MAD, explicitamente robusto a assimetria e caudas, reduzindo risco de limiares artificiais em distribuições assimétricas; e (v) LOF, não paramétrico e local, adequado quando a normalidade depende da densidade local e quando existe multimodalidade dentro de células.

Em termos práticos, cada método tem hiperparâmetros e decisões associadas (por exemplo, limiares de corte, número de vizinhos no LOF, escolhas de variáveis) que são selecionadas de forma sistemática por Nested CV.

## Função-objetivo de otimização (WR) e racional de ponderação

O sistema é calibrado para maximizar uma métrica de recall ponderado que reflete confiança diferencial na âncora por categoria. A função-objetivo é:

$$WR = 0.2 \cdot \text{Recall}_{\text{under}} + 0.6 \cdot \text{Recall}_{\text{fit}} + 0.2 \cdot \text{Recall}_{\text{over}}$$

Esta opção traduz três escolhas substantivas. Primeiro, o objetivo não é apenas “detetar mismatch”; é classificar corretamente estados de adequação, concordando com a métrica normativa, evitando que melhorias numa cauda sejam obtidas à custa de degradação do núcleo. Segundo, a categoria *fit* recebe peso maior (0.6) porque a âncora normativa é considerada mais fiável para identificar adequação no domínio ISCED 5–8; a calibração deve, por isso, priorizar estabilidade e consistência nessa categoria, sob pena de perder credibilidade. Terceiro, os pesos de 0.2 em *under* e *over* preservam incentivos para não negligenciar as caudas, reconhecendo que são substantivamente relevantes e que o indicador deve manter capacidade de as identificar, mesmo que com menor grau de confiança normativa.

Para transparência, a WR é acompanhada de métricas por categoria (recall, precisão e  $F_1$ ), evitando que um valor agregado oculte assimetrias no desempenho.

## Validação cruzada aninhada

A calibração é realizada com Nested CV em três fases decisionalmente separadas. Esta estrutura é desenhada para impedir *data leakage* e para produzir métricas finais não enviesadas, em linha com padrões de avaliação exigentes. As observações são separadas aleatoriamente em treino (para o loop interior), validação (loop exterior) e teste (avaliação final).

### *Loop interior: seleção de hiperparâmetros (calibração)*

No loop interior utiliza-se exclusivamente *cal\_data* (2010–2012). Este conjunto é dividido em treino (70%) e validação interna (30%) com amostragem estratificada por escolaridade, de modo a preservar a distribuição de níveis educacionais e, em particular, a presença relativa de ISCED 5–8. Para cada método e para cada combinação de hiperparâmetros (e, quando aplicável, conjunto de variáveis), o modelo é ajustado no treino interno e avaliado na validação interna, selecionando-se a configuração que maximiza WR (Eq. [eq:wr]). O resultado do loop interior é um conjunto de “melhores hiperparâmetros” por método, obtidos sem recorrer a qualquer informação da janela de validação ou de teste.

### *Loop exterior: seleção do método (validação)*

No loop exterior, cada método é re-treinado na totalidade de *cal\_data* com os melhores hiperparâmetros identificados no loop interior. Em seguida, todos os métodos são avaliados em *val\_data*, que é um conjunto completamente independente do processo de afinação. O método final é o que maximiza WR na validação. Esta etapa separa explicitamente a decisão “qual a melhor configuração dentro do método” (loop interior) da decisão “qual o melhor método” (loop exterior), reduzindo a probabilidade de selecionar um método que apenas se ajusta bem ao ruído específico do período de calibração.

### *Avaliação final: teste não enviesado (teste)*

Finalmente, o método selecionado no loop exterior (com a sua configuração) é avaliado em *test\_data*. Este conjunto não é utilizado em nenhuma decisão. As métricas neste teste são as únicas que suportam afirmações quantitativas sobre desempenho fora da amostra (face à âncora do EUROSTAT).

## 3. Resultados da aplicação aos Quadros de Pessoal

A calibração baseou-se numa pesquisa exaustiva em grelha (*grid search*) sobre múltiplas dimensões do desenho experimental, incluindo:

1. **Famílias de detetores:** cinco métodos de deteção de outliers com propriedades estatísticas distintas

2. **Configurações de variáveis:** desde apenas escolaridade até combinações de até quatro variáveis
  - 1) Escolaridade (`years_of_education`)
  - 2) concentração ocupacional na empresa (índice HHI com ocupações a 3 dígitos, `hhi_occup`)
  - 3) dimensão da empresa (volume de negócios, `log_turnover`)
  - 4) eficiência produtiva (produtividade aparente do trabalho, `log_labour_productivity`)
3. **Estratégias de agrupamento:** três níveis de granularidade na definição de células ocupação–setor
  - 1) Só ocupação
  - 2) Ocupação e CAE a 2 dígitos
  - 3) Ocupação e classificação de intensidade tecnológica/conhecimento do sector
4. **Hiperparâmetros específicos:** limiares de corte e tamanhos de vizinhança

O processo envolveu **452 configurações** experimentadas no loop interior (seleção de hiperparâmetros), seguidas de **5 configurações finalistas** no loop exterior (seleção de método) sendo a **configuração final** avaliada em teste.

## Espaço de Hiperparâmetros por Família de Detetores

### 1. *Mean ± SD (Média ± Desvio-Padrão)*

#### Hiperparâmetros testados:

- multiplicador de desvio [0.8, 1.0, 1.2, 1.4, 1.5, 1.6, 1.8, 2.0]

**Combinações testadas:** 8 limiares × 8 configurações de variáveis × 3 estratégias de agrupamento = **192 configurações**

### 2. *MAD (Median Absolute Deviation)*

#### Hiperparâmetros testados:

- multiplicador de desvio: [2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]
- fator de normalização [1.4826, 1.5]

**Combinações testadas:** 7 limiares × 2 fatores de escala × 8 configurações de variáveis × 3 estratégias de agrupamento = **336 configurações** (estimado)

### 3. *Mahalanobis Distance*

**Racional:** Extensão multivariada da distância normalizada, considera correlações entre variáveis (e.g., escolaridade e HHI, escolaridade e produtividade). Adequado quando desejamos capturar padrões conjuntos, não apenas desvios univariados.

**Hiperparâmetros testados:**

- nível de significância para threshold: [0.01, 0.05, 0.08, 0.1, 0.15, 0.2]

**Combinações testadas:** 6 valores de alpha × 8 configurações de variáveis × 3 estratégias de agrupamento = **144 configurações**

#### 4. MCD (*Minimum Covariance Determinant*)

**Racional:** Versão robusta de Mahalanobis, estima a covariância com base numa fração dos pontos menos outlying. Adequado quando existem outliers que distorcem a matriz de covariância clássica.

**Hiperparâmetros testados:**

- nível de significância para threshold: [0.01, 0.05, 0.08, 0.1, 0.15, 0.2]
- fração de pontos para estimar covariância: [0.9]

**Combinações testadas:** 6 valores de alpha × 1 support\_fraction × 8 configurações de variáveis × 3 estratégias de agrupamento = **144 configurações**

#### 5. LOF (*Local Outlier Factor*)

**Hiperparâmetros testados:**

- tamanho da vizinhança para densidade local: [3, 5, 7, 10, 15, 20]
- contaminação: ['auto'] - Modo automático (estima proporção de outliers a partir dos dados)

**Combinações testadas:** 6 valores de n\_neighbors × 8 configurações de variáveis × 3 estratégias de agrupamento = **144 configurações**

### Configurações de Variáveis Testadas

Todas as famílias de detetores foram testadas com **8 combinações de variáveis:**

5. **Baseline:** years\_of\_education (apenas escolaridade)
6. **+ HHI:** years\_of\_education, hhi\_occup (concentração ocupacional na empresa)
7. **+ Turnover:** years\_of\_education, log\_turnover (dimensão da empresa)
8. **+ Produtividade:** years\_of\_education, log\_labour\_productivity (eficiência produtiva)
9. **+ HHI + Turnover:** years\_of\_education, hhi\_occup, log\_turnover
10. **+ HHI + Produtividade:** years\_of\_education, hhi\_occup, log\_labour\_productivity
11. **+ Turnover + Produtividade:** years\_of\_education, log\_turnover, log\_labour\_productivity
12. **Full:** years\_of\_education, hhi\_occup, log\_turnover, log\_labour\_productivity

## Estratégias de Agrupamento (Células) Testadas

Cada combinação detector + variáveis foi testada com **3 granularidades de célula**:

1. *occupation\_cell (Ocupação apenas)*
  - **Granularidade:** ~100 células
2. *occupation\_sector2d\_cell (Ocupação x Setor 2-dígitos)*
  - **Granularidade:** ~7,500 células
3. *occupation\_sector\_cell (Ocupação x Intensidade Tecnológica)*
  - **Granularidade:** ~1,100 células

## Resultados do Loop Interior: Seleção de Hiperparâmetros

O loop interior foi executado sobre *cal\_data* (2010–2012), com divisão treino/validação interna (70%/30% estratificada). Para cada família de detetores, foi identificada a configuração que maximizou **WR (Weighted Recall)** na subpopulação ISCED 5–8.

*Configurações selecionadas por método (melhores hiperparâmetros)*

Método	Hiperparâmetros	Variáveis	Agrupamento	WR (cal)	F <sub>1</sub> (cal)
<b>Mean±SD</b>	<i>z_threshold=1.6</i>	<i>years_of_education</i>	<i>occupation_cell</i>	<b>0.725</b>	0.866
<b>MAD</b>	<i>z_threshold=2.0, scale_factor=1.4826</i>	<i>years_of_education</i>	<i>occupation_cell</i>	0.601	0.417
<b>Mahalanobis</b>	<i>alpha=0.05</i>	<i>years_of_education, log_turnover</i>	<i>occupation_cell</i>	0.662	0.476
<b>MCD</b>	<i>alpha=0.01, support_fraction=0.9</i>	<i>years_of_education, log_turnover</i>	<i>occupation_cell</i>	0.633	0.594
<b>LOF</b>	<i>n_neighbors=3, contamination=auto</i>	<i>years_of_education, hhi_occup</i>	<i>occupation_cell</i>	0.600	0.413

## Resultados do Loop Exterior: Seleção de Método

Os cinco métodos finalistas (com hiperparâmetros já fixados no loop interior) foram re-treinados na totalidade de `cal_data` e avaliados em `val_data` (2013–2015), conjunto completamente independente.

### Desempenho na Validação (`val_data`, 2013–2015)

Método	WR (Edu 5-8)	F <sub>1</sub> Overall	F <sub>1</sub> (Edu 5-8)	Recall (Edu 5-8)	Precisão (Edu 5-8)
<b>Mean±SD</b>	<b>0.670</b>	<b>0.428</b>	<b>0.737</b>	0.712	0.824
<b>Mahalanobis</b>	0.642	0.345	0.642	0.632	0.792
<b>MCD</b>	0.633	0.308	0.594	0.598	0.802
<b>MAD</b>	0.601	0.251	0.417	0.502	0.852
<b>LOF</b>	0.600	0.236	0.413	0.500	0.352

## Avaliação Final: Teste Não-Enviado

O método selecionado (Mean±SD, `z_threshold=1.6`, `years_of_education`, `occupation_cell`) foi avaliado em `test_data` (2016–2018), conjunto nunca utilizado em decisões.

### Métricas no Teste (`test_data`, 2016–2018)

Métrica	Valor
<b>WR (Edu 5-8)</b>	<b>0.681</b>
<b>F<sub>1</sub> (Edu 5-8)</b>	<b>0.507</b>
<b>Precisão</b>	0.562
<b>Recall</b>	0.491
<b>Accuracy</b>	0.819

### Matriz de Confusão (Overall)

	Predito: UNDER	Predito: FIT	Predito: OVER
Real: UNDER	142,117	1,161	1,616
Real: FIT	100,558	18,230	0
Real: OVER	9,529	0	2,948
















O método empírico (Mean±SD) foi calibrado para aproximar a âncora normativa EUROSTAT (ISCED 5-8 ↔ ISCO 1-3) na população de ensino superior. Os resultados em teste indicam:

- **Concordância global** (accuracy 81.9% em Edu 5-8) razoável, mas não perfeita
- **Precisão 56.2%** indica que 44% das classificações empíricas como mismatch não concordam com a âncora
- **Recall 49.1%** sugere que o método empírico é mais conservador: deixa de detetar ~51% dos casos que a âncora classifica como mismatch

Esta discrepância era **esperada e desejável**: a âncora normativa é fixa e não captura evolução ocupacional; o método empírico adapta-se a mudanças estruturais. A calibração garante que, *em média*, o indicador mantém interpretabilidade próxima da prática EUROSTAT, mas com flexibilidade temporal.

## Notas de Análise

As Notas de Análise são publicações de curta/média dimensão dedicadas a um tema ou questão específica de políticas públicas, podendo revestir a forma de revisão de literatura, análise estatística, sistematização de conhecimento e metodologias existentes

1. Sustentabilidade demográfica e políticas de família 
2. Impacto macroeconómico do choque de inflação importada 
3. Os Salários em Portugal: evolução na última década 
4. A produtividade das empresas em Portugal - Determinantes intrínsecas e de contexto 
5. Semana de quatro dias - Revisão de Literatura e Estudos-Piloto 
6. Como o mercado de trabalho e a igualdade de género influenciam a fecundidade em Portugal 
7. Organização do tempo de trabalho na Administração Pública Central: Inquérito e análise dos resultados 
8. Planeamento de Recursos Humanos em Saúde - Radiografia dos Instrumentos de Planeamento 
9. Trabalho, Liderança e Género no SNS - Nota de Análise 
10. Agenda Nacional de Avaliação - Lições das experiências de outros países 
11. Evolução dos salários por nível de ensino em Portugal: dinâmicas recentes 
12. Envelhecimento na população: perceções, recursos de saúde e respostas sociais 
13. Dinâmica da produtividade do trabalho em Portugal: efeito de produtividade setorial vs. alteração da composição setorial 
14. A força de trabalho do SNS e do Serviço de Saúde da Região Autónoma da Madeira: análise comparativa 
15. Desajuste educacional no mercado de trabalho: incidência e impacto salarial em Portugal 



[www.planapp.gov.pt](http://www.planapp.gov.pt)



PLANAPP



@planapp\_



Newsletter



PLANAPP