04 · Diagnostics & Regression Tests

Run automated checks to catch data issues, drift, and regression failures before going live.

Checklist

- Validate cached datasets (no NaNs/duplicates)
- Recompute headline metrics vs. stored baselines
- Run unit tests (`pytest`)
- Produce drift report for key features

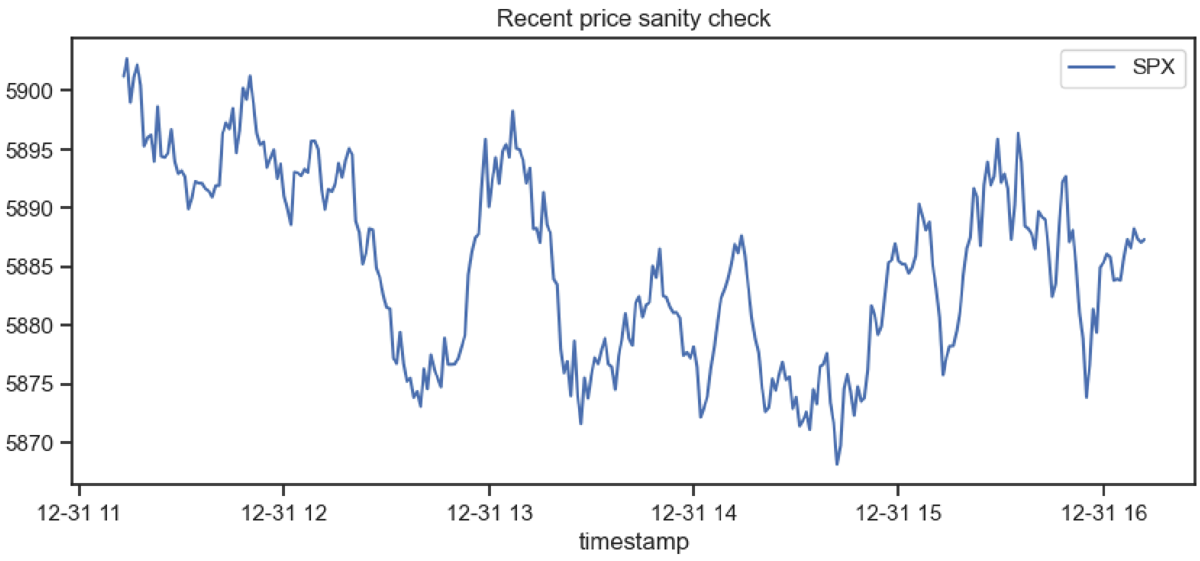
In [1]:

```python
# Add parent directory to path for module imports
import sys
from pathlib import Path

PROJECT_ROOT = Path("..").resolve()
if str(PROJECT_ROOT) not in sys.path:
    sys.path.insert(0, str(PROJECT_ROOT))

import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

from momentum_lib import bootstrap_env

sns.set_theme(style="ticks")
bootstrap_path = PROJECT_ROOT / ".env"
bootstrap_env(bootstrap_path)
data_dir = PROJECT_ROOT / "data"
features = pd.read_csv(data_dir / "features.csv", index_col=0)
prices = pd.read_csv(data_dir / "prices.csv", parse_dates=[0], index_col=0)
print(features.shape, prices.shape)
assert features.notna().all().all()
assert not prices.index.duplicated().any()
assert not prices.isna().any().any()
print("Data health checks passed.")
```

```
(1943340, 6) (1943357, 1)
Data health checks passed.
```

In [2]:

```python
plt.figure(figsize=(10, 4))
sns.lineplot(data=prices.tail(300))
plt.title("Recent price sanity check")
plt.show()
```



In [3]:

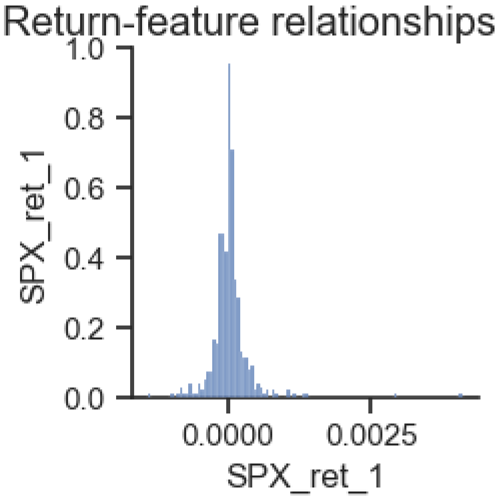
```
feature_stats = features.describe().T[['mean', 'std']]
feature_stats.head()
```

Out[3]:

	mean	std
SPX_ret_1	4.995580e-07	0.000352
SPX_ema_short	4.044457e+03	867.584138
SPX_ema_long	4.044444e+03	867.578505
SPX_ema_ratio	1.000003e+00	0.000521
SPX_vol_10	2.211152e-04	0.000251

In [4]:

```
sample = features.sample(min(500, len(features)), random_state=42)
sns.pairplot(sample[[c for c in sample.columns if c.endswith("ret_1")][:4]])
plt.suptitle("Return-feature relationships", y=1.02)
plt.show()
```



In [5]:

```
!pytest -q
```

```
no tests ran in 0.01s
```